AFHRL-TR-88-3

AIR FORCE HUMAN RESOURCES LABORATORY

AD-A201 329

MULTIATTRIBUTE DECISION MODELING TECHNIQUES: A COMPARATIVE ANALYSIS

Jonathan C. Fast

Metrica, Incorporated
8301 Broadway
Suite 215
San Antonio, Texas 78209

Larry T. Looper

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

August 1988

Final Report for Period September 1986 - September 1987

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

WILLIAM E. ALLEY, Technical Director
Manpower and Personnel Division

HAROLD G. JENSEN, Colonel, USAF
Commander

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S) AFHRL-TR-88-3
6a. NAME OF PERFORMING ORGANIZATION The MAXIMA Corporation 6b OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Manpower and Personnel Division
6c. ADDRESS (City, State, and ZIP Code) 8301 Broadway, Suite 212 San Antonio, Texas 78209	7b ADDRESS (City, State, and ZIP Code) Air Force Human Resources Laboratory Brooks Air Force Base, Texas 78235-5601
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Human Resources Laboratory 8b. OFFICE SYMBOL (If applicable) HQ AFHRL	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F33615-83-C-0030
8c. ADDRESS (City, State, and ZIP Code) Brooks Air Force Base, Texas 78235-5601	10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
62703F	7719	20	07

11. TITLE (Include Security Classification)
Multiattribute Decision Modeling Techniques: A Comparative Analysis

12. PERSONAL AUTHOR(S)
Fast, J.C.; Looper, L.T.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM Sep 86 TO Sep 87	August 1988	72

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
12	04	
05	09	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
decision analysis
decision models
hierarchical decision models
multiattribute decision models
utility analysis

19. ABSTRACT (Continue on reverse if necessary and identify by block number)
This research developed a taxonomy of decision modeling techniques and accomplished a comparative analysis of two techniques developed and used by the Air Force in the areas of personnel selection, job classification, and promotion. The two techniques, policy capturing and policy specifying, were shown to have several characteristics which allowed them to be aligned with existing decision analytical theory. A set of criteria for evaluating the usefulness of a particular technique in a particular decision context was developed, and four techniques were selected for more detailed study and evaluation. The four were: policy capturing, policy specifying, Simple Multiattribute Rating Technique (SMART), and Hierarchical Additive Weighting Method (HAWM). The last two were, respectively, examples of utility assessment and hierarchical decision models. A panel of experts rated each technique over the set of 16 criteria, first without regard to decision context. The panel then rated the match between each of the four techniques and the need for particular criteria in each of three Air Force decision contexts: person-job match, promotion, and research and development project prioritization. Results of the ratings are presented and suggestions made for enhancing the Air Force's ability to conduct decision analytical studies.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL Nancy J. Allin, Chief, STINFO Office	22b TELEPHONE (Include Area Code) (512) 536-3877 22c OFFICE SYMBOL AFHRL/TSR

DD Form 1473, JUN 86 Previous editions are obsolete SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

SUMMARY

This task evaluated two decision modeling techniques used by the Air Force Human Resources Laboratory (AFHRL). These two techniques, policy specifying and policy capturing, were developed by AFHRL and have been used in a variety of decision modeling contexts. However, the relationship between the two techniques and other decision modeling analysis techniques had not been previously investigated.

As part of this task, the research team produced a taxonomy of decision modeling techniques and found that there was a place in the decision modeling literature for the two AFHRL techniques. Policy capturing fell clearly within the well-founded and empirically tested field of statistical/holistic decision modeling. However, policy specifying fits only roughly into the class of direct estimation techniques, but was not similar enough to any technique to conform to an existing axiomatic base.

The task also produced a set of criteria for evaluating the potential usefulness of a decision modeling technique in a particular context. These criteria were applied to four modeling techniques, at first without regard to context, in order to determine their strengths and weaknesses. The four techniques studied were: SMART (Simple Multiattribute Rating Technique), HAWM (Hierarchical Additive Weighting Method), policy specifying, and policy capturing. Policy capturing was found to have many characteristics that make it a very useful decision modeling tool; however, since it is primarily a holistic technique, it should be applied to decision contexts different from the other three methods. Policy specifying was found to be a technique which, although possessing some unique characteristics, could be modified to make it more useful.

Using a rating scheme developed to determine the utility of each technique in any decision context, the four techniques were then evaluated in three separate decision contexts. The three decision contexts studied were: person-job match, research and development project prioritization, and an Air Force promotion board. The criteria and resulting rating scheme proved to be useful for determining the utility of each technique in each decision context.

PREFACE

The work documented in this report is a component of the Force Acquisition and Distribution System Subthrust of the Manpower and Personnel Division's research and development program. The evaluation and comparative study of decision modeling tools will improve the conduct of person-job match and promotion system research, and will provide tools for personnel managers and force planners to make more informed resource allocation decisions to achieve the Air Force's defense mission.

The contract team working on this project included Mr. Jonathan Fast of Metrica, Inc.; Dr. William Stillwell, and Mr. Thomas Martin, of the MAXIMA Corporation; Dr. Detlof Von Winterfeldt, of Decision Insights; Dr. David Seaver, of General Physics Corporation; Dr. Joe H. Ward, Jr., independent consultant; and Dr. Patrick T. Harker, of the Wharton School at the University of Pennsylvania. Dr. Von Winterfeldt substantially authored section II and Appendix A to this report. Mr. Fast and Mr. Larry T. Looper of the Air Force Human Resources Laboratory contributed sections III, IV, V, and VI. The opinions expressed in this report do not necessarily represent the views of all the authors. Metrica, Inc., the MAXIMA Corporation, or the United States Air Force.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONCLUDED)

LIST OF FIGURES

Figure | **Page**

LIST OF TABLES

Table **Page**

I. INTRODUCTION

This is the final report of a task to examine and improve two policy modeling techniques developed at the Air Force Human Resources Laboratory (AFHRL). These techniques, policy capturing and policy specifying, have been very useful in the military personnel decision contexts for which they were developed, but their relationship to the recognized fields of utility and value theory has not been well established. Decision theorists and decision analysts have developed a number of closely related decision modeling approaches such as multiattribute utility/value assessment and hierarchical analysis and have applied these techniques to a number of non-military problems. Although advancements are still being made in these techniques, there is a need to compare and assess these techniques with those being used in military contexts in order to enable analysts and decision makers to make more informed choices among the available techniques in a particular decision context. The examination of these techniques needs to be extended to include the effect of problem context on the determination of usefulness of a technique. This report contains the results of this context-dependent evaluation of the techniques. The ultimate goal of this research and development (R&D) effort is an intelligent computer-based system that contains alternate techniques, and guides the user in making the most appropriate choice of technique for a particular application.

During this research four objectives were accomplished:

1. Survey and review of the available theories and techniques.

2. Development of criteria for comparison of procedures and methods of measuring technical performance.

3. Specification of the relative strengths and weaknesses of each approach.

4. Evaluation of the usefulness and applicability of each technique in selected typical problem contexts.

Nine methods representing four theoretical areas were reviewed during this task, and are reported in Section II of this report. For Section III of this report, four more commonly used techniques (including policy capturing and policy specifying) were selected for further analysis. This section also contains a description of the development of criteria and rating scales for evaluating the four methodologies, and discusses the results of evaluating these techniques without regard to context. Section IV describes the methodology used to extend this analysis to three Air Force decision making contexts. Section V contains the results of the methodology's being applied through the use of expert judgment. Drawing upon this analysis, Section VI contains suggestions for modifying policy capturing and policy specifying in order to make more appropriate and defensible applications of these two techniques. Appendix A includes a discussion of a theoretical underpinning for the judgments made in policy specifying, using difference measurement theory.

II. TAXONOMY OF TECHNIQUES

This section of the report provides a summary of nine multiattribute decision modeling and evaluation procedures that may be useful for military applications such as person-job-matching and other personnel utilization decisions in the Air Force. The procedures are classified into four groups:

1. Riskless indifference procedures
2. Risky indifference procedures
3. Direct estimation procedures
4. Statistical procedures.

Each procedure will be described by its historical origin and main references, elicitation techniques, model forms, and assumptions.

Before discussing the separate procedures, it is helpful to identify the commonalities among them. Jointly, the procedures are often called multiattribute procedures, because they attempt to evaluate decision alternatives on a set of pre-specified value relevant attributes. Most multiattribute procedures go through several or all of the following steps:

1. Identify the objects that are to be evaluated and the purpose of the evaluation; these objects are labelled O(j), j = 1,m.

2. Identify and structure the dimensions on which the objects are to be evaluated; this may be done in a hierarchy with general objectives at the top and specific attributes at the bottom; the attributes are labelled A(i), i = 1,...,n.

3. Develop scales for each of the lower level attributes on which the objects are to be evaluated; the scales are assumed to be measured numerically (real values) and they are labelled X(i).

4. Develop a formal value or utility model that quantifies the tradeoffs among the attribute scales and the attributes; this may be done by developing a utility (value) function u(i) for each scale X(i) and by assigning scaling factors (weights) w(i) to each scale.

5. Recombine the pieces developed in steps 1-4 through some formal aggregation rule that assigns a single value or utility to each object; the most common form is the weighted additive aggregation rule but more complex polynomial forms have been used.

6. Perform sensitivity analyses and select the object(s) with the highest overall value or utility.

The main differences among the procedures discussed in this paper occur in steps 4 and 5, the quantification of tradeoffs and re-aggregation. Techniques for tradeoff assessment (step 4) range from simple rating and weighting techniques based on indifference procedures to construct single-attribute utility functions and scaling factors, to deriving models from holistic ratings of objects via regression or similar model fitting techniques. Aggregation rules (step 5) vary from extremely simple additive rules to complex interactive rules. Yet, in spite of these differences, the structural similarities among various approaches are very strong, and many appliers of multiattribute decision modeling techniques are convinced that the structuring steps (1-3) drive much of the subsequent analysis, independently of the techniques for quantifying tradeoffs or aggregation rules.

Riskless Indifference Procedures

Riskless indifference procedures attempt to evaluate objects in the absence of any risk. They are also applicable in those instances in which risk is present, but the decision maker is risk neutral; i.e., if it appears reasonable to select risky objects according to the expected value of their riskless evaluations. Riskless indifference procedures construct utility functions and scaling factors (step 4 above) by observing tradeoff and indifference judgments. They aggregate these judgments in a variety of forms, including additive, multiplicative, multilinear, and, in the case of conjoint measurement, polynomial rules.

Two broad classes of riskless indifference procedures exist. The first is built on the notion of "strengths of preference" or "preference intensities" and leads to interval scale value functions. It is therefore often called "value measurement." The second is built on the simpler notion of "preference" or "indifference," and leads to weaker representations, which have interval quality only in a restricted sense. This class is usually called "conjoint measurement." Of the two, value measurement has gained increasing acceptance in decision and management science.

Value Measurement

History and Main References. Value measurement has its origin in difference measurement theories first created by Suppes and Winet (1955), and further developed in Krantz, Luce, Suppes, and Tversky (1971). Fishburn (1970) provided an additive extension of difference measurement for the simple case of homogeneous attributes (i.e., all attributes are measured on the same scale, as in time streams of income). Dyer and Sarin (1979), using results by Krantz et al. (1971), Fishburn (1970), and Keeney and Raiffa (1976) provided some generalizations, including multiplicative versions. Although there has been much literature on the model forms and the axiomatic basis of the value model, the status of the judgments required, namely "strengths of preferences," has remained somewhat obscure. Farquahr (1984) provides some discussion of "strengths of preferences" as do Von Winterfeldt and Edwards (1986).

Elicitation Techniques. Two types of elicitations are required for the value measurement models. The first creates a single-attribute value function v_i for each scale X_i. The most common technique for this step is bisection; i.e., finding the point on the scale that is just midway in value between the best and worst scale levels. Arbitrarily assigning the best and worst scale levels a value of 100 and 0, the midpoint then receives a value of 50. Further bisections can refine the value function to any level of detail. Another, less common technique is standard sequences (Krantz et al., 1971) in which segments of equal value differences are "pieced together." Approximation techniques which are, strictly speaking, not appropriate in the indifference framework are direct rating and category estimation (Torgerson, 1960).

The second elicitation creates the scaling factors required for the aggregation of the single-attribute value functions v_i. In value measurement these are obtained by comparing value differences created by stepping from the worst to the best levels in the attribute scales X. The decision maker is first asked to rank order these value increments. In addition, he or she has to make a judgment as to how much the increments differ. This creates a ratio scale of value differences which are mapped into weights w_i by renormalization. A variant of this method is the "swing weight" procedure (Von Winterfeldt & Edwards, 1986). In it, the decision maker is asked to imagine he or she is to be "stuck" with an alternative that is worst on all scales. Which scale level would he or she most like to change from worst to best; which second, third, etc.? This procedure again creates a rank order of value differences, which are then formed into ratio scales and re-normalized.

Model Forms and Assumptions. The most simple difference value model is the additive model:

$$v(o_j) = \sum_{i=1}^{n} w_i v_i (X_j)$$

It assumes that value differences judged in one attribute are independent of the scale levels in other attributes. This assumption is sometimes called "difference independence" (Dyer & Sarin, 1979) or "additive difference independence" (Von Winterfeldt & Edwards, 1986).

A slightly more complex model is the multiplicative model:

$$1 + wv(o_j) = \prod_{i=1}^{n} [1 + ww_i v_i (X_{ij})].$$

It assumes that relative value differences judged in any subset of attributes are independent of fixed scale levels in the other attributes. This assumption is sometimes called "weak difference independence" (Dyer & Sarin, 1979) or "multiplicative difference independence" (Von Winterfeldt & Edwards, 1986).

The most general difference value model is the multilinear model of the form:

$$v(o_j) = \sum_{i=1}^{n} w_i v_i(X_{ij}) + \prod_{i<k} w_{ik} v_i(X_{ij}) v_k(X_{kj}) + \ldots + w \prod_{i=1}^{n} W_{1\ldots n}\, v_i\,(X_{ij}).$$

It assumes that the relative value differences judged in any single attribute are unaffected by fixed scale levels in the remaining attributes.

Weights for the additive model are constructed exactly as described above. For the multiplicative and the multilinear models, additional scaling factors have to be assessed; namely, the w, the w_i's, the w_{ik}'s, etc. These assessments are described in Dyer and Sarin (1979), and they are completely analogous to the assessments for Keeney and Raiffa's (1976) multiplicative and multilinear utility functions.

Conjoint Measurement

History and Origin. Conjoint measurement, as currently applied, is a derivative of the measurement theory model developed by Luce and Tukey (1964) and by Krantz (1964). Independently, these authors developed what was at that time a new theory of fundamental preference measurement, based merely on indifference judgments. Luce and Tukey built this theory essentially "from scratch" whereas Krantz built it on a reduction of the theory of extensive measurement. In its original form, conjoint measurement theory was additive; later extensions by Krantz et al. (1971) included the multiplicative form as well as simple polynomials. Conjoint measurement theory has not been applied as frequently as difference value measurement. However, ideas and concepts of conjoint measurement theory have found their way in a technique called "conjoint measurement" by Green, Carmone, and Wind (1972) and Green and Rao (1974) which basically fits an additive value function to overall evaluations of objects in an orthogonal design of attribute levels.

Elicitation Techniques. In classical conjoint measurement theory, there is no separation of procedures for assessing single-attribute value functions and weights. Instead, the procedure guarantees that the resulting value functions are appropriately matched in their units of scales. This procedure is described by Fishburn (1967) as "lock step," by Krantz et al. (1971) as "dual standard sequence," and by Keeney and Raiffa (1976) as "toothsaw." It begins by identifying an arbitrary step in value on an arbitrary scale X(i). Subsequently a series of steps are laid off on other scales X(k) such that the subjective value increments among these steps are identical to the original arbitrary increment. Since now each scale X(k) is subdivided into steps of equivalent value, the single-attribute value functions can be fitted, and since all value functions are calibrated against the same step, there is no need for weighting.

Model Forms and Assumptions. The most common form is again the additive form:

$$v(O_j) = \sum_{i=1}^{n} f_i(X_{ij}).$$

It assumes that preferences among objects that vary only on a subset of attributes do not change if the remaining (fixed) scale levels are changed. This assumption is called "joint independence" (Krantz et al., 1971).

The multiplicative form, as well as simple polynomials, has been explored by Krantz et al. (1971). Table 1 summarizes the functional forms that have been explored in the conjoint measurement context. As one moves to the more complicated polynomials, the independence assumptions become somewhat unintuitive, and they are therefore not discussed here.

Table 1. Polynomial Forms of Conjoint Measurement (3 Attributes)

ADDITIVE:	$v = f_1 + f_2 + f_3$
MULTIPLICATIVE:	$v = f_1 \cdot f_2 \cdot f_3$
DISTRIBUTIVE:	$v = (f_1 + f_2) \cdot f_3$
DUAL DISTRIBUTIVE:	$v = f_1 \cdot f_2 + f_3$
GENERAL:	

$$v = \sum_{k=1}^{m} a_k f_1^{b_{1k}} f_2^{b_{2k}} f_3^{b_{3k}}$$

Risky Indifference Procedures

Risky indifference procedures require the decision maker to state preferences or indifferences among probabilistic gambles which have multiattributed outcomes. Thus, the evaluation objects are uncertain, and the purpose of constructing a utility function is both to map preferences among outcomes and to map preferences among gambles. The risky indifference procedures discussed here all assume that the decision maker wants to maximize expected utility. In other words, a utility is attached to the outcomes of the gambles and the gambles themselves are ordered by taking the expectation of the utilities of their outcomes. If the decision maker was risk neutral, it would be perfectly appropriate to construct a value function and to use it to calculate expected values to guide preferences under uncertainty. Utility theorists argue, however, that most decision makers are risk averse or risk prone and that property is not captured in the value functions resulting from applying riskless indifference procedures or the value functions these procedures produce.

The Expected Utility Model

History and Origin. Although the expected utility (EU) model has many possible founders, Von Neumann and Morgenstern (1947) are usually credited for the first axiomatic foundation of expected utility measurement. Savage (1954) extended the EU model to include subjective probabilities. Edwards (1954, 1961) introduced the EU model to psychologists and led a series of experimental investigations into its descriptive validity. Today, the EU model is widely used as the logical normative cornerstone of decision analysis (e.g., Holloway, 1979; Keeney & Raiffa, 1976).

Elicitation Techniques. The expected utility model requires elicitation of two entities: probabilities for events and utilities for outcomes. It is standard practice to use direct numerical estimation techniques for eliciting probabilities, and it is equally customary to use indifference techniques for the elicitation of utilities. The direct estimation of probabilities usually takes the form of asking an expert: What is the probability of this event? Or: What are the odds?

The two indifference methods to elicit utilities are the variable probability method and the variable certainty equivalent method (Von Winterfeldt & Edwards, 1986). In the variable probability method, the decision maker is presented with a gamble created by an (unspecified) probability p of obtaining the best possible outcome versus a probability (1-p) of obtaining the worst outcome. For each intermediate outcome, the decision maker is asked to specify a level of p such that he or she would be indifferent between playing the gamble or taking the intermediate outcome for sure. Setting the utility of the worst and best outcomes to be 0 and 1, respectively, the expected utility calculus implies that the utility of the intermediate outcome must be p.

The variable certainty equivalent procedure is similar to the bisection procedure in difference value measurement. The decision maker is presented with a 50-50 gamble for the best versus the worst outcome, and

has to identify an intermediate outcome such that he or she is indifferent between gambling or taking the intermediate outcome for sure. Having obtained this "midpoint," a utility of .5 is assigned to it, which is implied by the EU calculus. The bisection procedure is then followed by offering 50-50 gambles between the worst outcome and the midpoint versus a "sure thing" that lies in between, etc.

Model Form and Assumptions. If a gamble G can be described by k events E(l)...E(k) which are associated with unique (possibly multiattributed) outcomes O(l)...O(k), and if the utility function is denoted by u and the probabilities of the k events are p(l)...p(k), then the expected utility model can be expressed by

$$EU[G] = \sum_{j=1}^{k} P_j u(O_j).$$

The main structural assumption in this model is the sure thing principle. It says that preferences among gambles that vary only in a subset of events should be unaffected by the (common) outcomes in the remaining events. Other assumptions are more technical. Of those, the substitution principle is perhaps the most important one. It requires that it not matter whether a gamble is presented in stages or whether it is presented in one stage by multiplying the probabilities down through the possible paths of the multistage version. Both substitution and sure thing principles are consistently violated as descriptive principles of preferences.

Additive, Multiplicative, and Multilinear Utility Functions

Origin and History. Multiattribute extensions of the expected utility model date back to Fishburn's seminal article (1965a) in which he proved the additive form given certain strong independence assumptions. Pollak (1967), Keeney (1968, 1972), and Raiffa (1969) developed multiplicative models. Keeney (1974) extended the multiplicative models to multilinear ones, and more exotic forms involving independent product terms were later introduced by Farquahr (1975) and Fishburn (1976).

These extensions all begin with the assumption that the expected utility model is valid, and that outcomes have multiple attributes and scales. They then employ independence assumptions of varying degrees of restrictiveness and depending on the validity of these assumptions result in additive, multiplicative, multilinear or other, more exotic aggregation rules being valid.

Elicitation Techniques. As in the difference value models, multiattribute utility functions require construction of single-attribute utility functions and scaling factors. To construct utility functions, the variable probability or variable certainty equivalent methods described previously are used. To construct scaling factors, the following variable probability method is used. The decision maker is presented with a gamble with (unspecified) probability p(i) of winning the outcome which has the best scale values in all attributes and probability 1-p(i) of obtaining the outcome with the worst scale values in all attributes. This gamble is compared to a sure thing that has the worst scale values on all but the i-th attribute, where it has the best scale value. The decision maker is asked to adjust the probability p(i) until he or she is indifferent between the gamble and the sure thing. Given that the utility for the most desirable outcome is 1 and for the least desirable outcome is 0, the expected utility calculus implies that the scaling factor for the i-th attribute (i.e., its weight) is exactly p(i). This solves the problem for the additive case, since it can be shown that the sum of the p(i)'s must be 1 if the model is indeed additive.

In the multiplicative and multilinear cases, additional parameters have to be assessed. These are derived from indifference judgments similar to those made to obtain the p(i)'s.

Model Forms and Assumptions. The additive utility model is very similar in structure to the additive value model:

$$u(O_j) = \sum_{i=1}^{n} k_i u_i (X_{ij}).$$

This model assumes that gambles should be indifferent (have equal utility) whenever they have identical marginal (single-attribute) probability distributions. This assumption is called "marginality" (Raiffa, 1969), "additive independence" (Keeney & Raiffa, 1976), and "additive utility independence" (Von Winterfeldt & Edwards, 1986).

The multiplicative model has the form:

$$1 + ku(O_j) = \prod_{i=1}^{n} [1 + kk_i u_i(X_{ij})].$$

It requires that preferences among gambles which vary only on a subset of attributes be independent of fixed levels in other attributes. This assumption has been called "utility independence" (Keeney & Raiffa, 1976) and "multiplicative utility independence" (Von Winterfeldt & Edwards, 1986).

Finally, the multilinear model has the form:

$$u(O_j) = \sum_{i=1}^{n} k_i u_i(X_{ij}) + \prod k_{ik} u_i(X_{ij}) u_k(X_{kj}) + \ldots + \prod_{i=1}^{n} k_{1\ldots n}\, u_i(X_{ij}).$$

It assumes that preferences among single-attribute gambles are unaffected by fixed values of the outcome in the remaining attributes. In particular, it requires that certainty equivalents for gambles which are uncertain in only one attribute do not depend on the levels of the outcomes in the other attributes. This assumption is called "utility independence" (Keeney & Raiffa, 1976) or "multilinear utility independence" (Von Winterfeldt & Edwards, 1986).

Direct Estimation Procedures

The common thread among the direct estimation procedures is that all parameters of the value/utility function are directly estimated as numbers, ratios, and the like. These procedures tend to lack the axiomatic base of indifference techniques, and are instead grounded in the theory and practice of psychophysical judgments. Their advocates claim that there exist practical advantages of these procedures over the more elegant, yet more complex indifference methods.

SMART

Origin and References. Edwards (1971) developed the Simple Multiattribute Rating Technique (SMART) as a direct response to Raiffa's (1969) article on multiattribute utility theory, which Edwards found extremely stimulating but of limited practical usefulness because of the complexities in model forms and elicitation techniques. SMART was meant to capture the spirit of Raiffa's multiattribute utility procedures, while at the same time being simple enough to be useful for practical-minded decision makers. Through the years Edward's procedure went through several metamorphoses, so that today SMART stands more for a collection of techniques rather than a single procedure (Von Winterfeldt & Edwards, 1986). The most recent versions of SMART are extremely close to the value measurement techniques but still retain much of the simplification spirit that motivated the early version.

Elicitation Techniques. In its simplest form, SMART uses direct rating and ratio weighting procedures for constructing utility functions (Gardiner & Edwards, 1975). First, scales are converted into value functions, either by rating the scale values (if scales are discrete) or by linear approximations (if scales are continuous).

Next, attributes are rank ordered in the order of their importance. The lowest ranked attribute is given an importance weight of 10; the importance of the others is expressed in terms of multiples of 10. The resulting "raw" weights are normalized to add to 1. Because of the range insensitivity of importance weights (see Gabrielli & Von Winterfeldt, 1978; Keeney & Raiffa, 1976), recent SMART weighting methods have been changed to include "swing weighting," which is virtually identical to the weighting methods described in the value difference measurement models.

SMART applications have often used value trees, rather than building the multiattribute model simply on the level of the attributes. In tree applications of SMART, weights are elicited at all levels in the value tree and the final weights for attributes are calculated by "multiplying down the tree." This procedure has a number of advantages (see Stillwell, Von Winterfeldt, & John, 1987) as it facilitates the judgments and allows separation of weighting tasks in an organization between experts (lower level weights) and policy makers (higher level weights).

Model Form. The only model form that has been applied in the SMART context is the weighted additive model:

$$v(O_j) = \sum_{i=1}^{n} w_i v_i(X_{ij}).$$

The Analytic Hierarchy Process

Origin and References. Saaty (1977, 1980) developed, apparently independently of utility theory approaches, his Analytic Hierarchy Process (AHP). It is structurally similar to SMART, but elicitation methods are different and there are several algorithms for reconciliation of inconsistent judgments and for consistency checks that are not available in any of the utility procedures. The AHP has been applied vigorously since its first practical exposition in Saaty's (1980) book, and the widespread application appears to be pushed further along by the introduction of commercially available software packages that implement the AHP algorithms.

Elicitation Methods. The AHP builds heavily on value trees, which Saaty calls "analytic hierarchies." There are no attempts to define or operationalize attributes in terms of scales. Instead, the lowest level of the analytic hierarchy is further split into the alternatives that are to be evaluated. Thus the tree is a mixture of ends (upper levels) and means (lower level).

At each level, a complete set of pairwise comparisons is made between attributes. This comparison first establishes which of two attributes is more important, and secondly, rates the relative importance on a nine-point scale. This rating is interpreted as the ratio of the importance between the two attributes. The procedure elicits more pairwise comparisons than would be necessary to solve for a unique set of weights analytically, and typically produces an inconsistent set of n(n-1)/2 ratio weight assessments. The procedure then goes through an eigenvalue computation, to find a set of weights that best fits the weight ratios provided by the decision maker.

This procedure is repeated at each level of the tree. The process is most different from utility and value elicitation procedures in its elicitation of "weights" at the lowest level of the tree. Recalling that the lowest level consists of the alternatives O(j), the procedure elicits pairwise judgments of how much more of the next level attribute one alternative possesses than another. These judgments are again reconciled using the eigenvalue procedure. Thus, the lowest level weights most closely correspond to single-attribute value judgments or ratings in utility theory.

Model Form and Assumptions. The model form is the simple additive model, as used in SMART. The theoretical assumptions are similar to those of ratio measurement, although, with the exception of an attempt by Vargas (1984), they have not been spelled out explicitly.

Policy Specifying

Origin and References. Policy specifying was developed specifically for analyzing complex hierarchical models for which simple additive forms are inadequate (Ward, 1977; Ward, Pina, Fast, & Roberts, 1979). It combines elements of value and utility models with those of analytic hierarchies. Its main application areas are in Air Force person-job-matching problems.

Elicitation Techniques. Ward et al. (1979) described the policy specifying procedure in some detail. First, a hierarchy is constructed which is very similar to a value tree. One constraint is that there should be only two branches at each node. This constraint is due to the limits of the practicability of the elicitation procedure for more than two attributes.

Assessments are then made for each pair of attributes in the hierarchy. The assessment process begins by specifying worst and best levels for each attribute and assigning ratings (from 0 to 100) to all four corner points. Next, the functional form for the value function in each attribute is specified, and finally, the aggregation rule is defined that fits the four corner points. Once all functional forms and value functions at each lower level pair are constructed, the analysis moves on to higher levels and creates functions of functions, and so on. At the higher levels, the judgmental task becomes almost identical to the holistic procedures discussed in the statistical models.

Model Forms and Assumptions. The model forms can be substantially more complex than any of the additive, multiplicative, or multilinear ones discussed earlier. This is due to the nested process of model building, which generates, even at moderate degrees of higher level model complexities, rather complicated overall model structures. Although there are no theoretical restrictions of the complexities of the model forms in the policy specifying process, it will be rare to see model forms that are different from simple polynomials (Krantz et al., 1971).

There are no explicit assumptions for the model forms developed in the policy specifying context. However, as long as the model forms remain in the context of simple polynomials, two theoretical underpinnings are possible: the conjoint measurement theory of simple polynomials (Krantz et al., 1971) and a hierarchical theory of multilinear dependence in the difference value measurement sense. This second theory is explored further in Appendix A to this report. If further developed, it would seem in principle to be able to provide independence tests for the policy specified models.

Statistical/Holistic Procedures

Statistical procedures all attempt to build a linear statistical model that relates holistic judgments of the overall worth of alternatives to their attribute levels. All such procedures rely heavily on obtaining large numbers of subjective holistic judgments. Component scales for attributes of the alternatives are usually observable, physical features; however, some procedures include scaling of component attributes when the number of possible levels is small.

Holistic evaluations of alternatives or hypothetical alternatives are obtained in various ways, usually via some subjective-estimate method, such as rating scales or magnitude estimation. Two primary characteristics distinguishing one statistical procedure from another are: (a) whether the component attribute scales are discrete and not clearly ordered with respect to overall alternative worth, and (b) the number of holistic judgments used to build the statistical model.

Both of these considerations combine to determine the exact statistical model employed. In general, analysis of variance or regression analysis is used whenever attribute scales are discrete or nonmonotonic; of course, these analysis techniques can also be used for continuous, monotonic scales by choosing fixed points along the continuum to represent discrete levels of the dimension. Assuming that there are no serious three- or more-way interactions, the total number of holistic judgments can be substantially reduced by substituting either a fractional replication design or an orthogonal design for the complete factorial design normally employed.

Policy Capturing

History and Origin. Policy capturing owes its conceptual heritage to Egon Brunswik's probabilistic functionalism approach to psychology (Hammond, 1966). Although there are certain prescriptive components of policy capturing, the primary historical concern has been descriptive. Brunswik's "lens model" has spawned a great deal of laboratory research on how people combine information about different aspects of a stimulus to form an overall evaluation of the stimulus. Slovic and Lichtenstein (1971) provided a good historical overview of this research, known as the "multiple cue probability learning" (MCPL) paradigm. In a typical MCPL experiment, the relationship between stimulus dimensions and overall worth is manipulated, and the experimenter studies subjects' learning of the (arbitrary) relationship.

The lens model was also applied in real-world settings to study how expert decision makers combine information on relevant dimensions to form an overall evaluation. Various terms were used to describe this research, including "policy capturing" and "bootstrapping." Initially, the goal of policy capturing research was also descriptive (Christal, 1967). One robust conclusion of the policy capturing literature is that a linear model built from the expert's judgments will outperform the expert when applied to a new stimulus sample (Goldberg, 1968, 1970a, 1970b; Meehl, 1954, 1965; Sawyer, 1966). Other work in policy capturing has utilized this finding to focus on the normative properties of policy capturing models (Hammond, Stewart, Brehmer, & Steinmann, 1975; Jones, Mannis, Martin, Summers & Wagner, 1975).

Elicitation Techniques. There is no formal procedure for determining the relevant attributes for inclusion in the policy capturing analysis. Once they have been determined, scales are built for each of the attributes. In most cases, these are (a) continuous (i.e., interval level measures); (b) observable, physical characteristics; and (c) linearly (or at least monotonically) related to overall worth. Although the policy capturing approach can accommodate attributes that are (a) less than interval level, (b) subjectively judged, and/or (c) nonmonotonically related to overall worth, these cases are by far the exception to the rule.

The next step is to collect a sample of either real or hypothetical alternatives that are in some sense representative of the population of alternatives likely to be evaluated with the resulting model. One or more experts are then asked to judge these alternatives with respect to some aggregate criterion, such as "overall desirability." These judgments are normally obtained via a rating scale or subjective estimate response mode. There is virtually no empirical work in the policy capturing literature on the effects of the exact response mode used to obtain holistic judgments; what little research there is derives from study of the effects of response modes in perceptual research.

Once an expert's holistic judgments are obtained, a statistical parameter estimation technique such as multiple linear regression is applied to the data. The expert's holistic judgments are treated as the dependent (or criterion) variable, and the attribute scales are treated as the predictor variables. The formulae for obtaining least squares estimates of the model parameters (weights) are well known, and many computer programs are available for this purpose.

In some cases, model predictions and/or model parameters may be fed back to the expert decision maker, and new holistic judgments may be obtained. In any event, the parameters and implications of the model are usually discussed with the expert in an informal manner to check for any obvious errors, either in determination of relevant attributes, specification of attribute scales, or in holistic judgments of overall worth. Once the expert's model is agreed to, it may be used to evaluate alternatives quantitatively, with the decision rule being to choose those alternatives with the highest model scores.

Model Forms and Assumptions. A schematic of Brunswik's lens model is presented in Figure 1, adapting the notation used by Dudycha and Naylor (1966) to policy capturing terminology.

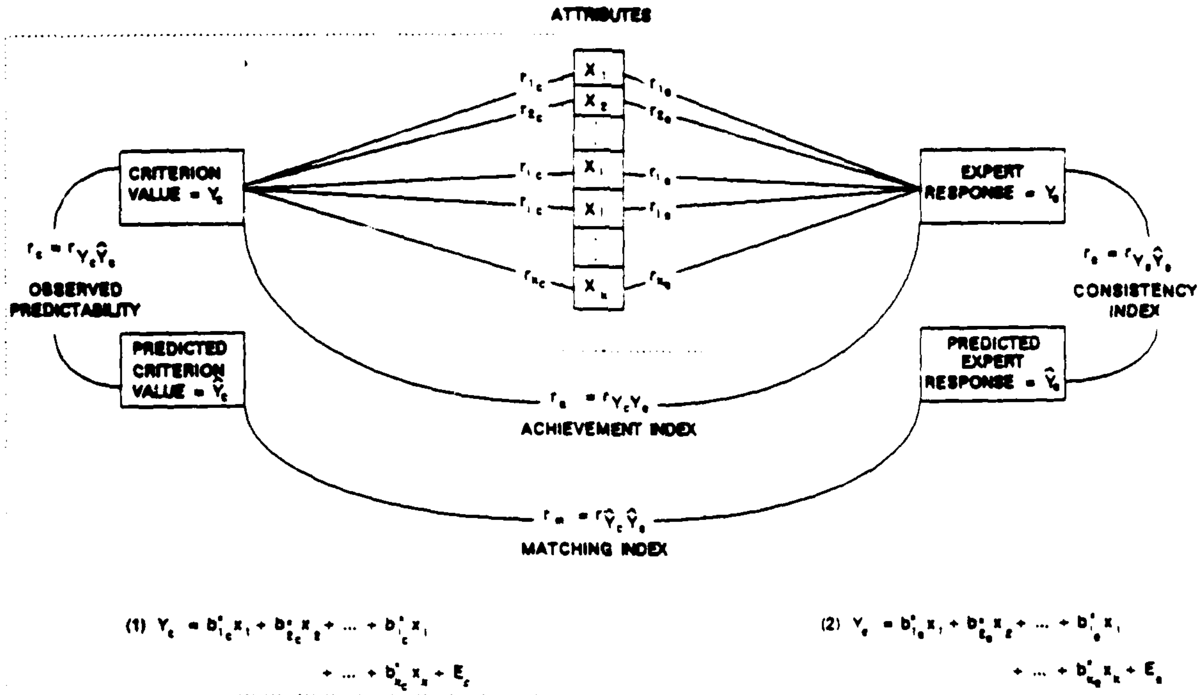

Figure 1. Modified Lens Model.

In the application of the lens model to policy capturing, it is assumed that the criterion value (Y_c) is not observable; otherwise, one would presumably build a statistical model based on the observed criterion values and subjective holistic judgments would not be necessary. In addition to the Y_c it is not possible to calculate indices within the dotted line; specifically, matching (r_m), observed predictability (r_c), or achievement (r_a).

Of course, a consistency index (r_e)' as well as predicted expert responses ($\hat{Y}_e$), can be calculated. The usual policy capturing procedure is to calculate the least squares estimates of the weighting parameters in:

$$y_j = b_0 + \sum_{k=1}^{n} b_k X_{jk} + e_j$$

where y_j is the expert's judgment for profile j, n is the number of attributes, b_k is the raw score regression weight on attribute k, X_{jk} is the value of attribute k on alternative j, b_0 is a constant term, and e_j is the residual error from the model of the expert for alternative j.

Variations on the standard least squares regression formulation include nonlinear transformations on the attribute scales and estimation methods other than least square. The usefulness of the derived model is a function of how closely the predicted expert responses (Y_e) match the (unobservable) criterion values (Y_c) (as measured by r_e).

Holistic Orthogonal Parameter Estimation (HOPE)

History and Origin. Barron and Person (1979) first proposed HOPE as an alternate method for eliciting multiplicative value and utility functions, as defined by Keeney and Raiffa (1976). HOPE requires the decision

maker to provide holistic interval level evaluations of a relatively small subset of m choice alternatives. The alternatives are chosen to form an orthogonal design (e.g., latin square, greco-latin square), an extreme type of fractional replication design. These judgments are then used to generate n equalities, from which n parameter estimates can be solved. The essential difference between HOPE and the methods proposed by Keeney and Raiffa (1976) is in the flexibility afforded by HOPE in selecting the alternatives upon which judgments are based. The basic strategy, however, of solving n unknowns is the same.

HOPE shares assessment of holistic judgments and use of orthogonal designs with the more traditional approaches of conjoint measurement techniques. Because of the large number of judgments often required in complete factorial designs, many applications of conjoint measurement have used fractional replication designs in place of complete factorials when higher order interactions are deemed unimportant (Green, 1974).

The orthogonal designs of HOPE are simply "highly fractional replications" in which all interaction effects are nonrecoverable. The mathematics necessary to solve the n equations for n unknowns is identical to that resulting from applying a standard ANOVA (or regression with dummy coded variables) to the orthogonal design.

Elicitation Techniques. The HOPE methodology closely follows the value and utility elicitation techniques described earlier, but differs in terms of the judgments required to estimate single-attribute value (utility) functions and scaling parameters. Once value-relevant attributes have been identified, discrete levels representative of the range of available alternatives are determined for each attribute. An orthogonal design of choice alternatives is then constructed based on the attributes and discrete levels identified. There is always more than one orthogonal design, and Barron and Person (1979) suggested choosing one comprised of "believable" alternatives.

Holistic evaluations of the alternatives in the orthogonal design are then obtained from the decision maker. Barron and Person (1979) suggested assigning a rating of 0 to the alternative worst on all attributes and 100 to the alternative best on all attributes, and some rating in between to the remaining alternatives. Any of the elicitation methods used in conjoint measurement or policy capturing is acceptable. Of course, since many fewer judgments are required, the decision maker may be able to spend more time reflecting on each alternative than is the case when a full factorial design is used.

Barron and Person (1979) suggested that risky utility functions can be assessed by treating each alternative in the orthogonal design as a sure thing consequence in comparison to a standard gamble with alternatives best on all dimensions and worst on all dimensions as the uncertain outcomes. The indifference probability then plays the role of the riskless rating scale holistic judgments.

In addition to the judgments of alternatives in the orthogonal design, one extra evaluation is required of an alternative that is the complement of one of the alternatives in the orthogonal design. (For example, the complement of an alternative worst on attribute A(j) and best on all others is that alternative best on attribute A(j) and worst on all others.) The complementary alternatives are used to estimate the value of w (or k) in the multiplicative value (or utility) model.

Model Forms and Assumptions. HOPE is applied within the general framework of a multiplicative value (or utility) model, and thus the notation and equation are the same as in the above discussion of multiplicative value and utility models. If the sum of the evaluations of the two complementary alternatives is equal to the evaluation of the alternative best on all attributes, then w (or k) is zero or near zero, and the multiplicative model can be reduced to the additive model.

Once w (or k) is estimated and the model form is selected, the elicited holistic judgments are used to generate an equation for each of the scaling parameters, w(i) or k(i), in the model. In addition, equations are generated to estimate the single-attribute value (utility) associated with each level of each attribute in the orthogonal design.

In the additive value case, for example, w(i) is simply the difference between the average evaluation of alternatives best on attribute A(i) and the average evaluation of alternatives worst on attribute A(i). Likewise,

the product of w(i) and the single-attribute value of an intermediate level on attribute A(i) is the difference between the average evaluation of alternatives at that intermediate level on attribute A(i) and the average evaluation of alternatives worst on attribute A(i).

When the value model is determined to be multiplicative, w or k can be estimated directly from the w(i) or indirectly from the estimates of the overall value of the complementary alternatives.

Unlike policy capturing, there are no replications and no interaction terms, hence no way to estimate error terms in the linear model. In most cases, the number of judgments (hence independent equations) will equal (or only slightly outnumber) the number of unknown parameters.

III. EVALUATION OF TECHNIQUES

In this section discussion will focus on four of the techniques that will be examined in more detail. These techniques are: policy capturing, policy specifying, SMART, and a software implementation of AHP called the Hierarchical Additive Weighting Method (HAWM). Policy capturing and policy specifying are included because they are two techniques often used by the Air Force in many decision contexts. SMART was chosen because it is a representative multiattribute utility/value assessment methodology; and HAWM, because it represents a hierarchical weighting technique. SMART and HAWM were also selected because they are gaining widespread acceptance as techniques that are very useful in modeling policies and decisions across a variety of contexts. Both SMART and HAWM are also relatively easy to use and understand, and many decision analysts are depending on one or both of these techniques for solutions to their decision analytic problems.

Common Example

In order to clarify the exact form of the techniques that are being evaluated in this study, this section provides a more detailed discussion of each technique, with an example of the application of that technique to a common personnel problem--promotion decisions. Each of the methodologies to be evaluated in this study has received significant applied use; but disagreement exists, even among experts, as to the best form in which to apply them. In addition, different practitioners will be more or less familiar with the specific details of alternative approaches. This section is directed toward making clear exactly the form and application of each technique being evaluated in this study.

The example selected is that of a hypothetical Air Force enlisted promotion system, a system which would be used by the Air Force to determine which individuals eligible for promotion will in fact be selected for promotion. This system uses a number of personnel attributes to determine a rating for each candidate. In this example, six attributes were chosen for use in developing these decision models: (a) scores from a job knowledge test (JKT), (b) scores from the general organizational knowledge test (GOK), (c) time in service (TIS), (d) time in grade (TIG), (e) awards and decorations (AD), and (f) an individual performance rating (IPR).

The JKT and GOK would be tests of the airmen's knowledge of their area of specialization and of general military subjects and management practices at their level, respectively. Each test results in a score on a percent correct scale (that is, 0-100). In this promotion system, TIS and TIG will be measured in months. The score for awards and decorations is assigned according to the order of precedence of the award or decoration. For example, combat-related decorations receive higher value scores than do non-combat service awards. The IPR, given by the airman's supervisor, results in a rating score which is averaged over several recent ratings.

SMART

We will describe the most recent version of SMART as discussed in Von Winterfeldt and Edwards (1986). This version merges the original SMART technique proposed by Edwards (1971, 1977) with the difference measurement theory proposed by Dyer and Sarin (1979).

Elicitation of Single-Attribute Value Functions

The first step in SMART involves development of single-attribute value functions which are constructed by arbitrarily assigning the "worst" level of a single-attribute scale a value of 0 and the "best" level a value of 100. Levels in between are rated on a continuous scale between 0 and 100, with instructions to consider carefully the difference in value between levels. If the underlying scale is numerical and continuous, curve drawing procedures are often substituted for this rating technique.

In the context of this promotion example, consider the attribute AD with several levels ranging from "no award or decoration" to the highest level consisting of numerous awards. As illustrated in Figure 2, the level "no award or decoration" would receive a value of 0. The highest level, consisting of several examples of exemplary award combinations, would receive a value of 100. Next, the analyst would pick any of the intermediate levels on the scale, including individual awards and combinations of awards, and ask:

```
0          5                                                                  100
|----------|------------------------------------------------------------------|
No         Purple                                Combination 1.  Medal of Honor
Decoration Heart                                                  Purple Heart,
                                                                    Silver Star

                                                 Combination 2.   Distinguished
                                                                  Service Cross,
                                                                   Purple Heart
                                                                 (Three Awards)
```

Figure 2. Example of a Rating Scale for Assessing a Single-Attribute Value Function.

"On a scale from 0 to 100, where would 'Purple Heart' fall between 'no decoration' and the exemplary combination levels?" The decision maker may feel that a 'Medal of Honor/Purple Heart' combination is much more valuable than the 'Purple Heart' alone and assign a value of 5 to 'Purple Heart' and an 80 to the 'Medal of Honor/Purple Heart' combination. Similarly, other levels can be rated in between 0 and 100, thereby providing the full underlying value functions.

To illustrate the curve drawing procedure, consider the attribute TIG, ranging from 0 months to 120 months. A decision maker may be asked to draw a curve reflecting the relative value of different levels of TIG between 0 and 100. As the illustration in Figure 3 indicates, the relative value increments may initially be small, since to consider promotion possibilities, an airman must at least have served a minimum length of time in the present period. After a period of acceleration, the value of additional time in grade may level off. The nature and implications of such curves are discussed in detail with the decision maker in order to arrive at a final shape.

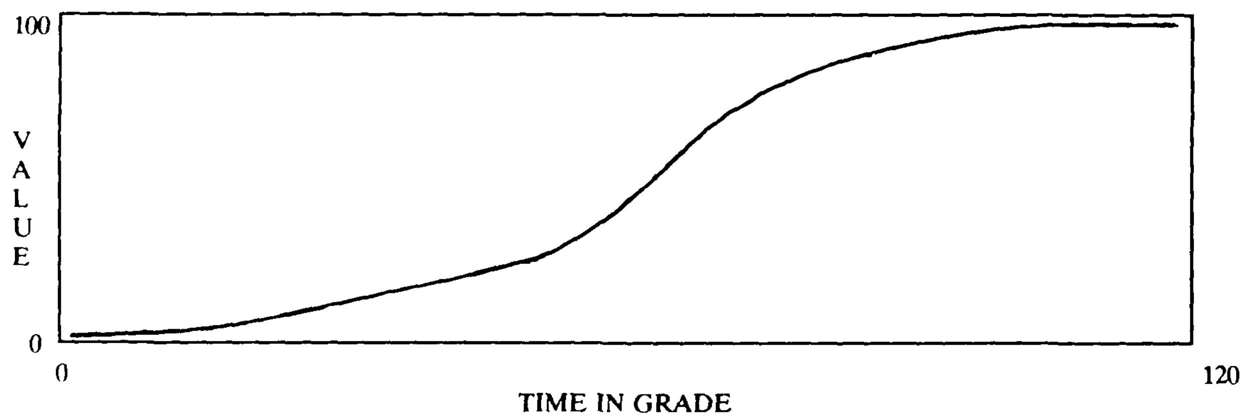

Figure 3. Example of the Curve Drawing Technique to Assess a Single-Attribute Value Function.

Elicitation of Weights

Weights in SMART are assessed by the "swing weighting" method, in which the analyst presents the decision maker with a profile of a hypothetical alternative that has the worst level on each attribute and another hypothetical alternative that has the best level on each attribute. The decision maker is then asked to assume that he or she is "stuck" with the worst alternative, but has an opportunity to move one (and only one) attribute level from its worst to its best level. Which attribute would be most desirable to move? In other words, which change from worst to best level would add the most overall value in terms of determining the promotability of individuals? After identifying the attribute that provides this largest change or "swing," the decision maker identifies the attribute with the second largest change, the third largest, etc. This process provides a rank order of the weights in SMART.

Next, the decision maker is asked to consider the value difference created by stepping from the worst to the best level in the most important attribute (i.e., the one that was chosen first), and to arbitrarily assign that value difference a score of 100. Similarly, an attribute for which the swing would make no difference in value at all is assigned a weight of 0. All other attribute swings are then given weights between 0 and 100. For example, an attribute that has the potential of adding half the overall value of the highest ranked attribute would receive a weight of 50. The resulting "raw" weights are summed up and each weight is divided by the total sum of the weights to create normalized weights that sum to one. When attributes are hierarchically structured, weights are assigned at each level of the hierarchy, and final attribute weights are obtained by multiplying the upper level weights by the lower level weights.

The swing weight method in the promotion example would be accomplished by asking the decision maker to rank order the desirability of moving an attribute from its worst to its best level. The decision maker might likely rank IPR score as the number 1 attribute, as a low IPR score would essentially make the candidate unpromotable. Following this change, the next most desired change may be in JKT, GOK, and AD. all of which may be considered to add approximately equal value to the promotion decision model. Next comes TIS, and TIG is last.

The swing in value in the IPR attribute would then be given a weight of 100 points. All other weights are expressed in values between 0 and 100. Hypothetical results are shown in column 3. These raw weights are highly skewed, because the IPR attribute produces an extreme swing in value (in practice one might worry about the definition of the endpoints of that scale, or refine this attribute by breaking it down into subattributes). Normalization of these weights is done mechanically. At the bottom of column 3 of Table 2 is the sum of the raw weights and in column 4 are the normalized weights, which, of course, total 1.00.

Table 2. Illustration of the Swing Weighting Technique

Attribute	Rank of swing	Raw weight of swing	Normalized weight
JKT	2	10	.07
GOK	2	10	.07
TIS	3	5	.04
TIG	4	1	.01
AD	2	10	.07
IPR	1	100	.74
		sum: 136	sum: 1.00

To illustrate hierarchical weighting, consider the tree structure in Figure 4. In this case it might be logical to first weight JKT versus GOK with respect to the knowledge (KNOW) objective only, then to weight TIS versus TIG with respect to the time (TIME) objective only. This can be done with the swing weighting procedure exactly as described above, and it would produce the results indicated in Table 3a. Next, weighting of the relative swings of the four higher level objectives KNOW, TIME, AD, and IPR is done by asking the

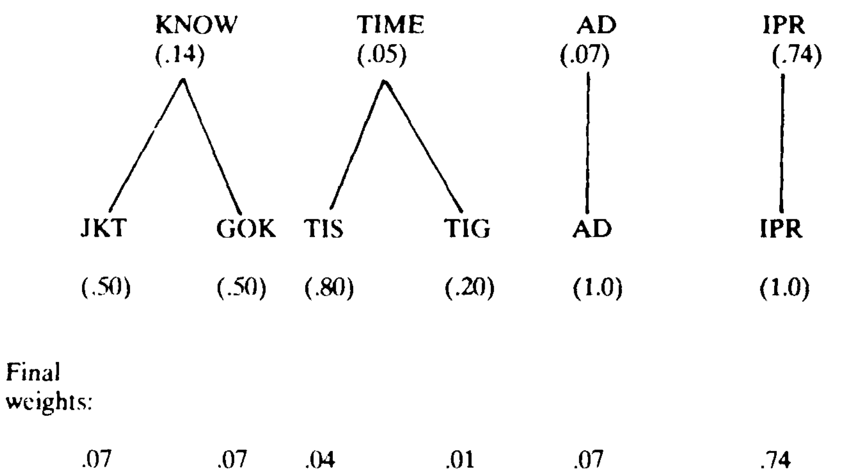

Figure 4. Illustration of a Hierarchical Tree Structure with Hierarchical Weights.

decision maker to simultaneously consider swings of attributes under the objectives that are to be weighted. A specific question might be: "Would you rather change both JKT and GOK from their worst levels to their best levels or change both TIG and TIS from their worst to their best levels?" The answer to this question would provide a rank order of the weights for KNOW and TIME. The questions regarding the other two attributes (AD and IPR) would be identical to those illustrated in the non-hierarchical case. Together they might provide a rank order as shown in Table 3b. Raw and normalized weights are also shown in that table. The final weights for the lower level attributes JKT, GOK, TIS, and TIG are obtained by multiplying the upper normalized weight with the respective lower level normalized weight (see Figure 4).

Table 3. Illustration of Hierarchical Swing Weighting

	Attribute	Rank of swing	Raw weight	Normalized weight
3a (Lower level)				
KNOW				
	JKT	1	100	.50
	GOK	1	100	.50
TIME				
	TIS	1	100	.80
	TIG	2	20	.20

Table 3. (Concluded)

	Attribute	Rank of swing	Raw weight	Normalized weight
3b (Upper level)				
KNOW	JKT			
		2	20	.14
	GOK			
TIME	TIS			
		4	6	.05
	TIG			
	AD	3	10	.07
	IPR	1	100	.74

Aggregation of Weights and Single-Attribute Values

The aggregation of weights and single-attribute values is accomplished as follows. For each alternative O_j, a profile of attribute levels X_{ij} is generated which indicates the degree to which that alternative scores on the attributes. The X_{ij}'s are converted into single-attribute values $v_i(X_{ij})$ which are simply read off the value curves and graphs as shown in Figures 2 and 3. The overall value of the alternative is then calculated by the formula

$$v(O_j) = \sum_{i=1}^{n} w_i \, v_i(X_{ij}).$$

In the promotion example, a promotable candidate may have the profile described in column 2 of Table 4. The associated single-attribute values and weights might be as shown in columns 3 and 4. Multiplying weights and single-attribute values generates column 5, and adding these cross-products produces the overall value of 71.95 for this candidate.

Table 4. Illustration of the Computation of Aggregate Value for a Promotion Candidate

Attribute	X_{ij} Candidate O_j's Scoring profile	$V_i(X_{ij})$ Relative Single-Attr. values	W_i Weights of Attr.	$W_iV_i(X_{ij})$ Cross-Products
JKT	50 points	50	.07	3.50
GOK	75 points	75	.07	5.25
TIS	60 months	50	.04	2.00
TIG	12 months	25	.01	.25
AD	AF Commend.	25	.07	1.75
IPR	100 points	80	.74	59.20
			Total value:	71.95

The Hierarchical Additive Weighting Method (HAWM)

Like SMART, Saaty's Analytic Hierarchy Process (AHP) has undergone several metamorphoses (Saaty, 1977, 1980, 1986). The version of AHP discussed below is based on the techniques implemented in the Hierarchical Additive Weighting Method (HAWM) software that was developed for the IBM PC/XT by Kansas State University (Hwang & Yoon, 1981).

HAWM begins with a hierarchical structure of the evaluation problem, with top values that are very much like a SMART structure. However, at the bottom, the alternatives fan out under each attribute as yet another level of evaluation in the tree. Figure 5 presents the HAWM analog for the promotion example. Here promotable airmen are the alternatives (O_j) and are repeated at the bottom of the tree.

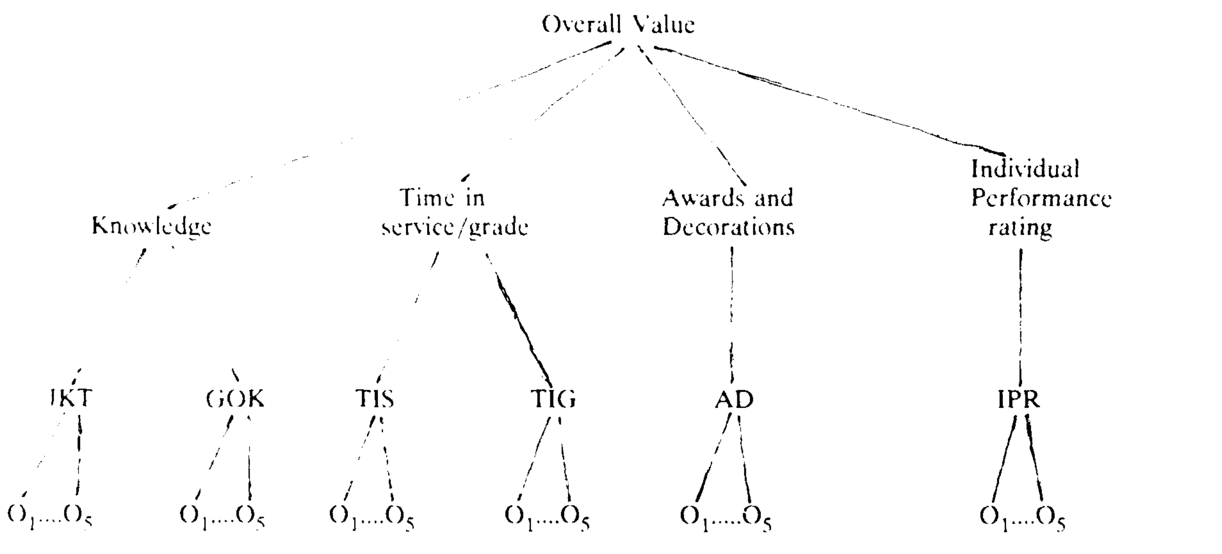

Figure 5. Illustration of an Analytic Hierarchy.

Elicitation of Weights

The HAWM process begins by eliciting weights in the upper part of the tree. Weights are also elicited for the bottom level (the alternatives) to indicate their relative desirability in achieving bottom level objectives or attributes. Since that step is somewhat similar to the value function assessment in other procedures, it will be discussed in a separate section.

In the upper part of the tree, weights are interpreted to reflect the "relative importance" of the attributes or objectives. Weights are assessed under each node, comparing possible attributes with pairwise weight judgments. The decision maker is presented with one pair and asked:

1. Which attribute do you think is more important?
2. On a scale from 1 to 9, how much more important is that attribute (1 meaning equally important, 9 meaning much more important)?

The numbers obtained from these weighting judgments are considered weight ratios and entered into an n x n (attributes by attributes) matrix of weight ratios in which the diagonals are set to 1. In the HAWM, the additional assumption is made that the weight ratios must be reciprocal. Thus, a set of n(n-1)/2 weight ratios fills the complete n x n matrix that defines weight ratios at each node.

Having obtained n(n-1)/2 weight ratios, the HAWM solves for the "best-fitting" set of normalized weights; i.e., those weights that can best reproduce the (possibly inconsistent) assessed weight ratios. The HAWM solves these weights as the eigenvector of the weight ratio matrix. In addition to providing the best-fitting weight solution, the HAWM also provides an index of (in)consistency which ranges from 0 (perfect consistency) to 1 (highly inconsistent weight ratio assessments).

In the promotion example, the upper level weight ratio assessment might produce a weight ratio matrix such as the one shown in Table 5. The circled numbers are the assessed ones. The others are inferred from the reciprocity assumption. The diagonals are simply assumed. The last column of Table 5 shows the weights derived from the HAWM program (as run in the HAWM software) and indicates that there is moderate consistency in the weight ratio assessments. After such an initial assessment, the decision maker is asked if the ratios should be revised or kept unchanged.

Table 5. Illustrative Weight Ratio Assessment

	KNOW	TIME	AD	IPR	Normalized weights
KNOW	1	3	2	1/9	.13
TIME	1/3	1	1/2	1/9	.06
AD	1/2	2	1	1/9	.09
IPR	9	9	9	1	.72
			Inconsistency score: .054		

If satisfied with the current assessment, the decision maker goes on to lower level nodes of the value tree, repeating the process described above. In the example, there are only two lower level nodes: JKT versus GOK and TIS versus TIG. The decision maker is asked to provide relative weight ratios for each of these pairs considering the contribution to achieving the next higher objective (KNOW or TIME). Both weight assessments would generate 2 x 2 matrices, with no possibility for inconsistencies. For example, the assessed weight ratio of JKT versus GOK might be 2, resulting in relative weights of .67 for JKT and .33 for GOK. Similarly, the assessed weight ratio for TIS vs. TIG might be 3, resulting in relative weights of .75 for TIS and .25 for TIG. Since there exists no possibility for inconsistency, the results are identical to those obtained by simply normalizing the raw weight ratios.

Preference Scores

Once the bottom level of alternatives is reached, the decision maker has two choices in HAWM: Either continue the judgments of relative importance or produce judgments of the relative preference of the alternatives with respect to achieving the lowest level attribute. Since in the context discussed here, the latter interpretation is more intuitive, only this variant of the HAWM will be discussed.

Under each lowest level node, and for each pair of alternatives, the decision maker is asked:

1. Which of the two alternatives do you prefer with respect to the attribute under consideration;
2. On a scale from 1 to 9 (1 meaning indifference, 9 meaning extreme preference), how much do you prefer this alternative on the attribute under consideration?

As in the importance weight assessment, the relative preference assessments are assumed to be reciprocal, so that n(n-1)/2 assessments are sufficient to fill out the complete n x n matrix. The final scores for each alternative are again the eigenvector of that matrix that best matches the relative preference ratios.

To illustrate this process in the promotion context, consider the attribute JKT and assume that five promotable candidates have differing levels of that attribute. A preference comparison of these five airmen might look like the one in Table 6. The last column in that table indicates the renormalized scores that each of the candidates receives as a result of the relative preference judgments. The consistency index shows that the assessments were somewhat inconsistent.

Table 6. Illustration of Relative Preference Assessments for Five Promotion Candidates on the JKT Attribute

	Candidate					
	O_1	O_2	O_3	O_4	O_5	Relative Score
O_1	1	6	3	2	1	.33
O_2	1/6	1	1/2	1/3	1/6	.05
O_3	1/3	2	1	2	1/3	.14
O_4	1/2	3	1/2	1	1/2	.14
O_5	1	6	3	2	1	.33

Inconsistency score: .034

Aggregation Rule

The results of weighting and preference assessments are aggregated in a form that is very similar to the SMART rule by multiplying down the tree and adding the multiplicative elements for each alternative. Consider the example set of relative weights and preference judgments in Figure 6. The overall value of alternative O_1 would be calculated by multiplying down all the normalized scale values for that alternative. Thus, for example, the weight on KNOW (.13) from Table 5 would be multiplied by the weight on JKT (.5), which in turn would be multiplied by the preference score of candidate O_1 on attribute JKT (.33) from Table 6. Having done similar calculations for each of the paths connecting the top of the tree with alternative O_j at the bottom, the analyst then simply adds these cross-products to generate the overall evaluation of alternative O_j. The overall values of the other alternatives are calculated in a similar way.

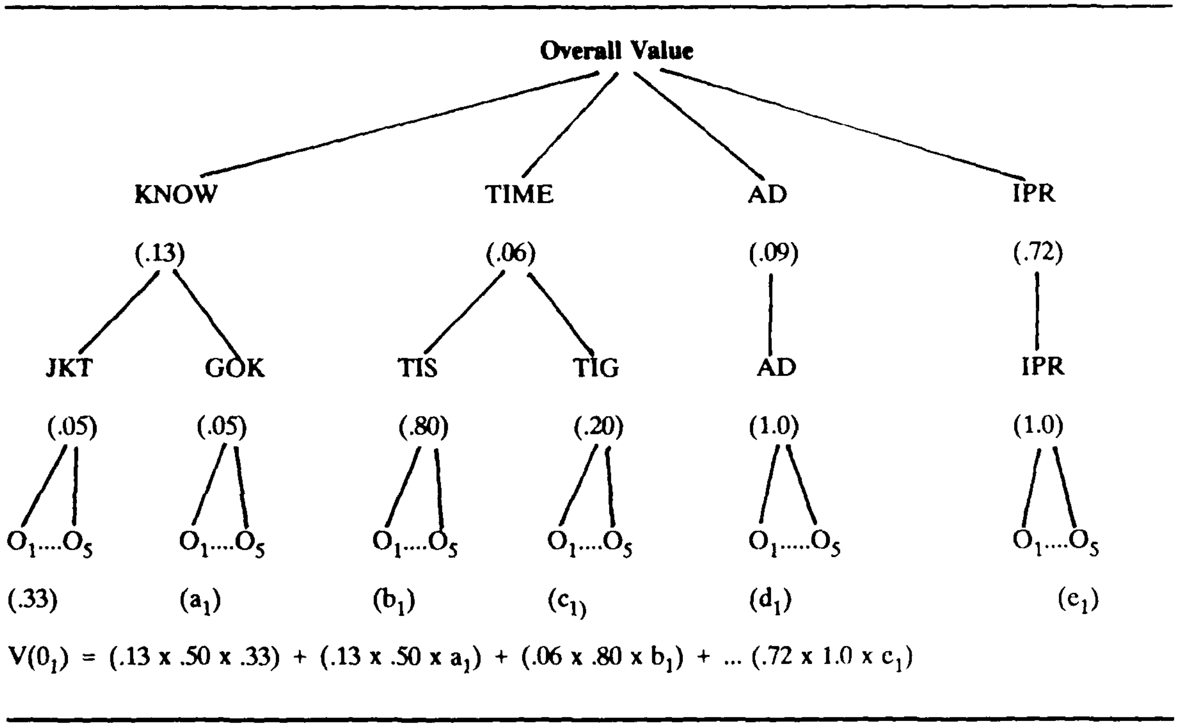

Figure 6. Illustration of the HAWM Aggregation Process.

Policy Capturing

The policy capturing approach discussed in this report utilizes ordinary least squares (OLS) regression to derive a linear (in the unknown parameters) model of the relationship between attributes and a criterion, from some number of holistic judgments of attribute profiles. The predictor forms, however, can be nonlinear. Although there are several recent variations, using experimental design considerations which seek to limit the number of judgments required (for example, Barron & Person, 1979), such techniques are not considered here. Neither are new model fitting approaches such as nonlinear regression parameter estimation or ridge regression. This more limited view is taken in order to remain consistent with the manner in which the technique is most often applied.

Elicitation of Holistic Judgments

In order to apply the policy capturing approach to the construction of a value model, the decision maker(s) is (are) presented with a number of "profiles," each of which describes an option in the decision problem. The description is complete with regard to the attributes thought to be relevant. The option is described in terms of an exact value for each continuously measured attribute and a category level for discrete attributes. The decision maker is asked to provide a judgment of the value for the option, either in naturally occurring units (dollars, for example), some arbitrary scale (for example, 0 to 100), or a rank order within the option set.

Selection of the profiles to be judged is one of the keys to proper utilization of the technique and takes several factors into account. First, the set of profiles should be comprehensive; that is, it should reflect the full range of each attribute. Second, no single area of the scale for a continuous attribute, nor single category for discrete variables, should be allowed to dominate the set. The set should be selected to be balanced across the range of each of the attributes. Third, the set of profiles should be reflective of the real world of the decision maker. The decision maker is likely to have difficulty making judgments about unrealistic profiles. Therefore, unlikely or impossible combinations of attributes should occur infrequently or not at all. The number of attributes is closely related to the decision maker's ability to process information. Generally, seven or fewer

attributes are the maximum number that are considered in applications of policy capturing. Finally, the set of profiles should be selected such that a variety of combinations of attributes are represented.

Clearly, some of these criteria for selection of an option set will be in conflict in many cases. For example, in resource allocation decisions cost is usually highly and positively correlated with some measure of time. It would thus be unrealistic to have a large number of low cost, long time options in the option set. Some options may be so unlikely as to appear ridiculous (e.g., a 1 year, full-time training course costing $1,000). In application, the analyst must balance the criteria against one another in order to arrive at a usable set of options for judgment.

There is no definitive guidance as to the number of options that need to be included in the option set. Since OLS regression is the method of analysis, the rule of thumb may apply that there should be a minimum of 10 cases (options) per variable (attribute) for which a weight parameter is to be estimated. There may also need to be additional cases when the stability of the parameter estimates is in question such as would be caused by a strong multicolinearity problem among the attributes.

In the context of the promotion example, a set of profiles, each representing a candidate for promotion, would be presented to the decision maker (or panel of decision makers) in a form similar to that shown in Figure 7. The decision maker would then be asked to respond with a score for each individual, representing the individual's relative "promotability." The score can be expressed on some arbitrary scale (for example, 0-100), in terms of the point scales used to measure each attribute, a rank ordering, or even categories (1 - 5 categories receiving 1 to 5 points each).

			Attributes			
Applicant	Job Knowledge Test Score	General Organizational Knowledge Score	Time In Service	Time In Grade	Awards and Decorations	Individual Performance Ratings
1.	86	62	47 months	18 months	Air Force Achievement Medal, Purple Heart	135
2.	77	81	106	8	Airman's Medal, Purple Heart (3 awards)	133

Figure 7. Example: Judgment Profile for Policy Capturing in Promotion Application.

The profiles would have been chosen to ensure that the entire set is both representative of the pool of applicants who would normally come before a promotion board and that it adequately covers the range of each attribute. The profile set would also be constructed to minimize the intercorrelations among attributes in the set. In addition, the set would be screened to eliminate unlikely or impossible applicants (for example, an applicant who has the Congressional Medal of Honor and three Purple Hearts, but only 1 month TIS).

Model Development

The value model is developed by performing a regression analysis on the judged profiles. The independent or predictor variables are the attributes, and the dependent variable is the decision maker's judgment of the promotability value of that profile. While almost any statistical package provides the means to analyze the

judgments in this manner, little or no software exists that facilitates other aspects of the use of this form of analysis for building value models. For example, problem structuring and sensitivity analysis are either not addressed in existing software, or are difficult and cumbersome to perform. Each requires multiple, sequential analysis, where a single attribute is removed or the profile set changed, and completely new analysis performed to evaluate impact.

On the other hand, the standard statistical packages, when applied in this context, provide several descriptive features that are unavailable to most other approaches. For example, the squared multiple correlation coefficient (R^2) provides an estimate of the completeness of the attribute set. In addition, most packages provide a significance test for the model but, given the independence assumptions of the sample statistic, it is usually unclear whether the significance can be taken as exact.

In a group situation, the individual decision maker's equations are clustered using a mathematical clustering routine, in order to arrive at a single equation of promotability. The regression equations are clustered using a clustering routine such as the Hier-Grp software (Ward, Treat, & Albert, 1985). The resulting clusters of equations are then examined to determine how many different rating patterns were evidenced by the rating panel. Aberrant raters can be removed so that the individual equations can be aggregated into a single equation, or feedback techniques can be used to eliminate group differences.

Policy Specifying

Unlike SMART and HAWM, which have seen extensive and widespread use, policy specifying is a decision analysis tool that has been used primarily by the United States Air Force within the personnel utilization decision context. The technique was developed in the mid-1970's (see Ward, 1977) by researchers at the Air Force Human Resources Laboratory to provide a preference modeling tool which explicitly considers the interaction of decision variables. Policy specifying provides the decision maker with a decision model within a hierarchical decision structure which does permit the decision maker's preference function to include variable interaction terms.

Hierarchical Structure

The decision maker, together with the assistance of a decision analyst, decides upon a decision objective and a set of decision variables (attributes). The variables themselves may be either quantitative or qualitative. The decision analysis process of policy specifying begins by having the decision maker decide which pairs of variables should logically interact. The decision maker then continues through the set of decision variables, forming interacting pairs when appropriate. Once all the logical pairs are formed, the decision maker moves up a level, in hierarchical fashion, to consider the interaction of decision variables with functional relationships (i.e., previously paired decision variables) or functional relationships with other functional relationships. Descriptive names are usually given to specific pairs at each level of the hierarchy until the overall decision objective is reached.

Using the promotion decision context as the example, the decision hierarchy as shown in Figure 8 could be derived from the problem. Although it may be seen that there are three pairs of decision variables producing three separate functions which are later combined into two additional functions, such an arrangement is somewhat arbitrary in this decision context. One might argue that some other order of combination is just as logical; perhaps, combining EXPERIENCE with POTENTIAL or IPR with JKT. At any rate, this tree will be used for the rest of the discussion of policy specifying.

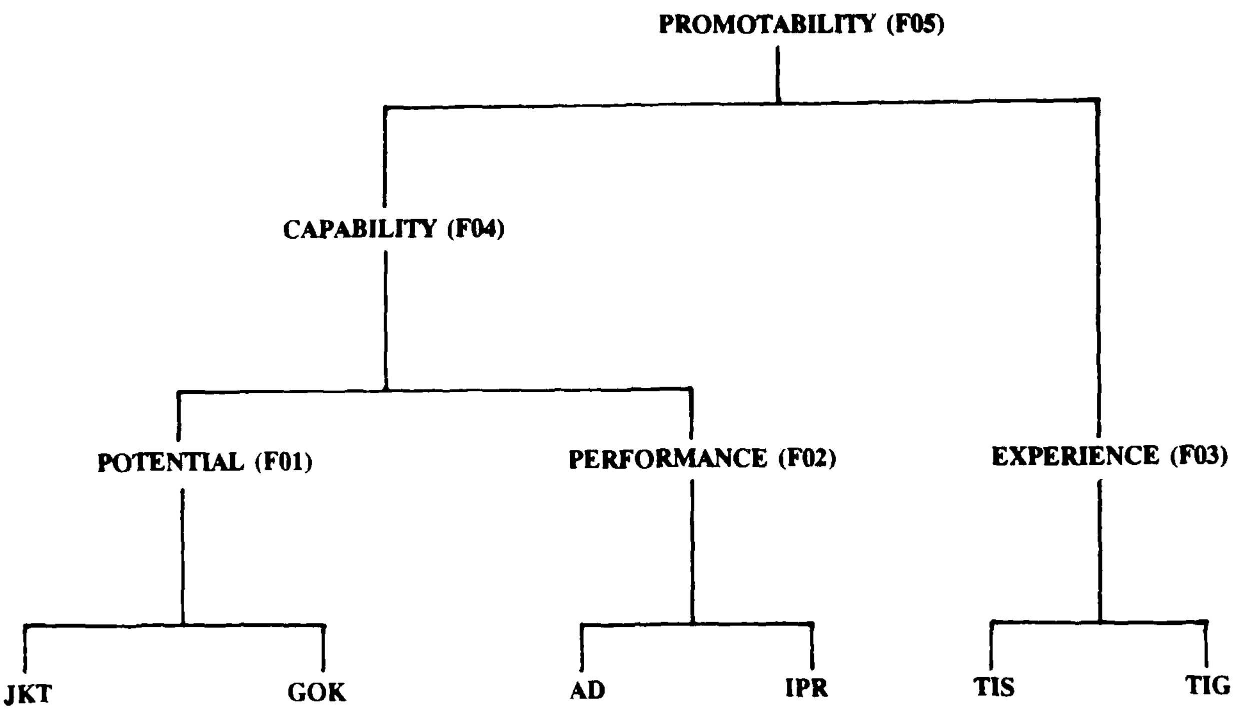

Figure 8. Promotion Policy Specifying Hierarchy.

Once the hierarchical decision tree is formed, the decision maker/decision analyst team must form the interactive functions for each of the pairs. In the current version of policy specifying, this is done by using the following general procedure:

Specifically let

b_j = the unknown weights to be determined by the policy specifying procedure

$j = 1......p$

p = number of terms in the function

b_0 = an unknown constant

X_1, X_2 = variables which are not vectors of data but are variables or combinations of variables which, when given a set of weights b_j and b_0 and a set of values for X_1 and X_2 will yield a composite value Y.

The general starting function is:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_1X_2 + b_4X_1^2 + b_5X_2^2 + ...b_pX^m{}_1X^n{}_2. \quad (1)$$

Prior to the policy specifying process, the range of possible values for X_1 and X_2 are known but the b_j and b_0 values are not known. Policy specifying proceeds by constructing the pairwise functional relationships between the decision variables which are the consequence of specifying Y values for stated X_1, X_2 combinations. These policy statements result in a set of equations (restrictions) in terms of b_j and b_0 so that the numerical values of the weights can be determined. Specification is completed when $p + 1$ independent restrictions are imposed. Once the values of b_j and b_0 are known, then predicted values Y can be calculated for any values of X_1 and X_2.

In order to simplify the choice from among a myriad of possible starting functional forms, the decision analyst has two starting models to work with (Ward, 1977). Each of these models attempts to capture the interaction between and among the decision variables. For this example, the following model will be selected as the starting model.

This model is defined in Ward (1977) as the following:

$$Y = b_0 + b_1X_1^a + b_2X_2^b + b_3X_1^aX_2^b \quad (2)$$

where:

b_0, b_1, b_2, b_3 are as defined previously,

Y is a composite measure of the strength (payoff) of the relationship between X_1 and X_2 (in this case, JKT and GOK).

The decision maker begins by deciding the extreme points of the relationship between JKT and Y (Payoff) or GOK and Y, at points where the other variable is at its worst or best. Figure 9 depicts what a decision maker might have decided about the JKT-GOK relationship. Note that the decision maker has also decided that the relationships are linear (a,b = 1) but that there is interaction between JKT and GOK since the payoff function for one of the variables changes depending on the level of the other variable. The decision maker has also scaled the decision variables and payoffs between 0 and 100.

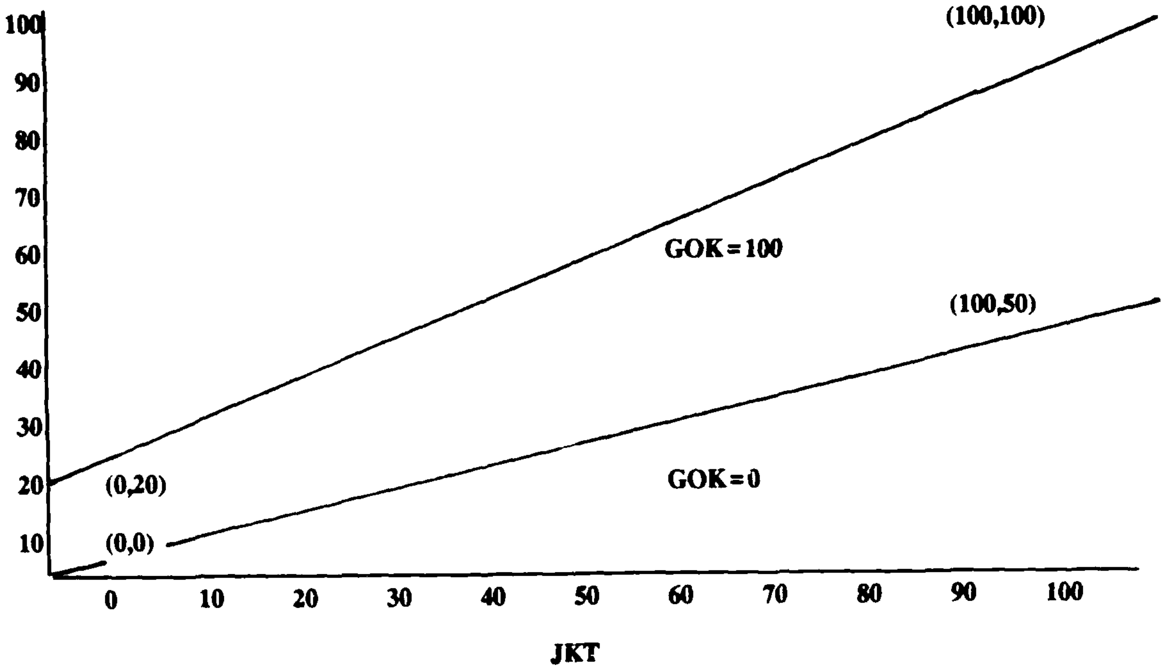

Figure 9. JKT-GOK Relationship

For this starting model, there are four unknown parameters to solve (b_0, b_1, b_2, b_3). Since the decision maker has decided that JKT and GOK are linearly related, a and b are set to 1.0. The decisions concerning the extreme points given the worst/best value of the other variable result in the payoff table in Figure 10.

		GOK 0	GOK 100
JKT	100	50	100
	0	0	20

Figure 10. JKT-GOK Worst/Best Payoff Table.

Using the promotion example translated into this starting model, Equation 2 becomes:

$$Y(JKT, GOK) = b_0 + b_1JKT + b_2GOK + b_3(JKT)(GOK). \quad (3)$$

Now there are four points (restrictions) for the payoff (Y) which are sufficient to allow the computation of the weights:

$$Y(0,0) = b_0 + b_1(0) + b_2(0) + b_3(0)(0) = 0$$

$$b_0 = 0$$

$$Y(100,0) = 0 + b_1(100) + b_2(0) + b_3(100)(0) = 50$$

$$b_1 = \frac{50}{100} = .5$$

$$Y(0,100) = 0 + b_1(0) + b_2(100) + b_3(0)(100) = 20$$

$$b_2 = \frac{20}{100} = .2$$

$$Y(100,100) = 0 + .5(100) + .2(100) + b_3(100)(100) = 100$$

$$b_3 = \frac{30}{10{,}000} = .003$$

Since each weight is now known, function F01 can be stated as:

$$F01 = .5JKT + .2GOK + .003\,(JKT)(GOK) \quad (4)$$

The process continues for each of the other pairs of decision variables or functions until function F05 (PROMOTABILITY) is formed. Fortunately, these computations are handled by the policy specifying software package.

At each stage of the model formulation, the decision maker is given feedback about what impact his/her interaction decisions have on the payoff at each point in the tree. For example, with function F01 (POTENTIAL), the decision maker might be given the payoff table in Table 7 as a result of applying the current functional relationship (Equation 4) for JKT and GOK. The decision maker could decide that the relationships depicted in the table do not adequately represent his/her policy and make changes in one or more of the parameters of the starting model and repeat the computation of F01.

Surface Fitting Procedure

An alternative to the use of specified starting models has also been developed. This alternative procedure allows the decision maker to supply various points in the payoff table. With a sufficient number of points, a surface is fit to the points and a function formed which represents the decision policy. Of course, this function has some statistical error of fit which should improve with more points. A variation on this approach is to allow the user to expand the model (number of terms and degree of interaction) and to evaluate the goodness of fit at each expansion. These alternative fitting procedures allow for more flexibility in function development.

Table 7. JKT-GOK Payoff Table

GOK

JKT	0	10	20	30	40	50	60	70	80	90	100
100	50	55	60	65	70	75	80	85	90	95	100
90	45	50	54	59	64	69	73	78	83	87	92
80	40	44	49	53	58	62	66	71	75	80	84
70	35	39	43	47	51	56	60	64	68	72	76
60	30	34	38	41	45	49	53	57	60	64	68
50	25	29	32	36	39	43	46	50	53	57	60
40	20	23	26	30	33	36	39	42	46	49	52
30	15	18	21	24	27	30	32	35	38	41	44
20	10	13	15	18	20	23	26	28	31	33	36
10	5	7	10	12	14	16	19	21	23	26	28
0	0	2	4	6	8	10	12	14	16	18	20

Evaluation Approach

The framework for assessing the methodologies included specification of evaluation attributes that were context dependent and a scale on which to score the techniques being evaluated. Attributes were designed to be generalizable; that is, they could be used to evaluate other techniques than those considered in this study. At the same time, the evaluation conducted for this study required attributes suitable for comparing all four techniques within a specific context. Thus, although specific comparisons between the four approaches to be evaluated in this study suggested some criteria, most criteria reflect concerns that are relevant to deciding in a particular context which technique to use.

A total of 15 attributes were defined and included in the technique scoring device and are described in detail in Figure 11. The attributes roughly divide into two categories: those related to scientific issues, and those related to empirical or elicitation issues.

Ability to model judgmental dependence. In some decision contexts, the problem attributes cannot be judged independently. It may be that there is an interaction among the attributes that must be considered, or that combinations of the attributes are more important than single attributes. This judgmental dependence can be modeled by techniques that allow the use of multiplicative, multilinear, or higher order polynomial relationships. In the most general case, some contexts require techniques that have the ability to model more complex forms of judgmental dependence among attributes.

Ability to model judgments holistically. In some decision contexts, the analyst uses a decision modeling technique to model and replicate the decision maker's decisions, rather than to study his decision process. In these contexts, the decision modeling technique is used to model the decision maker's decision holistically, since the decision maker is shown all the decision attributes for each decision option simultaneously, rather than sequentially, and is asked to evaluate each decision option separately.

Ability to aid understanding of the features of the problem, the decision maker's value structure, and the sensitivity of the problem to the modeling approach. For some decision problem areas, the analyst may need to adopt a role to help the decision maker understand more about the problem, rather than to provide the decision maker with a model that can be applied to solving the problem. In these cases, the analyst wants a decision modeling technique that will encourage detailed understanding of the substantive features of the problem. A technique that produces a detailed understanding of the decision maker's value structure, to aid the decision maker in this understanding, is also desired. The analyst might also require the technique to facilitate an understanding of how the decision problem might be sensitive to the modeling technique used and the attributes of the problem.

Ability to communicate the technical aspects of the problem, and the decision maker's preferences and decision logic. In some contexts, the analyst wants to use a technique that facilitates communication. The analyst might want to find a technique that provides easily communicable attribute scales and facilitates their understanding. The analyst might also want to use a method that requires explicit definition of the decision maker's values and preferences in an easily communicable form. Another area in which a technique might excel would be that it easily lends itself to displaying key problem components and their linkage to the decision recommendations.

Ability to use method with little or no training. This ability is concerned with how easy the method is to use. Some methods are very complex and difficult to understand, so that they are not usable by the non-expert. Others are more usable by the non-expert who has some experience or training. There are contexts in which the user has no experience using decision modeling techniques and therefore it would be important for the technique to be easily understood and require little experience or training for use.

Ability to develop a decision model that is theoretically defensible and scientifically well founded. This ability concerns how well researched the decision modeling technique is and how well it can be explained. In some contexts, this becomes important because the decision models will be scrutinized closely for accuracy and repeatability. Some decision modeling techniques have been well researched and have a great deal of empirical and psychophysical support in the published literature, while others have been less widely used and not much has been published about their use. Some techniques will have well-founded axiom systems, founded in utility measurement theory, and published and defended in well-known journals. Others will have been developed using an ad hoc approach without much thought given to utility measurement theory.

Ability to expand model to incorporate new problem information. This ability concerns the ease with which a decision model developed using a decision modeling technique can be changed when new problem information is added. In some contexts, the analyst uses a set of problem attributes and decision options to develop the decision model, and then wants to test how the introduction of new problem attributes would affect the decision model. Some techniques can handle the introduction of new information easily, and others require the process to be reaccomplished to incorporate new information.

Figure 11. Decision Attributes.

Ability to perform sensitivity analysis. After an analyst has developed a decision model, there arises the need in some decision contexts to perform a sensitivity analysis on the decision model. In a sensitivity analysis, the analyst will modify some part of the model to determine how sensitive the results of applying the decision model to a set of options are to changes in that part of the model. The analyst may wish to modify the weights on particular attributes or the form of the model or the way in which the value hierarchy was constructed, to determine if these changes impact the decision recommendations for a particular decision option set. Some decision modeling techniques will be more amenable to this form of analysis than will others.

Acceptability to the users of the final product, the decision model developed. In some contexts, the final product of the decision modeling exercise, the decision model, will be delivered to a customer who must implement and use it operationally to make decisions. In these contexts, several features of the decision modeling technique will affect how well the decision model is accepted. An acceptable technique will be essentially straightforward to use, and produce models that are easily understood. These techniques will produce a logical structure that is substantially acceptable, and will require judgments that are easily made by the decision maker.

Ability to develop decision model without using a computer. Some decision modeling techniques require extensive computer support, for problem structuring, for parameter elicitation, or for problem evaluation. Other techniques might be able to be used with or without a computer; still others might have to be used without a computer because no software exists that supports the methodological approach. In some decision contexts, computer resources are abundant and this will not be an issue; but in other contexts computer resources may be scarce or non-existent, and this will affect the choice of decision modeling technique.

Ability to model decision environment with many decision options. The number of decision options to be considered in developing a decision model will vary from one decision context to another. In some contexts, there will be only a few options to choose from, whereas in others the number of options may be large. Techniques will vary on their ability to handle different sizes of option sets. Some decision modeling techniques require many decision options in order to develop a reliable model (such as regression-based holistic modeling techniques) and might not produce reliable decision models with a small number of options. Others require pair-wise comparisons between all options, which becomes more tedious as the number of options increases.

Ability to apply decision model to new options. In some decision contexts, the decision model is developed using one set of decision options, and the analyst desires after the fact to apply this model to new options that were not considered in developing the decision model. Some techniques handle this relatively easily because they develop mathematically based models that take the form of an equation. This equation can be applied to a new decision option by simply plugging in the new attribute values and determining the results. Other techniques are not equipped for this type of application, because the model form is a hierarchy rather than an equation, or the technique requires all decision options to be known beforehand in order to develop a model.

Ability to develop model with little analyst involvement. In some contexts, there are no analysts available to aid the decision maker in using the decision modeling technique, or the only analysts available are too expensive to use. In these contexts, the decision maker may have to develop the decision model using no or little analyst involvement. Decision modeling techniques vary in their capability to be used without an analyst to aid the decision maker.

Ability to develop model with little decision maker involvement. In some contexts, the decision maker's time is limited or very expensive, and the analyst will have to develop the decision model with very little input from the decision maker. Techniques vary in their ability to be used by an analyst who has limited availability to a decision maker during the decision modeling process.

Ability to model a group decision making process. In some contexts, the decision maker is actually a group of people who must make the decision and therefore the group itself develops the decision model. In other contexts, either the group appoints one person to be spokesperson for the group, or there is a single decision maker who will be responsible for the decision problem, and this single person will develop the decision model. Some techniques have been explicitly developed to meet the problem of developing a group decision model, others can be adapted to meet this condition, and some techniques can be used only with a single decision maker.

Figure 11 (concluded)

These attributes were then used to develop the Technique Scoring Matrix shown in Figure 12. The scale used to judge the techniques is shown at the top of Figure 12, and ranges from "extremely capable" to "extremely incapable" of providing each of the 15 attributes. The judgment to be made in each case was whether the technique was able to provide the user with that capability and then to what degree it was able to do so.

Evaluation Results

The matrix was provided to 18 experts familiar with one or more of the four techniques. Each expert rated the technique they knew across each attribute and a single rating for each attribute for each technique was calculated using an average of all ratings. The agreement among experts was tested by calculating an interrater reliability for the expert's rating on each technique (see Christal & Weissmuller, 1976).

The interrater reliabilities (R_{kk}) ranged from .70 for the SMART ratings to .90 for the policy specifying ratings, corroborating the accuracy of using the averaging technique to smooth differences. The scores given for each of the techniques are shown in Table 8.

These scores reflect that SMART was considered to be best on 8 of the 15 attributes, whereas policy capturing scored highest on 3 of the attributes.

this technique is:	
8.	Extremely Capable
7.	Very Capable
6.	Moderately Capable
5.	Slightly Capable
4.	Slightly Incapable
3.	Moderately Incapable
2.	Very Incapable
1.	Extremely Incapable

Capable of modeling judgmental dependence	
Capable of modeling decisions holistically	
Capable of aiding understanding of the problem, the decision maker's value structure, and the sensitivity of the problem to the modeling approach	
Capable of communicating the technical aspects of the problem, and the decision maker's preferences and decision logic	
Capable of being used with little or no training	
Capable of developing a decision model that is theoretically defensible and well founded scientifically	
Capable of expanding model to incorporate new problem information	
Capable of performing sensitivity analysis	
Capable of producing a final product that is acceptable to the user	
Capable of being used without a computer	
Capable of modeling decision environment with many decision options	
Capable of being applied to a new decision option set	
Capable of being used to develop model with little analyst involvement	
Capable of being used to develop model with little decision maker involvement	
Capable of being used to model a group decision making process	

Figure 12. Technique Scoring Matrix.

Table 8. Technique Scoring Results

	SMART	HAWM	POLICY SPEC	POLICY CAPT
Ability to model judgmental dependencies	3.3	3.4	7.4	5.0
Ability to model judgments holistically	1.0	3.0	3.6	7.8
Ability to aid understanding	7.3	5.8	6.6	5.6
Ability to communicate the decision logic	7.0	5.4	5.8	5.6
Ability to use technique with no training	7.3	6.0	3.9	4.8
Ability to develop theoretically defensible model	7.3	5.6	4.0	6.7
Ability to expand model with new information	4.7	2.8	6.4	4.1
Ability to perform sensitivity analysis	6.0	4.0	5.8	5.9
Acceptability of final product	7.3	5.8	6.9	6.8
Ability to model decisions without a computer	6.0	3.4	1.8	2.2
Ability to model decision with many decision options	5.7	1.6	7.0	6.8
Ability to apply decision model to new option set	6.0	1.4	6.8	7.0
Ability to model w/ little analyst involvement	4.7	5.6	3.1	3.7
Ability to model with little DM involvement	3.7	3.0	3.4	3.3
Ability to model group process	4.7	3.8	5.1	7.6

In contrast, policy specifying was ranked first on only 3 attributes but last on 4 attributes. HAWM was ranked first only once and was last on 8 attributes.

SMART Evaluation

The strengths of SMART are reflected in the 8 attributes on which it ranked first: ability to aid understanding, ability to communicate the decision logic, ability to use with no training, ability to develop a theoretically defensible model, ability to perform sensitivity analysis, acceptability of the final product, ability to model decisions without a computer, and ability to model with little decision maker involvement. It is a fairly straightforward technique, easy to use and easy to understand. The judgments made in the hierarchy are naturally made and readily communicated. Decision makers believe the results and trust in its use, primarily because of the lack of complexity in the method. SMART has a computer implementation that works well and does not require much involvement on the part of a decision analyst. Even with its lack of complexity, SMART is well founded axiomatically, with the difference judgments that are required based on axioms from measurement theory.

The categories in which SMART scored lowest pinpoint its weaknesses: its inability to model judgmental dependencies and its inability to model judgments holistically. SMART has no built-in procedure for handling functional dependencies or correlations. In addition, all models in SMART must be additive and linear in all attributes, so that judgmental dependencies cannot be addressed. It is a hierarchial procedure by design, and has no capability for modeling judgments holistically.

Policy Capturing Evaluation

Policy capturing has some strengths also, as evidenced by the categories in which it rated highest: ability to model judgments holistically, ability to apply model to new decision option set, and ability to model group processes. Policy capturing is strong in these areas primarily because it is a descriptive technique based on holistic judgments, and was developed to be used in a group setting. Policy capturing also has the advantage of being considered in this study as a class of techniques which have been in the literature for over 25 years, and have been implemented in many different forms. The other three techniques are primarily prescriptive techniques, based on decomposition techniques developed to be used by single decision makers. The other techniques have also been judged in this report as single implementations of a body of literature, and therefore cannot take advantage of advancements or variations which have been described by others but not implemented in these particular software packages.

Because of its descriptive nature and the fact that the final output is a policy equation, policy capturing rates highest in its ability to be applied to a new decision option set. It naturally also has the most empirical literature published, because of its age and the fact that it is a class of techniques. It scores well on the group attribute since it was developed to systematically handle group decision making situations.

Policy capturing has some weaknesses, even though it was rated lowest on only one of the attributes: ability to aid understanding. There is no one computer implementation of the class of techniques known as policy capturing. Most applications of the technique use hard copy records to elicit judgments, paper and pencil to record them, and a statistical software package to determine the regression equations. Some implementations have automated portions of this procedure, but only on mainframes which lack the flexibility of traveling to where the user is located. Policy capturing, by its design and use, has no facility to incorporate a hierarchy, and because of its descriptive nature, offers little insight into decision problems, or the rationale for a particular decision, because the only outputs available for examination are the regression weights. The technique also has the weakness of involving many resources. A great deal of time and resources are required before and after a policy capturing session, in setting up judgment samples, and in analyzing the results.

HAWM Evaluation

HAWM was ranked first on only one attribute: ability to model decision with little analyst involvement. On the negative side, HAWM was ranked last in nine categories: ability to aid understanding, ability to communicate decision logic, ability to expand with new information, ability to perform sensitivity analysis,

acceptability of final product, ability to model decision with many options, ability to apply decision model to a new option set, ability to model with little decision maker involvement, and ability to model a group process. HAWM was rated last in ability to model decisions with many options and ability to be applied to a new option set, for very nearly the same reason; namely, the technique is extremely sensitive to the introduction of new information. The addition of new alternatives can upset the rational prescriptions of the technique primarily because the alternatives in HAWM are introduced as part of the hierarchy. The introduction of new alternatives can cause rank reversals for alternatives that had been previously judged. The introduction of any new information requires that the judgment process be reaccomplished since every member of the hierarchy must be pairwise compared to every other member of the next level in the hierarchy. HAWM ranked last in the attribute of modeling a group process and low in the ability to model judgmental dependencies for nearly the same reason. This implementation of the technique was not equipped to handle a group decision process, and could not handle any functional forms except linear additive models. However, the literature discusses versions of Saaty's AHP process which theoretically could handle non-additive and nonlinear models of the decision process (Harker & Vargas, 1985; Saaty & Takizawa, 1985).

Policy Specifying Evaluation

Policy specifying was ranked highest on three attributes: ability to model judgmental dependencies, ability to expand model with new information, and ability to model decisions with many decision options. The ranking of policy specifying as highest on the judgmental dependency attribute is significant. Policy specifying was developed to meet a perceived void in the decision modeling field--the inability to handle judgmental dependencies (sometimes called interactions). Policy specifying was the only technique of the four studied which allowed for this ability to specify more complex models through the introduction of curvilinearity and interaction. Because it is built using a very flexible pairwise hierarchy, policy specifying was also ranked first on the ability to expand with new information, and the ability to model a decision with many decision options.

Policy specifying was ranked lowest on four attributes: ability to use technique with no training, ability to develop a theoretically defensible model, ability to use without a computer, and ability to model with little analyst involvement. Because it is a technique used almost exclusively by the Air Force Human Resources Laboratory, it has the smallest body of published literature of the four techniques. Like the other two prescriptive techniques, it suffers from the lack of features to handle functional dependencies or group decision processes. However, unlike the other two techniques, policy specifying in this implementation is a very complex procedure that is difficult to communicate to users. Its low ranking in terms of analyst involvement reflects the amount of time required by the decision analyst and the decision maker to arrive at a pairwise hierarchy and a pairwise function for each element in the hierarchy.

IV. EVALUATION OF CONTEXT/TECHNIQUE MATCH

In this section discussion will focus on the evaluation of these four techniques as applied to three Air Force decision contexts. The contexts studied were: an Air Force promotion board, an enlisted person-job-match system, and the prioritization of R&D projects. Before a discussion of the evaluation methods is carried out, two of the three decision contexts will be discussed in more detail. The Air Force promotion board decision context has already been discussed in detail as the common example in Section III.

Decision Contexts

In the Air Force enlisted person-job match (PJM) decision context, the Air Force is faced with making a classification decision about each individual entering the Air Force. The classification decision is two-tiered in that the Air Force first decides whether or not to allow the individual to enlist, and then participates with the individual in deciding what job he/she should fill in the Air Force. The nature of this decision context is that it is a sequential process: as one job is filled another person with unknown characteristics arrives requesting a job. Thus the decision options (each person-job match) are unknown and many, as the decision maker is formulating a decision model. There are six decision attributes used in this context: Individual Preference (IP), Fraction of jobs Filled (FF), Technical School predicted Success (TSS), Time Remaining to fill job (TR), Job

Requirement (JR), and Applicant Aptitude (AA). These attributes are divided into two categories: job characteristics, and person characteristics. Among the person characteristics are the IP and AA attributes. The IP attribute is collected from the individual using a crude measuring instrument which allows the individual to express interest in the four aptitude areas on a scale from 0 to 10. The four aptitude areas in this context are: Mechanical (M), Administrative (A), General (G), and Electronics (E). The AA attribute is the person's scores in each of the M, A, G, and E aptitude areas of the Armed Services Vocational Aptitude Battery (ASVAB). The job characteristics are the JR, FF, and TR attributes. JR is the aptitude requirement of each job, as reflected by the minimum score required on the ASVAB in order to qualify for each particular job. Each job will have its own requirement score in one of the four aptitude areas, and in many cases will have a combination of scores required. The FF attribute reflects the number of jobs of any one kind that are presently available divided by the number initially available during a specific period of time, indicating the fraction of jobs currently filled. The TR attribute is related to the FF attribute in that TR reflects the length of time remaining to fill the particular job. Of course, the intention of this variable is to reflect the impact on the job assignment system of leaving a job empty. The TSS attribute is both a person and job attribute, since it is the result of predicting a person's success in a technical school using demographic data about the individual. A regression equation is developed using actual technical school results for each technical school as the dependent variable. For more information on this decision context, the reader is referred to Ward, Haney, Hendrix, and Pina (1977).

In the R&D project context, the Air Force is faced with making a decision about how R&D funds should be programmed over a fiscal year. There are more projects requiring funding than funds available to pay for them, and a scheme is necessary for determining how to best allocate the funds. In this decision context, the options are few and known before the decision has to be made, and the decision is typically made by a team of experts within the research organization. Many different formulations have been tried for this problem, but here the reference is to the formulation by DeWispelare (1983).

In this context, there are four attributes to be considered in making the decision: Technology Base (TB), Sponsorship Potential (SP), Cost (C), and Time to Project Fruition (TPF). The TB attribute is measured using a 10-point scale that reflects how much of the R&D effort has already been expended on any particular project area, with a Delphi procedure used to achieve consensus among the judges. The SP attribute is an estimate of the number of potential sponsors for the project, as determined through survey or telephone interview. The attribute C reflects an estimate of the cost to complete the project within a 5-year planning horizon. The attribute TPF is an estimate of the actual time required to achieve the research objective. This attribute is estimated using expert opinion and past history.

Evaluation Approach

The framework for evaluating the methodologies in specific contexts starts with the 15 attributes that were developed for the technique evaluation in Section III of this report (shown in Figure 12). A similar evaluation tool was developed for each context, using the same 15 attributes to form the rating form shown in Figure 13. The eight-point scale on this form reflects the need for the technique used to have the 15 abilities shown on the form. The rater first decides on whether or not the ability is important to the decision context and then to what degree it is or is not important. A second group of 13 experts were sent these forms (a group of experts distinct from those who evaluated the techniques in Section III). These judges were asked to score the need for these abilities within a context with which they were familiar. The results were collected and then averaged to form an average context need for each ability. As with the technique ratings, the interreliability for each set of raters' ratings was calculated. These reliabilities ranged from .75 for the promotion board context to .85 for the PJM context, corroborating the accuracy of using average ratings for the group of experts. These scores were then matched with the technique scores from the first group of experts, and a payoff (utility) table was generated using the policy specifying technique. This table is shown as Table 9, and reflects the utility, to the context, of each combination of context need and technique capability. The payoff table generated here reflects the opinion of the experts that the technique that meets or exceeds the needs of the context will always receive the maximum payoff of 100. As soon as a technique falls below the minimum needs of a context, the payoff is reduced substantially until the curve flattens out at the lower scores.

In this context this ability is:	8. Extremely Important 7. Very Important 6. Moderately Important 5. Slightly Important 4. Slightly Unimportant 3. Moderately Unimportant 2. Very Unimportant 1. Extremely Unimportant

Ability to model judgmental dependence	
Ability to model judgments holistically	
Ability to aid understanding of the features of the problem, the decision maker's value structure, and the sensitivity of the problem to the the modeling approach	
Ability to communicate the technical aspects of the problem, and the decision maker's preferences and decision logic	
Ability to use method with little or no training	
Ability to develop a decison model that is theoretically defensible and scientifically well founded	
Ability to expand model to incorporate new problem information	
Ability to perform sensitivity analysis	
Acceptability to the users of the final product, the decision model developed	
Ability to develop decision model without using a computer	
Ability to model decision environment with many decision options	
Ability to apply decision model developed to a new decision option set	
Ability to develop model with little analyst involvement	
Ability to develop model with little decision maker involvement	
Ability to model a group decision making process	

Figure 13. Context Scoring Matrix.

Table 9. Payoff of Context/Technique Match

Context Need		Technique Capability 1	2	3	4	5	6	7	8
	1	100	100	100	100	100	100	100	100
	2	64	100	100	100	100	100	100	100
	3	53	62	100	100	100	100	100	100
	4	42	52	62	100	100	100	100	100
	5	32	42	52	63	100	100	100	100
	6	21	32	43	54	65	100	100	100
	7	11	22	34	45	57	69	100	100
	8	0	12	24	37	49	61	73	100

A hierarchical set of evaluation criteria was developed, based on the same 15 abilities measured in the context and technique scoring matrices. The purpose of the hierarchy was to allow the decision maker to give relative weights to the 15 abilities for each context. This hierarchy is shown in Figure 14.

The structure was designed to be comprehensive while ensuring independence between criteria whenever possible. There are two main branches in the structure representing scientific issues versus issues that reflect empirical and/or applied considerations. The scientific branch focuses primarily on validity and versatility issues, examining the quality of a technique's axiomatic foundation, understandability of methods and results, the technique's ability to model complex value structures, and the technique's complexity and/or difficulty of understanding. Empirical considerations are broken into issues such as acceptability to users, and issues involved with model elicitation, including a technique's usefulness in group decision making, and its requirements for resources and decision maker involvement.

For each of the "twigs" shown at the far right of the value structure--labeled context/technique match--the same group of experts who scored the contexts were asked to use the SMART procedure with the swing weighting methodology described earlier to generate the relative weights for each alternative technique context match. Weighting was accomplished using the Context Weighting Matrix, shown in Figure 15. First the expert assigned a rank of 1 to the most important context/technique match and 2 to the second most important, until the eight twigs of the scientific portion of the hierarchy were rated. Once all eight were ranked, a percentage of 100 was assigned to the highest ranked match; then relative weights were assigned to the other matches, reflecting how much less important they were than the top ranked match. These weights were then normalized for use in the SMART procedure.

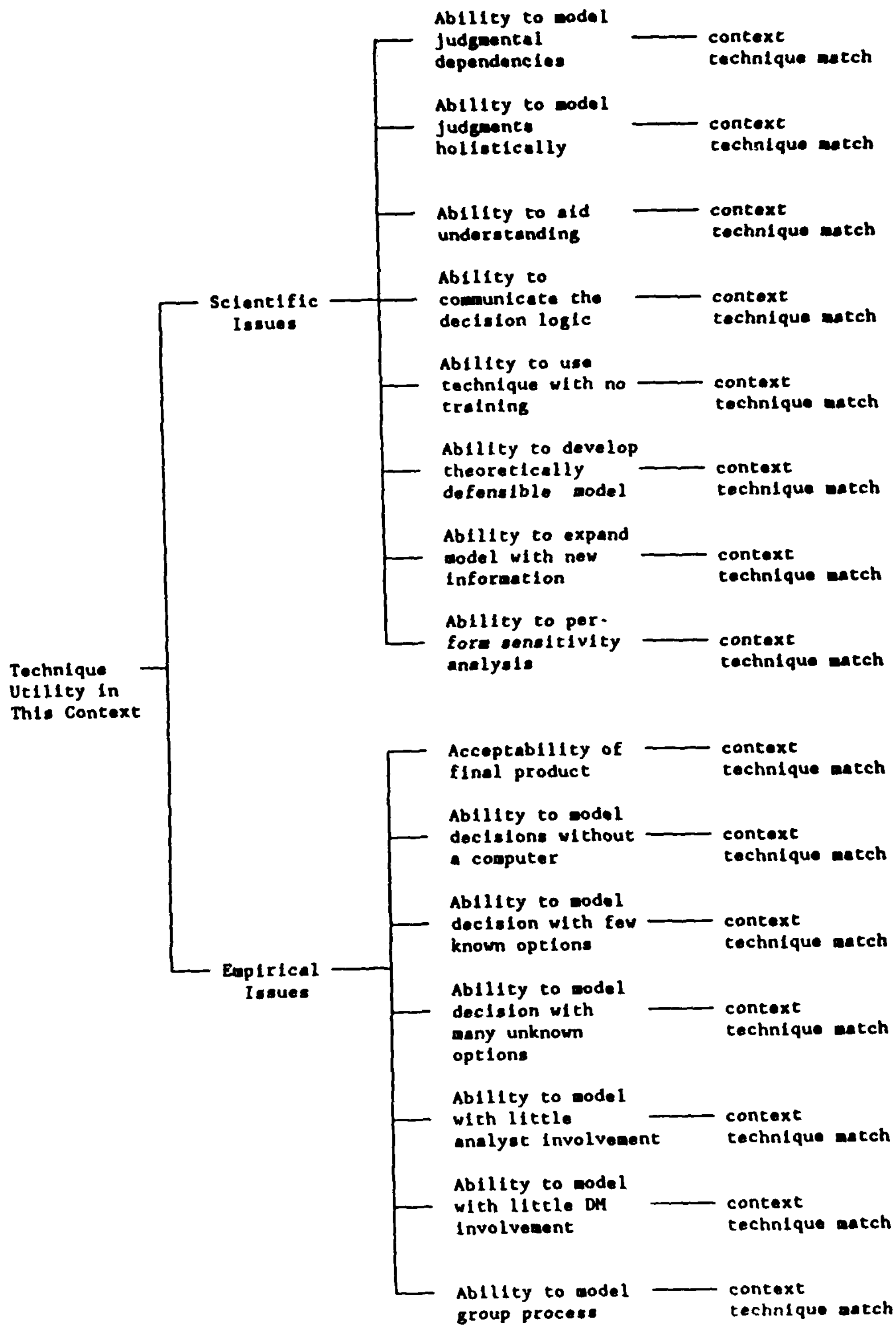

Figure 14. Hierarchical Structure.

CONTEXT

	Worst	Best	Rank	Weight
Ability to model judgmental dependencies	Poor context technique match	Good context technique match		
Ability to model judgments holistically	Poor context technique match	Good context technique match		
Ability to aid understanding	Poor context technique match	Good context technique match		
Ability to communicate the decision logic	Poor context technique match	Good context technique match		
Ability to use technique with no training	Poor context technique match	Good context technique match		
Ability to develop theoretically defensible model	Poor context technique match	Good context technique match		
Ability to expand model with new information	Poor context technique match	Good context technique match		
Ability to perform sensitivity analysis	Poor context technique match	Good context technique match		
Acceptability of final product	Poor context technique match	Good context technique match		
Ability to model decisions without a computer	Poor context technique match	Good context technique match		
Ability to model decision with many options	Poor context technique match	Good context technique match		
Ability to apply decision model to new option set	Poor context technique match	Good context technique match		
Ability to model w/ little analyst involvement	Poor context technique match	Good context technique match		
Ability to model with little DM involvement	Poor context technique match	Good context technique match		
Ability to model group process	Poor context technique match	Good context technique match		

Figure 15. Context Weighting Matrix.

V. CONTEXT EVALUATION RESULTS

Person-Job-Match Context

The results from employing the context scoring matrix for the PJM context are shown in Table 10. This table shows the average rating for each ability given by the group of experts that participated in the study. Within this context, the experts believed that the three most important abilities a technique should have were acceptability of the final product to the users, ability to model in a decision environment with many decision options, and ability to apply the decision model to a new decision option set. These last two abilities reflect the sequential nature of the PJM decision context. In the analysis process, the average scores were then combined with the average scores for the techniques previously collected, using the payoff matrix of Table 9. The resulting payoffs are shown in Table 11. These payoffs were then used as the scores for the matches in the SMART procedure.

This group of experts also gave relative weights to the 15 abilities using the swing weighting method described in Section III. The average relative weights used in the utility determination are shown in Table 12. These weights were used as the twig weights in SMART, with the assumption that the scientific and empirical sides of the hierarchy would receive equal weight in the utility determination. Multiplying by the payoff scores determined above resulted in the relative utilities for the four techniques shown in Table 13.

Table 13 reflects the final results of the utility determination procedure. Policy capturing was ranked highest in the PJM context, with policy specifying a close second choice. In the actual decision context, the Air Force has used a mixture of policy specifying and a utility weighting procedure similar to SMART, but combinations of techniques were not studied in this evaluation.

Table 10. PJM Context

In this context this ability is:	8.	Extremely Important
	7.	Very Important
	6.	Moderately Important
	5.	Slightly Important
	4.	Slightly Unimportant
	3.	Moderately Unimportant
	2.	Very Unimportant
	1.	Extremely Unimportant

Ability to model judgmental dependence	6.4
Ability to model judgments holistically	4.7
Ability to aid understanding of the features of the problem, the decision maker's value structure, and the sensitivity of the problem to the modeling approach	5.9
Ability to communicate the technical aspects of the problem, and the decision maker's preferences and decision logic	6.4
Ability to use method with little or no training	5.4
Ability to develop a decison model that is theoretically defensible and scientifically well founded	5.4
Ability to expand model to incorporate new problem information	6.4
Ability to perform sensitivity analysis	5.3
Acceptability to the users of the final product, the decision model developed	7.9
Ability to develop decision model without using a computer	2.9
Ability to model decision environment with many decision options	7.4
Ability to apply decision model to new decision option set	7.3
Ability to develop model with little analyst involvement	3.7
Ability to develop model with little decision maker involvement	5.1
Ability to model a group decision making process	6.7

Table 11. Payoffs for PJM Context

PJM	POLICY SPECIFYING	POLICY CAPTURING	SMART	HAWM
Ability to model judgmental dependencies	100	61	42	43
Ability to model judgments holistically	61	100	35	55
Ability to aid understanding	100	72	100	74
Ability to communicate the decision logic	70	68	100	65
Ability to use technique with no training	58	67	100	100
Ability to develop theoretically defensible model	53	100	100	100
Ability to expand model with new information	100	51	58	37
Ability to perform sensitivity analysis	100	100	100	60
Acceptability of final product	71	70	76	58
Ability to model decisions without a computer	61	65	100	100
Ability to model decision with many options	76	74	61	13
Ability to apply decision model to new decision set	74	77	65	12
Ability to model with little analyst involvment	65	100	100	100
Ability to model with little DM involvement	55	54	58	51
Ability to model group process	60	100	55	45

Table 12. PJM Context Weights

Ability to model judgmental dependence	.099
Ability to model judgments holistically	.059
Ability to aid understanding of the features of the problem, the decision maker's value structure, and the sensitivity of the problem to the modeling approach	.054
Ability to communicate the technical aspects of the problem, and the decision maker's preferences and decision logic	.061
Ability to use method with little or no training	.039
Ability to develop a decison model that is theoretically defensible and scientifically well founded	.051
Ability to expand model to incorporate new problem information	.077
Ability to perform sensitivity analysis	.053
Acceptability to the users of the final product, the decision model developed	.105
Ability to develop decision model without using a computer	.032
Ability to model decision environment with many decision options	.088
Ability to apply decision model to new decision option set	.095
Ability to develop model with little analyst involvement	.048
Ability to develop model with little decision maker involvement	.053
Ability to model a group decision making process	.075

Table 13. Overall Utilities: PJM Context

POLICY CAPTURING	76.2
POLICY SPECIFYING	75.9
SMART	72.1
HAWM	53.7

Promotion Board Context

The same procedure was then applied to the promotion board context. Table 14 contains the average expert opinions as to the importance of the 15 attributes in this context. The experts' ratings show that the highest rated attributes for this context were the ability to model judges holistically, the acceptability of the final product to the user, the ability to model an environment in which there are many decision options, the ability to apply the model to a new option set, and the ability to model a group process. Even though PJM and this context have two attributes rated highest in common, the total ratings are quite different from the ratings given in the PJM context. Based on this, it would be expected that the rank ordering of the four techniques should be different.

These ratings resulted in the payoffs shown in Table 15, when combined with the technique ratings. These payoffs were then used in the SMART procedure using the context weights shown in Table 16. These weights, when multiplied by the payoffs, resulted in the relative utilities for the four techniques shown in Table 17.

Table 17 reflects the final result of the utility ranking procedure. Policy capturing procedures were found to be most appropriate for the Air Force promotion system context. In the actual decision context, the Air Force has used the policy capturing technique for promotion board work, thus providing some support for the utility procedure used in this analysis.

Table 14. Promotion Board Context

In this context this ability is:	8.	Extremely Important
	7.	Very Important
	6.	Moderately Important
	5.	Slightly Important
	4.	Slightly Unimportant
	3.	Moderately Unimportant
	2.	Very Unimportant
	1.	Extremely Unimportant

Ability to model judgmental dependence	4.2
Ability to model judgments holistically	7.4
Ability to aid understanding of the features of the problem, the decision maker's value structure, and the sensitivity of the problem to the modeling approach	6.0
Ability to communicate the technical aspects of the problem, and the decision maker's preferences and decision logic	5.6
Ability to use method with little or no training	5.2
Ability to develop a decison model that is theoretically defensible and scientifically well founded	6.0
Ability to expand model to incorporate new problem information	5.0
Ability to perform sensitivity analysis	4.8
Acceptability to the users of the final product, the decision model developed	7.6
Ability to develop decision model without using a computer	2.8
Ability to model decision environment with many decision options	7.2
Ability to apply decision model to new decision option set	7.4
Ability to develop model with little analyst involvement	4.0
Ability to develop model with little decision maker involvement	5.0
Ability to model a group decision making process	7.6

Table 15. Payoffs for Promotion Board Context

PROMOTION BOARD	POLICY SPECIFYING	POLICY CAPTURING	SMART	HAWM
Ability to model judgmental dependencies	100	100	63	64
Ability to model judgments holistically	37	100	6	30
Ability to aid understanding	100	71	100	73
Ability to communicate the decision logic	100	100	100	72
Ability to use technique with no training	59	69	100	100
Ability to develop theoretically defensible model	54	100	100	71
Ability to expand model with new information	100	63	69	50
Ability to perform sensitivity analysis	100	100	100	64
Acceptability of final product	73	72	78	61
Ability to model decisions without a computer	62	66	100	100
Ability to model decision with many options	78	75	62	15
Ability to apply decision model to new option set	73	76	64	11
Ability to model w/ little analyst involvement	63	68	100	100
Ability to model with little DM involvement	56	55	59	52
Ability to model group process	53	100	47	37

Table 16. Promotion Board Context Weights

Ability to model judgmental dependence	.044
Ability to model judgments holistically	.079
Ability to aid understanding of the features of the problem, the decision maker's value structure, and the sensitivity of the problem to the modeling approach	.069
Ability to communicate the technical aspects of the problem, and the decision maker's preferences and decision logic	.069
Ability to use method with little or no training	.050
Ability to develop a decison model that is theoretically defensible and scientifically well founded	.068
Ability to expand model to incorporate new problem information	.056
Ability to perform sensitivity analysis	.062
Acceptability to the users of the final product, the decision model developed	.094
Ability to develop decision model without using a computer	.036
Ability to model decision environment with many decision options	.077
Ability to apply decision model to new decision option set	.089
Ability to develop model with little analyst involvement	.057
Ability to develop model with little decision maker involvement	.055
Ability to model a group decision making process	.087

Table 17. Overall Utilities: Promotion Board Context

POLICY CAPTURING	80.6
SMART	74.0
POLICY SPECIFYING	73.0
HAWM	55.5

Research and Development Project Context

The last context studied was the R&D project context, using the same procedures as were used on the other two contexts. Table 18 contains the average expert opinions as to the importance of the 15 attributes in this context. The two attributes rated highest by the experts as being most important to this context were: ability to aid understanding and acceptability of the final product. This combination of attributes is different from the two decision contexts previously studied and should have resulted in a different rank ordering of the four techniques. These ratings resulted in the payoffs shown in Table 19, when combined with the technique ratings. The experts also rated the relative weights for the 15 attributes in this context, resulting in the twig weights shown in Table 20. Multiplying by the payoffs gave the relative utilities shown in Table 21.

Table 21 shows the final results of the assessment. The SMART procedure was found to be most appropriate for use in the R&D project context. In the actual decision context, the Air Force uses many different procedures since this prioritization occurs at every R&D organization within the Air Force. The AFHRL has used policy specifying with some success, and DeWispelare (1983) has suggested a more intricate form of multiattribute utility function development than the SMART procedure studied in this report.

Table 18. R&D Project Context

In this context this ability is:	
	8. Extremely Important
	7. Very Important
	6. Moderately Important
	5. Slightly Important
	4. Slightly Unimportant
	3. Moderately Unimportant
	2. Very Unimportant
	1. Extremely Unimportant

Ability to model judgmental dependence	6.6
Ability to model judgments holistically	3.0
Ability to aid understanding of the features of the problem, the decision maker's value structure, and the sensitivity of the problem to the modeling approach	7.0
Ability to communicate the technical aspects of the problem, and the decision maker's preferences and decision logic	6.6
Ability to use method with little or no training	5.0
Ability to develop a decison model that is theoretically defensible and scientifically well founded	6.4
Ability to expand model to incorporate new problem information	6.0
Ability to perform sensitivity analysis	6.6
Acceptability to the users of the final product, the decision model developed	7.0
Ability to develop decision model without using a computer	2.8
Ability to model decision environment with many decision options	5.0
Ability to apply decision model to new decision option set	6.4
Ability to develop model with little analyst involvement	3.2
Ability to develop model with little decision maker involvement	3.2
Ability to model a group decision making process	6.2

Table 19. Payoffs for R&D Project Context

R&D PROJECT	POLICY SPECIFYING	POLICY CAPTURING	SMART	MAUM
Ability to model judgmental dependencies	100	59	41	42
Ability to model judgments holistically	100	100	53	100
Ability to aid understanding	75	63	100	65
Ability to communicate the decision logic	68	66	100	64
Ability to use technique with no training	61	70	100	100
Ability to develop theoretically defensible model	50	100	100	68
Ability to expand model with new information	100	55	61	41
Ability to perform sensitivity analysis	68	69	71	48
Acceptability of final product	78	77	100	65
Ability to model decisions without a computer	62	66	100	100
Ability to model decision with many options	100	100	100	38
Ability to apply decision model to new option set	100	100	72	21
Ability to model w/ little analyst involvement	70	100	100	100
Ability to model with little DM involvement	100	100	100	69
Ability to model group process	64	100	59	50

Table 20. R&D Project Context Weights

Ability to model judgmental dependence	.072
Ability to model judgments holistically	.041
Ability to aid understanding of the features of the problem, the decision maker's value structure, and the sensitivity of the problem to the the modeling approach	.080
Ability to communicate the technical aspects of the problem, and the decision maker's preferences and decision logic	.069
Ability to use method with little or no training	.041
Ability to develop a decision model that is theoretically defensible and scientifically well founded	.068
Ability to expand model to incorporate new problem information	.059
Ability to perform sensitivity analysis	.069
Acceptability to the users of the final product, the decision model developed	.104
Ability to develop decision model without using a computer	.035
Ability to model decision environment with many decision options	.082
Ability to apply decision model to new decision option set	.086
Ability to develop model with little analyst involvement	.055
Ability to develop model with little decision maker involvement	.057
Ability to model a group decision making process	.081

Table 21. Overall Utilities: R&D Project Context

SMART	83.7
POLICY CAPTURING	82.0
POLICY SPECIFYING	80.3
HAWM	59.9

VI. IMPLICATIONS FOR POTENTIAL IMPROVEMENTS TO DECISION MODELING TECHNIQUES

In this section suggestions are made for extensions and/or modifications to the policy specifying and policy capturing techniques and are directed toward improving the value of these techniques in terms of the evaluation criteria used in the present effort. Results of the analysis clearly show that the group of experts believed that a policy specifying capability serves a purpose not fulfilled by the other three techniques, but that the current implementation of the technique could be improved. Policy capturing as a tool also serves a purpose not met by the other three techniques, but the current implementation should be modified and enhanced.

Improvements to Policy Specifying

Improvement of policy specifying as a decision modeling technique should be directed toward eliminating problems of user acceptance and understanding underscored in the evaluation. It is suggested that a new version of policy specifying be developed which uses a SMART-like approach to define ordinary single-attribute utility functions and to aggregate the separate utility functions. The ability to specify interaction terms in the utility functions should be retained when the decision maker feels the context warrants it. This would be done by allowing the analyst or decision maker to access through the software package a special modeling component which specifies the nature of the relationships among attributes. This modeling component should allow for a form of curve fitting, which would likely be more acceptable and easier to communicate than is the current practice of specifying corner points on a linear equation template.

The policy specifying tool should also incorporate two additional major features. The first would be some form of consistency determination or comparison of model results to a desired set of results, similar in a sense to the R^2 measure in policy capturing. More importantly, the second is the capability for using the tool in a group decision making mode. Such a facility would have to be capable of analyzing parallel policies, rather than merging the policies.

Improvements to Policy Capturing

The suggested improvements to the policy capturing implementation are directed toward strengthening the weaknesses described in Section III. A microcomputer version of the technique should be developed that allows for all steps of the technique to be accomplished within one software package. This would include: a user-friendly interface, a profile generation module, a judgment data collection module, a regression analysis module, a clustering module, and a reporting module. The profile generation module would automatically generate profiles to be judged, using inputs from the decision analyst regarding attributes, distributions, and sample sizes. The judgment data collection module would provide the profiles to the decision makers in an automated form, and store the resultant judgments. The regression analysis module would be used to analyze the judgments and, together with the reporting module, would feed the information back to the analyst and the decision makers. The clustering module would be available for performing clustering of decision makers' equations, if desired.

Implementation of such a software package on the microcomputer should directly address the problem of intensive decision maker and analyst time and computer resource demands from using policy capturing. With

improved ease of use, speed, and feedback such a policy capturing model should be a great aid to problem understanding and problem solution.

Suggested Improvements to the Overall AFHRL Policy Modeling Capability

In the overall analysis, it is clear that neither policy specifying nor policy capturing answers all the needs of decision modeling, especially if context specific applications are considered. However, neither do the SMART and HAWM techniques fill every need. What is called for is development of a composite policy analysis/development tool, which makes available to the user all four techniques within an integrated framework. This framework would be developed in modular fashion, with modules for each of the functions required in a policy modeling tool: intelligent interface, problem structuring, policy development, policy analysis, sensitivity analysis, feedback, reporting, and batch profile generation and scoring.

TECHNICAL REFERENCES

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

Christal, R.E. (1967). *Selecting a harem--and other applications of the policy-capturing model* (PRL-TR-67-1, AD-658 025). Lackland AFB, TX: Personnel Resources Laboratory.

Christal, R.E., & Weissmuller, J.J. (1976). *New CODAP programs for analyzing task factor information* (AFHRL-TR-76-3, AD-A026 121). Lackland AFB, TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory.

DeWispelare, A. R. (1983). *The extended multi-attribute decision analysis model* (EMADAM) (AU-AFIT-EN-TR-83-3). Wright-Patterson AFB, OH: Air Force Institute of Technology.

Dudycha, L.W., & Naylor, J.C. (1966). Characteristics of the human inference process in complex choice behavior situations. *Organizational Behavior and Human Performance*, 1, 110-126.

Dyer, J.S., & Sarin, R.A. (1979). Measurable multiattribute value functions. *Operations Research*, 22, 810-822.

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

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

Edwards, W. (1971, June). Social utilities. In *Engineering Psychologist*, Proceedings of the Symposium on Decision and Risk Analysis Series. U.S. Naval Academy, MD.

Edwards, W. (1977). How to use multiattribute utility measurement for social decision making. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-7, 326-240.

Farquahr, P.H. (1975). A fractional hypercube decomposition theorem for multiattribute utility functions. *Operations Research*, 23, 941-967.

Farquahr, P.H. (1984). Utility assessment methods. *Management Science*, 30, 1283-1300.

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

Fishburn, P.C. (1965b). Analysis of decisions with incomplete knowledge of probabilities. *Operations Research*, 13, 217-237.

Fishburn, P.C. (1967). Methods of estimating additive utilities. *Management Science*, 7, 435-453.

Fishburn, P.C. (1970). *Utility theory for decision making*. New York: Wiley.

Fishburn, P.C. (1976). Utility independence on subsets of product sets. *Operations Research*, 25.

Gardiner, P., & Edwards, W. (1975). Public values: Multiattribute utility measurement for social decision making. In M.F. Kaplan, & S. Schwartz (Eds.), *Human judgment and decision processes*. New York: Academic Press.

Gabrielli, W., & Von Winterfeldt, D. (1978). *Are importance weights sensitive to the ranges of alternatives in multiattribute utility measurement?* (SSRI-TR-78-6). University of Southern California, CA: Social Science Research Institute.

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

Goldberg, L.R. (1968b). Seer over sign: The first "good" example? *Journal of Experimental Research in Personality,* 3, 168-171.

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

Goldberg, L.R. (1970b). Man versus model of man: Just how conflicting is that evidence? *Organizational Behavior and Human Performance,* 16, 13-22.

Green, P.E. (1974). On the design of choice experiments involving multifactor alternatives. *Journal of Consumer Research,* 1, 61-68.

Green, P.E., Carmone, F.J., & Wind, Y. (1972). Subjective evaluation models and conjoint measurement. Behavioral Science, 17, 288-299.

Green, P.E., & Rao, V.R. (1974). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research,* 8, 355-363.

Hammond, K.R. (1966). Probabilistic functionalism: Egon Brunswik's integration of the history, theory, and method of psychology. In K.R. Hammond (Ed.), *The psychology of Egon Brunswik.* New York: Holt, Rinehart, & Winston.

Hammond, K.R., Stewart, T.R., Brehmer, B., & Steinmann, D.O. (1975). Social judgment theory. In M.F. Kaplan & S. Schwartz (Eds.), *Human judgment and decisions processes.* New York: Academic Press.

Harker, P.T., & Vargas, L.G. (1985). *The theory of ratio scale estimation: Saaty's analytic hierarchy process.* Philadelphia: The Wharton School.

Holloway, C.A. (1979). *Decision making under uncertainty: Models and choices.* Englewood Cliffs: Prentice Hall.

Hwang, C., & Yoon, K. (1981). *Multiple attribute decision making: Methods and applications.* Berlin: Springer-Verlag.

Jones, K.M., Mannis, L.S., Martin, L.R., Summers, J.L., & Wagner, G.R. (1975). *Judgment modeling for effective policy and decision making.* Bolling AFB, Washington, DC: Air Force Office of Scientific Research.

Keeney, R.L. (1968). Quasiseparable utility functions. *Naval Research Logistics Quarterly,* 15, 551-565.

Keeney, R.L. (1972). Utility functions for multiattributed consequences. *Management Science,* 21, 276-287.

Keeney, R.L. (1974). Multiplicative utility functions. *Operations Research,* 22.

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

Krantz, D.H. (1964). Conjoint measurement: The Luce-Tukey axiomatization and some extensions. *Journal of Mathematical Psychology,* 1, 248-277.

Krantz, D.H., Luce, D., Suppes, P., & Tversky, A. (1971). *Foundation of measurement* (1). New York: Academic Press.

Luce, P., & Tukey, J.W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology,* 1, 1-27.

Meehl, P.E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press.

Meehl, P.E. (1965). Seer over sign: the first good example. *Journal of Experimental Research in Personality,* 1, 27-32.

Pollak, R.A. (1967). Additive Von Neumann and Morgenstern utility functions. *Econometrica,* 35, 485-494.

Raiffa, H. (1969). *Preferences for multiattributed alternatives.* (RAND Memorandum No. RM-5868-DOT/RC). Santa Monica, CA.

Saaty, T. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology,* 15, 234-281.

Saaty, T. (1980). *The analytic hierarchy process.* New York: McGraw Hill.

Saaty, T. (1986). Axiomatic foundation of the analytic hierarchy process, *Management Science,* 32, 841-855.

Saaty, T.L., & Takizawa, M. (1985). Dependence and independence: From linear hierarchies to nonlinear networks. *European Journal of O.R.*

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

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin,* 66, 178-200.

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

Stillwell, W., Von Winterfeldt, D., & John, R. (1987). Value tree analysis of energy alternatives. *Management Science,* 33, 442-450.

Suppes, P., & Winet, M. (1955). An axiomatization of utility based on the notion of utility differences. *Management Science,* 1, 259-270.

Torgerson, W. (1960). *Theory and methods of scaling.* New York: Wiley.

Vargas, L.G. (1984). *Preference relations, transitivity and the reciprocal property.* Working paper, Graduate School of Business, University of Pittsburgh, PA.

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior.* Princeton: Princeton University Press.

Von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research,* Cambridge University Press.

Ward, J.H., Jr. (1977). *Creating mathematical models of judgment processes: From policy capturing to policy-specifying* (AFHRL-TR-77-47, AD-A048 983). Brooks AFB, TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory.

Ward, J.H., Jr., Pina, M., Fast, J., & Roberts, D.K. (1979). *Policy specifying with application to personnel classification and assignment.* Paper presented at the Military Testing Association Meeting, San Diego, CA.

Ward, J.H., Jr., Treat, B.R., & Albert, W.G. (1985). *General applications of hierarchical grouping using the HIER-GRP computer program* (AFHRL-TP-84-42, AD-A150 266). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

BIBLIOGRAPHY

Allardyce, L.B., Amey, D.M., Feuerwerger, P.H., & Gulick, R.M. (1979). *Documentation of decision aiding software: Evaluation users manual.* McLean, VA: Decision and Designs.

Anderson, N.H. (1971). Information integration and attitude change. *Psychological Review,* 78, 171-206.

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

Anderson, N.H. (1981). *Foundation of information integration theory.* New York: Academic Press.

Anderson, N.H. (1982). *Methods of information integration theory.* New York: Academic Press.

Asbeck, E.L., & Haimes, Y.Y. (1984). The partitioned multiobjective risk method (PMRM). *Large Scale Systems,* 6, 13-38.

Barron, F.H. (1977). Using Fishburn's techniques for analysis of decision trees: Some examples. *Decision Sciences,* 4, 247-267.

Barron, F.H. (1973). Axiomatic conjoint measurement. *Decision Sciences,* 8, 548-599.

Barron, F.H. (1980). Concepts, theory, and techniques: Empirical validation of elicited utilities for prediction and decision making. *Decision Sciences,* 11(3), 411.

Brown, C.A., & Valenti, T. (1983). *Multiattribute tradeoff system: User's and programmer's manual.* Denver, CO: Bureau of Reclamation, Engineering and Research Center.

Chankong, V., & Haimes, Y.Y. (1978). *The interactive surrogate worth trade-off method for multiobjective decision making.* New York: Springer.

Chankong, V., & Haimes, Y.Y. (1983). *Multiobjective decision making: Theory and methodology.* New York: North Holland.

Dantzig, G.B. (1963). *Linear programming and extensions.* Princeton, NJ: Princeton University Press.

Dyer, J.S. (1972). Interactive goal programming. *Management Science.* 19, 62-70.

Feinberg, A. (1972). *An experimental investigation of an interactive approach for multicriterion optimization with an application to academic resource allocation.* (Working Paper No. 186). Los Angeles: University of California, Western Management Science Institute.

Fishburn, P.C. (1973). Bernoullian utilities for multiple-factor situations. In J.L. Cochrane & M. Zeleney (Eds.), *Multiple criteria decision making.* Columbia, SC: University of South Carolina Press.

Geoffrion, A.M., Dyer, J.S., & Feinberg, A. (1972). An interactive approach for multicriteria optimization with an application to an academic department. *Management Science,* 19, 357-368.

Goicoechea, A., Hansen, D.R., & Henrickson, R.L. (1981). A hierarchical-multilevel approach to the development and ranking of systems in the tanning industry. *Cumpt. and Industrial Engineering,* 5, 33-47.

Goicoechea, A., Stakhiv, E.Z., & Moy, W.S. (1984, May). *Risk and uncertainty analysis in water resources planning.* Paper presented at the Urban 84 Specialty Conference of the ASCE Water Resources Planning and Management Division, Baltimore, Maryland.

Green, P.E., & Carmone, F.J. (1970). *Multidimensional scaling and related techniques in marketing analysis.* Boston: Allyn and Bacon.

Green, P.E., & Wind Y. (1973). *Multiattribute decisions in marketing: A measurement approach.* New York: Holt, Reinhart and Winston.

Haimes, Y.Y. (1979). The surrogate worth trade-off method and its extensions. In G. Fandel & T. Gal (Eds.), *Multiple criteria decision making theory and applications* (pp. 85-108). Berlin, West Germany: Springer.

Haimes, Y.Y., & Hall, W.A. (1974). Multiobjectives in water resources systems analysis: The surrogate worth trade-off method. *Water Resources Research,* 10, 614-624.

Haimes, Y.Y., Hall, W.A., & Freedman, H.T. (1975). *Multiobjective optimization in water resources systems: The surrogate worth trade-off method.* Amsterdam: Elsevier.

Haimes, Y.Y. (1981). Risk benefit analysis in a multiobjective framework. In Y.Y. Haimes, (Ed.), *Risk/benefit analysis in water resources planning and management* (pp. 89-122). New York: Plenum.

Haimes, Y.Y., & Hall, W.A. (1980). Sensitivity, responsivity, stability, and irreversibility as multiple objectives in civil systems. *Advances in water resources,* 16, 465-475.

Haimes, Y.Y., & Tarvainen, K. (1981). Hierarchical-multiobjective framework for large scale systems. In P. Nijkamp & J. Spronk (Eds.), *Multicriteria analysis in practice* (pp. 202-232). London: Gower.

Hall, W.A., & Haimes, Y.Y. (1975). The Surrogate worth trade-off method with multiple decision-makers. In M. Zeleny (Ed.), *Multiple criteria decision-making* (pp. 207-233). New York: Springer.

Jusko, M.J., & Whitfield, R.G. (1980). A guide to IDAP, version 2: *An interactive decision analysis procedure.* Argonne, IL: Argonne National Laboratory.

Keeney, R.L., & Sicherman, A. (1976). An interactive computer program for assessing and using multiattribute utility functions. *Behavioral Science,* 21, 173-182.

Lavalle, I. (1973). On admissability and Bayesness, when risk attitude but not the perference ranking are allowed to vary. In J.L. Cochrane, M. Zeleny, (Eds.), *Multiple criteria decision making.* Columbia, SC: University of Columbia Press.

Lapin, L.L. (1985). *Quantitative methods for business decisions.* New York: Harcourt Brace Jovanovich.

Mumpower, J., & Bollacker, L. (1981). *User's manual for the evaluation and sensitivity analysis program (ESAP).* Vicksburg, VA: U.S. Army Engineer Waterways Experiment Station.

Nijkamp, P., & Vos, J.B. (1977). A multicriteria analysis for water resource and land use development. *Water Resources Research,* 13, 513-518.

Roy, B. (1971). Problems and methods with multiple objective functions. *Mathematical Programming,* 1(2).

Roy, B. (1973). How outranking relation helps multiple criteria decision making. In J. Cochrane & M. Zeleny (Eds.), *Multiple criteria decision making* (pp. 179-201). Columbia: University of South Carolina Press.

Roy, B. (1977). Partial preference analysis and decision-aid: The fuzzy outranking relation concept. In D.E. Bell, R.L. Keeney, & H. Raiffa, (Eds.), *Conflicting objectives in decisions* (pp. 40-75). Great Britain: Wiley, Chicheser.

Roy, B., & Vincke, P.H. (1984). Relational systems of preference with one or more pseudo-criteria: Some new concepts and results. *Management Science,* 30, 1323-1335.

Tarvainen, K., & Haimes, Y.Y. (1981). Hierarchical multiobjective framework for energy storage systems. In J. Morse (Ed.), *Organizations: Multiple agents with multiple criteria* (pp. 424-446). Berlin: Springer.

Tarvainen, K., & Haimes, Y.Y. (1982). Coordination of hierarchical multiobjective systems: theory and methodology. *IEEE transactions of systems, man, and cybemetics, T-SMC* (pp. 751-764).

Turban, E., & Meredith, J.R. (1985). *Fundamentals of management science* (3rd ed.). Plano TX: Business Publications.

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

Ward, J.H., Jr., Haney, D.L., Hendrix, W.H., & Pina, M. (1977). *Assignment procedures in the Air Force procurement management information system* (AFHRL-TR-78-30, AD-A056 531). Brooks AFB, TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory.

Zionts, S., & Wallenius, J. (1976). An interactive programming method for solving the multiple criteria problem. *Management Science*, 22, 551-562.

APPENDIX:
AXIOMATIC BASIS FOR POLICY SPECIFYING

There appear to be two theoretical areas in which the policy specifying technique could be improved substantially without major changes to the technique or loss of practicality: providing an axiomatic foundation for the judgments required in eliciting the utility functions and providing an axiomatic foundation for the polynomial model forms.

An Axiomatic Basis for Judgments Required in Policy Specifying

Policy specifying procedures usually employ numerical rating scales of hypothetical objects varying on two attributes on a scale from 0 to 100, where the rating reflects the strength of preference or relative degree of desirability or achievement. Typically, the hypothetical objects are chosen to array at the extreme points of a two-attribute plot of possible objects, or different ratings are performed for one attribute at differing levels of the other attribute. These two-attribute ratings are then fitted by some polynomial.

The rating task requires relative difference judgments; i.e., judging the relative spacing in desirability (or some other value-relevant aspect) of the two-dimensional objects. Thus, difference measurement theory (Krantz et al., 1971) is an appropriate basis for providing justification for these judgments. In simple terms, this theory requires that

1. pairs of objects can be ordered in terms of the relative preference of one over the other;

2. these strength of preference judgments are transitive;

3. strengths of preferences "add"; i.e., if a is preferred over b and b is preferred over c, then the strength of preference over c should be the "sum" of the strength of preference of a over b and the strength of preference of b over c;

4. for each tuple of pairs, it is possible to vary one of the elements so that the strength of preference in one pair matches the strength of preference in the other;

5. there are no infinitely desirable or undesirable objects in the set of alternatives that are to be evaluated.

These assumptions are formally stated by Krantz et al. (1971), who showed that the assumptions imply that a function must exist that maps objects a, b, c, and d into the real numbers such that

$$(a,b) \stackrel{\bullet}{\succsim} (c,d)$$
$$\text{if and only if}$$
$$v(a)-v(b) > v(c)-v(d)$$

where $(a,b) \stackrel{\bullet}{\succsim} (c,d)$ is interpreted as "the strength of preference of a over b is greater than or equal to the strength of preference of c over d."

Difference measurement has been shown in several studies to be the appropriate formal basis for interval scale measurements like the ones required in policy specifying. A possible complication arises in policy specifying, if the relation of strength of preference substantively changes as one moves from one part of the hierarchy to another; i.e, as one compares two-dimensional objects in different parts of the tree. Such shifts should be spelled out clearly in the instructions for making these judgments. They do not pose any problems for the formal application of difference measurement, except that, in principle, assumptions 1-5 would have to be tested in each specific case

Assuming that at all levels and places in the tree the relation $\stackrel{\bullet}{\succsim}$ has the same meaning is, of course, much more convenient; it has strong implications for the resulting functional forms. In particular, it prohibits the possibility of polynomials that include different utility functions defined over the same attribute. This issue will be discussed next.

An Axiomatic Foundation for Polynomial Model Forms

In most applications of policy specifying, a polynomial aggregation rule has been assumed. With n attributes, a simple polynomial is a linear combination of products of the attribute variables raised to non-negative, integral powers. For example, a typical polynomial for three attributes would be

$$v\,(x,y,z) = x^3\,y^2 + z^4\,z + x\,y^3\,z^2.$$

Providing an axiomatic basis for such polynomials within the framework of difference measurement theory requires that the aggregation rules be substantially more complicated than the additive or multiplicative forms usually developed in the literature. Furthermore, the same attribute domain can, in principle, occur with several different power coefficients, thus in essence creating the possibility for different utility functions defined on the same attribute.

A "simple" version of a polynomial difference model will be considered first. Each attribute can have only one power coefficient and is termed the simple multilinear polynomial model. In this model, the three-attribute polynomial

$$v(x,y,z) = x^2 + y = z^3 + x^2\,z^3 + y\,z^3 + x^2\,y + x^2\,y\,z^3 \tag{A-1}$$

would be admissible, but the following (seemingly simpler) polynomial would not be allowed:

$$v(x,y,z) = x + y + z + x^2\,z, \tag{A-2}$$

because in it x appears with two different power coefficients.

Simple multilinear polynomials follow directly from two assumptions: (a) The relation $\dot{\succsim}$ has the same meaning everywhere in the tree; and (b) at each pair-wise comparison, the aggregation rule must be a simple multilinear form:

$$v(x_1, x_2) = v_1\,(x_1) + v_2\,(x_2) + w\,v_1\,(x_1)\,v_2\,(x_2). \tag{A-3}$$

The "polynomial" part of the model would further restrict the v's to be positive integral power functions of the x's, or a power function of some positive linear transformation of the x's. Specifically, a simple polynomial form of (A-3) would be

$$v(x_1, x_2) = (ax_1 + b)^m + (cx_2 + d)^n + w(ax_1 + b)_m\,(cx_2 + d)^n \tag{A-4}$$

where m and n are positive integers.

It turns out that forms like (A-3) have a very straightforward axiomatic base in difference measurement, which is described, for example, in Dyer and Sarin (1979) and Von Winterfeldt and Edwards (1986). The key assumption is multilinear difference independence (Dyer and Sarin call it "weak difference independence"). It requires that the order of strengths of preferences made for pairs of objects that vary only in one attribute is unaffected by constant values of the other attribute. This assumption, together with assumptions 1-5 described earlier, justifies form (A-3). The simple polynomial form (A-4) is somewhat more restrictive, and there exist no necessary behavioral assumptions justifying it. However, (A-4) could simply be incorporated into a policy specifying model as a form of "curve fitting" routine, rather than an explicit behavioral assumption. The idea would be that it should be fairly easy to fit most judgments satisfying (A-3) with a polynomial of the form (A-4).

Examination of what happens to the terms of models (A-3) and (A-4) if one moves around in the hierarchy indicates that lateral movements do not affect the model. Since the domain of the single-attribute utility functions is changed, any power that is likely to fit the judgments that are provided can be selected. However, an important restriction occurs when an attempt is made to compare and integrate two higher level objectives

that are now expressed in terms of the v's. For example, consider the simplified structure in Figure A-1. As before, the assumption is that the aggregation rule is a simple multilinear polynomial; i.e.,

$$v(v_x, v_y) = v_x{}^m + v_y{}^n + kv_x{}^m v_y{}^n. \tag{A-5}$$

The problem is that by virtue of the assumption that the *relation at the second level of comparison* must be identical to that of the lower level, it is required that m = n = 1. In other words, a utility of a utility is a utility. The necessary implication of assuming m = 1 or n = 1 is that the strength of preference order at the higher level could be in contradiction with the strength of preference order at the lower level. By assuming identical relations at all levels, that contradiction is disallowed and thereby the possible polynomial forms for policy specifying substantially reduced.

To create the possibility for richer forms, $\dot{\gtrsim}$ must mean something different at different places in the tree --a somewhat messy, but possibly justifiable assumption. To consider an example, assume that in Figure A-1, x_1 and x_2 are two variables measuring an individual's knowledge about a subject, and y_1 and y_2 are measures of physical strength. When comparing the first pair, the decision maker may think of the "importance of knowledge." When comparing the second pair, the decision maker may think of the "importance of physical strength." Both individual functional forms may be extremely simple; e.g.,

$$v_x(x_1, x_2) = x_1 + x_2$$

and

$$v_y\,(y_1, y_2) = y_1 + y_2.$$

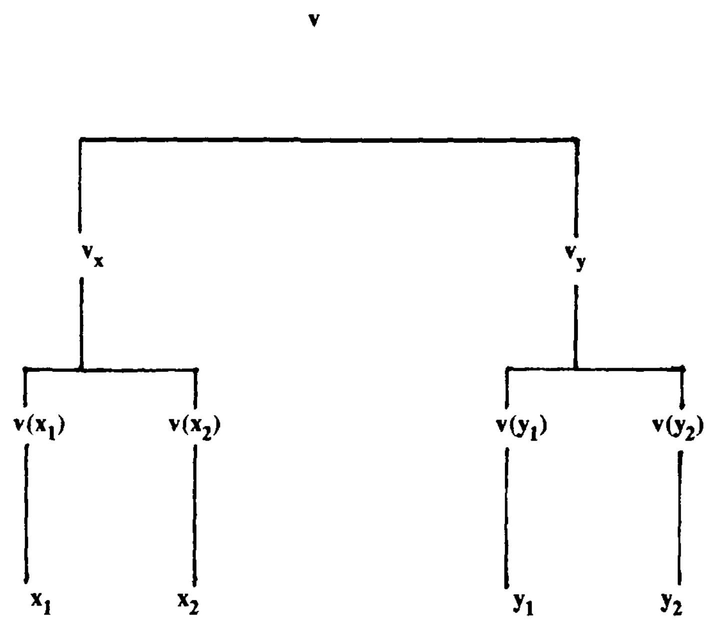

Figure A-1. Tree Structure to Illustrate Hierarchical Utility Assessment.

When comparing the "aggregate" knowledge score and the "aggregate" physical strength score, the evaluator may think in terms of how desirable it is for a person seeking a job to have some combination of v_x and v_y, and there is no reason not to assume a functional form like

$$v(v_x, v_y) = v_x^a + v_y^b + kv_x^a v_y^b, \tag{A-6}$$

thus

$$v(x_1, x_2, y_1, y_2) = (x_1 + x_2)^a + (y_1 + y_2)^b + w(x_1 + x_2)^a(y_1 + y_2)^b \tag{A-7}$$

which is a substantially more complex form than the proposed simple multilinear polynomials. To contrast these model forms, the model generated by (A-6) will be termed the general multilinear polynomial.

To construct an axiom system for the general multilinear polynomial model is not difficult in principle, but would theoretically require checks of axioms at each level of the tree. However, once this is accomplished, any form of polynomials can be generated.

* U. S. GOVERNMENT PRINTING OFFICE: 1988 - 561048 / 80038