# SOME BEHAVIORAL ASPECTS OF ELICITING UTILITY (USING THE MACCRIMMON-TODA METHOD FOR ORDINAL UTILITY AND THE STANDARD GAMBLE METHOD FOR CARDINAL UTILITY)

by

EUGENE WONG
BSc.,University of British Columbia, 1971

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN BUSINESS ADMINISTRATION
in the Management Science department of
the Faculty of Commerce

We accept this thesis as conforming to the
required standard

THE UNIVERSITY OF BRITISH COLUMBIA

October 1973 A.D.

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Management Science, Faculty of Commerce

The University of British Columbia
Vancouver 8, Canada

Date Oct 15, 1973.

Abstract

This study investigates some behavioral aspects and properties of eliciting utility. Previous investigations devoted to empirical utility measurement have stemmed from the work of experimentalists who have applied various utility models in an effort to measure utility. However, empirical studies devoted to investigation into behavioral factors which may bias the measurement are lacking and it is this gap in the utility literature that prompted our empirical study. We chose to examine the standard gamble method for deriving von Neumann-Morgenstern cardinal utility and the MacCrimmon-Toda method for deriving indifference curves. The domain of choice involved hospital days in bed with risk of additional days. The analysis consisted of identifying relationships between behavioral factors and properties of choice predictions obtained by the methods. Furthermore, the study also provided a means for comparing properties of the two methods for eliciting utility.

Among other findings, the results show that not all subjects expressed agreement with the appropriateness of specific axioms of behavior which underly some methods for eliciting utility and that not all people express constant sensitivity over all stimuli levels. The two results in themselves suggest that a priori assumptions regarding "rationality" and infinite sensitivity may have to be reexamined. The preferences elicited by both methods seem to suggest that the subjects follow a linear rule to trade-off

sure outcome and risk. Although correspondence between test-retest preferences predicted by the standard gamble was generally closer than that for the MacCrimmon-Toda method, the MacCrimmon-Toda method had generally better predictive ability. Our results also indicate that certain behavioral factors seem to affect preferences predicted by the methods as we hypothesized. This observation has implications for practical measurement of utility since "successful" application of methods for eliciting preferences depends upon our awareness of which behavioral factors may bias the measurement.

Table of Contents

List of Figures

Acknowledgements

I would like to especially thank Dr. Ilan Vertinsky, the chairman of my thesis committee, for his valuable assistance and guidance during the undertaking of this thesis project. Also, I would like to thank him for providing me with a research assistantship in the Health Systems Group of the Resource Science Centre during my graduate years and allowing me generous use of the resources available to this research group. I would like to thank the committee members Dr. Dean Uyeno and Dr. Jeff Sidney. Last but certainly not least I would like to extend my appreciation to the 23 subjects who volunteered their time and effort toward this study.

Dedicated to my parents

## Chapter 1 - An Introductory Discussion of Some Philosophical and Methodological Aspects of Utility Measurement

Our very destinies are determined by the decisions that we make in our daily lives. When confronted with choices, we may experience difficulty in reaching a decision due to our ignorance of the decision environment, i.e. our ignorance of the probabilities of outcomes, the scope of actions available, the decision framework, etc. However, more importantly, we may not be aware of what we actually "want" or at least are unable to present our preferences or desires in a communicable (eg. easily understandable, unambiguous, meaningful) manner although one often assumes that the nature of human needs and desires are obviously apparent.

In order to make meaningful recommendations concerning decisions to a decision-maker one must be able to elicit his subjective values. An important consideration is whether there is some basic "sensation" that determines for him the value that he attaches to objects, actions, etc.

Can feelings of pleasure and pain be used as a basis for value judgements? Bentham (1907) felt that pleasure and pain alone guide us to what we ought to do, as well as to what we shall do. However, the immediacy of pleasure sensations do not give us an adequate assessment of values since sensations may be felt but are difficult to communicate. Even if communication is possible, verification or validation of sensation is extremely difficult if not impossible. Therefore, the methodological problem is to

translate pleasure and pain response into some measure which can be communicated and manipulated. Philosophers who had originally advocated the principle of pleasure as a basis for value had not gone to the point of finding a reasonable way of measuring pleasure. Antithetically, economists had not been so concerned with the philosophical issues of value as the measurability or quantification of pleasure, or what the economists termed utility[1].

What do we mean by measurement? Measurement involves the assignment of numbers to objects, observations, outcomes, etc. whereupon "allowable" operations performed on these numbers will reveal new information about the entities measured. Alchian (1953) identifies three main aspects in the process of measurement: (1) the purpose of the measurement, (2) the method of measurement, and (3) the arbitrariness of the assigned numbers.

Underlying the three main aspects is the assertion that measurement is always invented and never discovered since measurement is not a property which is inherent in entities. The method of assignment of numbers to entities is determined merely by its convenience for the purpose. Thus, the arbitrariness of aspect 2 is limited by aspect 1, but for no other reason than convenience or manageability. The

---

[1] The term is inherited from Bentham and his utilitarian philosophy; in the context of utilitarianism, which is the doctrine that the greatest happiness of the greatest number shall be the end and aim of all social and political institutions, utility is the standard of morality - actions are right if they promote happiness.

arbitrariness of the assigned numbers (aspect 3) is inherent in the rule used for assigning numbers (aspect 2). In order to perform certain operations with these numbers in such a way as not to alter the information that these numbers were assigned to convey, the structure of the measurement must be isomorphic to some numerical structure which includes these operations. The following four basic measurement scales are of special interest here: (1) nominal, (2) ordinal, (3) interval, and (4) ratio.

The nominal or classificatory scale is the weakest of the four measurement scales. The scaling rule consists of partitioning a given class into a set of subclasses that are mutually exclusive and equivalent in the property being scaled. Numbers are then assigned to objects that belong to the same subclass. A one-to-one transformation will preserve equivalence. In other words, all scales derived by a one-to-one transformation have the same equivalence property.

In an ordinal or ranking scale, numbers are assigned to objects such that group membership and the ordering relationship (>) may be identified. Any monotonic transformation will preserve the equivalence and ordering relationship in an ordinal scale.

When the scale has the characteristics of an ordinal scale and furthermore when the distances between any two numbers on the scale are of known "size", i.e. when the same numerical difference between any two numbers on the scale reflect the same difference in the property measured,

an interval scale measurement has been achieved. The differences between numbers on the interval scale are isomorphic to the structure of arithmetic. It is this property that makes the interval scale a quantitative scale. An interval scale is unique up to a linear transformation.

A ratio scale has all the characteristics of an interval scale and in addition has a true "zero point" as its origin, i.e. the number zero is assigned to an object which possesses the true zero in the property being scaled. Multiplication by a positive constant is a linear transformation and thus will preserve the interval properties; in addition the true property "zero" will still be assigned the number zero (zero multiplied by a positive constant is still zero). Thus the ratio scale is unique up to multiplication by a positive constant. Furthermore, the scale is isomorphic to the structure of arithmetic and is thus a high order level of measurement.

In regards to the purpose of measurement, the normative purpose of utility measurement is to provide a basis for decision and policy. The decision-maker has to decide which type of measurement is most suitable for his purpose and his choice of measurement involves a tradeoff between economy of measurement and need for detail.

Historically, economists had been interested in utility as a descriptive theory; they were concerned with the measurability of utility in an attempt to explain consumer behavior. Economists had designated the term utility (out of

tradition as indicated in in the previous footnote) to the subjective value of commodities. Experimental psychologists, interested in psychophysics, had been measuring human response as a function of measurable physical stimulus whereupon the measurements are used to construct a subjective scale directly from the subject's own quantitative estimates of the scale values of a series of stimuli. From the functional relationship, measures of responses were derived. Fechner's law was an early attempt to express a functional relationship between the physical magnitude and the psychological magnitude of the sensation it arouses. Brightness may be measured in brils; loudness may be measured in sones[2]. Can utility be measured in utils?

It was in the light of this question that economists were searching for a measure of utility. Experimental economists were not so interested in the internal pleasure-feeling aspect as in the external choice behavior of the subject. The behavioristic interest is in the observed behavior which arises from simple sensations. Thus we have an observer interpreting observed behavior versus the subject introspecting his own feelings. For the behaviorist, the mind of the person is the observed behavior. For the introspectionist, there is no such equivalence. Obviously, direct empirical refutation of introspection is impossible. The closest we can come to validation of introspection is to

---

[2] Brils and sones are both subjective scales that are discussed by Stevens (1959).

examine its ability to derive theorems of behavior which it claims to enunciate. Introspectionism regards simple sensations such as pleasure as primitive notions which are not definable in any other terms. Therefore, in order to avoid interpreting or validating introspection it seems natural that values be operationally defined in terms of choice. However, even defining values by choice does not guarantee that behavioristic expressions will not be erratic, whimsical, careless, etc. Thus, although our conception of values spring from inner impulsive emotive feelings, our behavioristic expressions of our values must be guided by judgement of one sort or another in order to be useful information. Can standards of judgement be provided that will form not only ground rules for measurement of utilities but also norms by which one "should" abide in making judgements? In order to provide a framework for measurement of utility, judgement of values (eg. the exhibition of preferences) is restricted to that observed under so called rational behavior. A set of intuitively plausible assumptions (axioms) of rationality provide the formal conditions for rational patterns of preference. In this context, "inconsistent" behavior may be interpreted as behavior which contradicts one of the axioms of rationality. However, it is conceivable that a person's behavior contradicts one of the axioms yet he still considers himself to be consistent (or rational). Here, a person questions the axioms as criteria for rationality. Savage (1954) suggests that a person's acceptance of or agreement with an axiom may

be indicated by the degree to which he tries to act in accord with it or by his willingness to revise his behavior when made aware of his violation with the axiom.

The obvious interpretation of the axioms is that they provide the definitions of rationality. Secondly, in terms of providing ground rules for measurement of utility, the axioms of rationality may be interpreted as rules which must be obeyed before meaningful measurement is possible. Loosely speaking, this means that erratic or careless behavior cannot be interpreted in terms of values. In addition, these axioms may be translated into prescriptive statements regarding behavior, i.e. one's behavior should abide by these axioms; interpretation is also possible in terms of predicted rational behavior, i.e. one's behavior will be in accord with these axioms if choices are free and rational.

Existence of certain proposed utility functions is dependent on the axiomatic system upon which the proposition is based. In this sense, the axioms take on the same role as axioms in mathematics from which theorems are derived. Thus, the existence of a certain proposed utility function (or in other words, the possibility of finding a scale of numbers that express the utilities in a convenient way) is logically equivalent (<=>) to the hypothesis that behavior is consistent with the underlying axioms. Empirical test of the proposition that a given utility function exists is thus displaced to the axioms that imply it. It is a much easier task to test the axiomatic system statement by statement (in

isolation from the theorized utility function) than it is to question the validity of the utility function (in isolation from the axiomatic system which theorized it). It is a general observation that the stronger the measurement (ratio is stronger than interval, than ordinal, than nominal) of the proposed utility scale, the stronger the axioms needed to derive it.

The ordinal scale of utility essentially requires only the basic axiom of transitivity of preferences for its existence. It is therefore not unusual that the possibility of utility measurement in ordinal terms is fairly well agreed upon among value theory philosophers (see Davidson et al, 1955). In particular, Perry wrote:

> ... The important feature of preference is that it arranges the objects of any given interest in a order, relatively to one another, and in a manner that cannot be reduced either to the intensity or to the inclusiveness of the interest. This order of preference has its own characteristic magnitudes, which determine comparative values ... Preference generates a similar (his reference to the color spectrum) transitive, asymmetrical relation among its terms. It is transitive because if b is preferred to a and c to b, then c is preferred to a ... The three terms in order of preference may be said to constitute a "stretch" which is greater than any of its included stretches. Thus the stretch a-c is greater than the stretch a-b, or we may say that c lies beyond b. (1950, pages 635-636)

On the other hand he and other value theory philosophers, who had accepted ordinality, expressed scepticism about the possibility of measurability on a higher scale. One suspects that early (historically)

rejection of the cardinality (interval scale) of utility lie in the failure of expressing the cardinal proposition in terms of an acceptable axiomatic system. As a result, the argument for rejection revolved around the quantitative aspect of cardinality. Thus, the early cardinality proposition was rejected for mistaken reasons. As a typical example concerning this point, Whellwright, who rejected a higher-than-ordinal measure, wrote:

> We may on a particular occasion prefer reading a book to taking a walk: the former, then, we say, would give us (on this occasion) the greater pleasure. But is there any conceivable sense in which we could say that the intensity of the pleasure to be got from reading is twice rather than three times or one and a half times, the intensity of the pleasure to be got from walking? (the emphasis is mine) Would we not, by trying to make our comparison of intensities mathematically exact, reduce it to meaninglessness? (1949, page 87)

Along the same lines, Lewis wrote:

> ...numerical measure cannot be assigned to an intensity of pleasure, or of pain, unless arbitrarily. Intensities have degree, but they are not extensive or measurable magnitudes which can be added or subtracted. (the emphasis is mine) That is, we can- presumably- determine a serial order of more or less intense pleasure, more and less intense pains, but we cannot assign a measure to the interval between such. (1946, page 490)

However, for economists the question of measurability took on more pragmatic and less philosophical tones. The early objectors to any search for a cardinal utility claimed that ordinal utility (re marginal rates of substitution) could explain the aspects of economic behavior (eg. market behavior under certainty) which the proposed cardinal

utility (re marginal diminishing utility) could. However, cardinality revived itself in the face of welfare economics and risky[3] choices. A group of economists, of which Jevons, Menger, and Marshall are typical, asserted that the ordinal measurability of relative preferences at the level of introspection implies the possibility of an interval scale for utility[4]. In a Jevonsian[5] experiment, the person is asked to rank differences in utility of entities. Any numerical indices that preserve the ordinal relationship between utility differences will be related by a linear transformation. Thus the utility index is said to be cardinal. On the basis of a Jevonsian utility index, the rule of maximization of expected utility was used as a prescription for choices between wagers even though the

---

[3] In an attempt to develop an analytic framework for dealing with the degree of knowledge we have of our decision environment, various partitions between complete knowledge and complete ignorance have been attempted. The most accepted is a partition into unmeasurable uncertainty, measurable uncertainty, and certainty. For measurable uncertainty, which is sometimes called risk, the probability distribution of the uncertain event is known, while for unmeasurable uncertainty, a distribution is not given.

[4] The construction of a scale based upon ordered differences was first discussed by Coombs (1950). He then called such a scale an ordered metric and placed it between the ordinal and the interval scale. The possibility of an ordinal scale on intervals implies the possibility of a nominal scale on intervals, i.e. in principle the operations sufficient for an ordered metric scale ought to be sufficient to determine an interval scale. Stevens (1959) concludes that the ordered metric scale appears in practice to be a type of unfinished interval scale.

[5] The term was suggested and used by Weldon (1950) and had since been adopted by Ellsberg (1954).

[6] Formally, a wager is a set of alternative mutually exclusive outcomes, each of which occurs with stated probabilities. The terms "prospect" and "gamble" are often used in place of "wager".

index was not derived from any risk behavior[6]. This principle of maximization governing risk behavior is not implied by Jevon's concept of utility or by the methods of measuring it. In this regard, the emphasis upon mathematical expectation is arbitrary and any other normative criteria for choice is equally meaningful. It was not until the appearance of the von Neumann-Morgenstern utility theory that it was shown formally that the existence of an interval scale of utility, which is derived from exhibited choice behavior under risk, implies the norm of maximization of expected utility. Previously, there had been suggestions that such a norm was implicit in rational behavior. Ramsey (1926) had earlier hypothesized that a person acts upon his beliefs in such a way as to maximize the overall good. Bernoulli (1738) defined the moral expectation of a decision, which was to be maximized, as the mathematical expected value in accord with the expected utility model.

The hypothesis of the von Neumann-Morgenstern model is that the maximization of expected utility is sufficient for predicting the free choice behavior of a rational person. From the viewpoint of objective[7] probability, the expected utility concept seems plausible because the weighted sum of the outcome utilities is the expected long-run utility of the action. Using this utility model, the method of eliciting utility consists of observing an individual's

---

[7] The concept of objective probability will be discussed later in this chapter.

behavior in the very simplest of risky situations: a choice between a constructed wager and a certain outcome (Swalm, 1966).

From the observations of a person's choices in these simple constructed risk situations, von Neumann and Morgenstern theorized that it was possible to predict his choice in more complicated risk situations. From a person's introspection into his preferences for certain outcomes, the Jevons school assumed that the same predictive ability claimed by the von Neumann-Morgenstern model was possible. The obvious operational difference between the two cardinal measures is however masked by the similarity in their algebraic formulations; the two schools summarize their results by algebraic expressions (namely statements expressing maximization of expected utility) that are mathematically equivalent[8].

By employing wagers to indirectly elicit utilities, the von Neumann-Morgenstern theory avoided many of the objectionable methodological features of the Jevonsian and other direct methods[9] of measuring utility. In a psychophysics experiment, a series of visual intensities

---

[8] The confusion is also due to the use of the same term "cardinal utility" to denote the result of two different operations. On the operational approach, Bridgeman states: "If we have more than one set of operations we have more than one concept, and strictly there should be a separate name to correspond to each different set of operations." (1927, page 10)

[9] Some of these direct methods will be discussed in Chapter 2.

whose brightness is to be judged can be presented without attaching numerical indices to the visual intensities, but in a utility experiment magnitudes of wealth whose utility is to be judged cannot be presented without attaching numerical indices to the magnitudes of wealth. As a result, one of the two a priori arguments against the method of direct estimation of utility is that the subject may judge the physical magnitude of the stimulus and not the subjective effect thereby committing what the psychophysicists call the stimulus error. Also by limiting concern to choice behavior among wagers, the von Neumann-Morgenstern utility index is more firmly tied to a behavioristic basis than to an introspective basis and thus facilitates easier interpretations. Direct quantitative estimates may prove difficult to make because of their introspective nature (Stevens, 1959).

While for the direct methods of measuring utility in which riskless choices (or certain outcomes) are presented for which the utility alone is maximized, the introduction of risky choices into the measurement of utility presents an additional cue, namely that of probability.

Formally, objective probability measures the agreement between outcomes of repeatable physical events on the one hand and outcomes of hypothetical mathematically defined events on the other. This traditional interpretation of the probability concept is based on the relative frequency with which the event occurs in a long series of observations.

Alberoni (1962) described an experiment in which the subjects were shown a sequence of binary events and were told to predict future sequences. The subjects invariably associated the observed relative frequencies with probabilities and were increasingly confident that the events were due to "chance" as more event occurrences conformed to their expectations were observed. However, this interpretation of probability as frequency fails when one considers probability of events that cannot be repeated in order to yield long-run relative frequencies. Such is the case in practical decision problems. As an example, consider an event in which a certain company X will make more than one million dollars annual profit as a result of introducing a new product Y into the market. The management of company X will surely not consider stopping the introduction of product Y into the market because there is not enough data to construct long-run relative frequencies of outcomes. It is in situations such as this that the decision-maker must subjectively assess the likelihood of such an event occurring without recourse to the long-run frequency concept. It is this criticism of the objective concept of probability that led Savage (1954) to conclude that "the grounds for adopting an objectionistic view are not overwhelmingly strong". Is there empirical evidence to suggest that people do not always behave in accord with the objective view of probability?

In experiments of probability revision, subjects tended to revise their probabilities in the same direction as

dictated by Baye's theorem but in a much more conservative manner (Edwards, Lindman, Phillips, 1965). The experiments of Edwards and Phillips (1966) revealed that when subjects make datum-by-datum revisions throughout a sequence of data, the final subjective probability is far more conservative than the statistically calculated combination of series of single estimates made by subjects for each datum in the sequence. Alberoni (1962) had subjects estimate various binomial sampling distributions. The sums of the estimated probabilities for the different outcomes consistently totalled about 0.85, considerably less than the 1.0 required by probability theory. These experimental results lead one to adopt an additional concept of probability interpretation, that known as subjective probability.

The subjective view interprets probability as a measure of an individual's confidence in the truth of a particular proposition. Statisticians have devoted much effort in the development of formal procedures for dealing with fallible information in making inferences about prevailing and future states of the decision environment. Mathematical definitions of probability are meant to provide an analytic procedure for calculating probability. However, the intellectual process is much more subjective and in situations where the objective measurement of probability is unnecessary, uneconomical, unreasonable, or unattainable, the probability of the occurrence of an event may be described by the measure of the degree to which the belief of the occurrence of the event is substantiated by a group of people. A belief

is well substantiated if it is accepted by most people (breadth) and is intense (depth) (Churchman, 1961).

Therefore, while the von Neumann-Morgenstern theory has made the cardinal utility notion with its implied norm of maximization of expected utility a more meaningful concept than the Jevonsian theory, the empirical derivation of an utility index has introduced an additional stimulus that is subjectively interpreted by the subject, namely that of probability. Thus, value researchers who want to measure subjective estimates of outcome values (utilities) using models of risky choice must realize the inevitable confounding of subjective estimates of values with that of subjective interpretation of probabilities for which the unfortunate consequence is that the subject's estimates of probability and utility may be mutually biased. The difficulty of isolating probability from utility in experiments has been one of the methodological problems inherent in the construction of the von Neumann-Morgenstern utility index. Thus, experimenters (eg. Mosteller and Nogee, 1951) who have obtained utility indices have had to make assumptions regarding subjective estimates of probability, while experimenters (eg. Preston and Baratta, 1948) who have obtained subjective probability measures have had to make assumptions regarding the utility function.

In this chapter, some of the concepts of utility measurement which have been highlights in the history of utility theory have been discussed. In Chapter 2 we will

examine some of the numerous alternative ways of eliciting and measuring utility, each with its own assumptions regarding concepts of utility and capabilities of human judgement. In Chapter 3, we will return to the von Neumann-Morgenstern model and focus our attention upon this utility index and the indifference map concept, both of which have special theoretical and behavioral aspects that are of interest to us.

# Chapter 2 - Methods of Estimating Utilities

## Section I - Introductory Discussion

If at the outset of the decision-making process, the decision-maker does not possess clear and articulate notions of values that he places upon action-outcomes, one must create such notions by inducing him to consciously express a clear concept of worth for alternatives relevant to a particular decision. There are available a number of different methods that will provide some measure of utility from elicited value judgements. For each utility estimation method, the underlying model defines the conditions under which estimation is valid. The underlying premise of each model is that from a careful observation and analysis of an individual's stated preferences in comparisons between hypothetically constructed choices under controlled laboratory conditions, one may arrive at a prescriptive as well as predictive formulation of his choice behavior in real decisions from the deduced utility index.

Each method has its own assumptions regarding concepts of utility, type of decision alternatives, and capabilities of human judgement. Therefore, a classification of methods is useful to the decision-maker in helping him to decide which method is most suitable for his purposes.

The classification scheme that was devised resembles that of a decision tree. At each node in the tree the

decision-maker's choice of a set of method properties (indicated at the branch) directs him along one branch of the tree. By successively encountering a node, choosing a set of method properties, and branching, the decision-maker arrives at a method with the chosen properties. The methods appear at the very ends of the branches. An illustration of the numbering scheme is given here. Suppose that a branch is assigned the code X (for example, X is 1-1). Furthermore, suppose that the branch enters a node which branches N ways. Then each of these N branches is assigned a unique number X-I where I=1,...,N. Thus, a number code assigned to a branch will indicate which branches of the tree were its predecessors.

The decision tree is summarized by Figure 1 . A description of the tree is given in Section II.

## Section II - Classification of methods

Each of the classifying properties of the decision tree will be listed here.

### (1) Unidimensional Utility

Unidimensional utilities take into account only one factor. They are represented algebraically by U(X) and geometrically by a line in 2-dimensional space.

### (2) Multidimensional Utility

Multidimensional utility take into account the

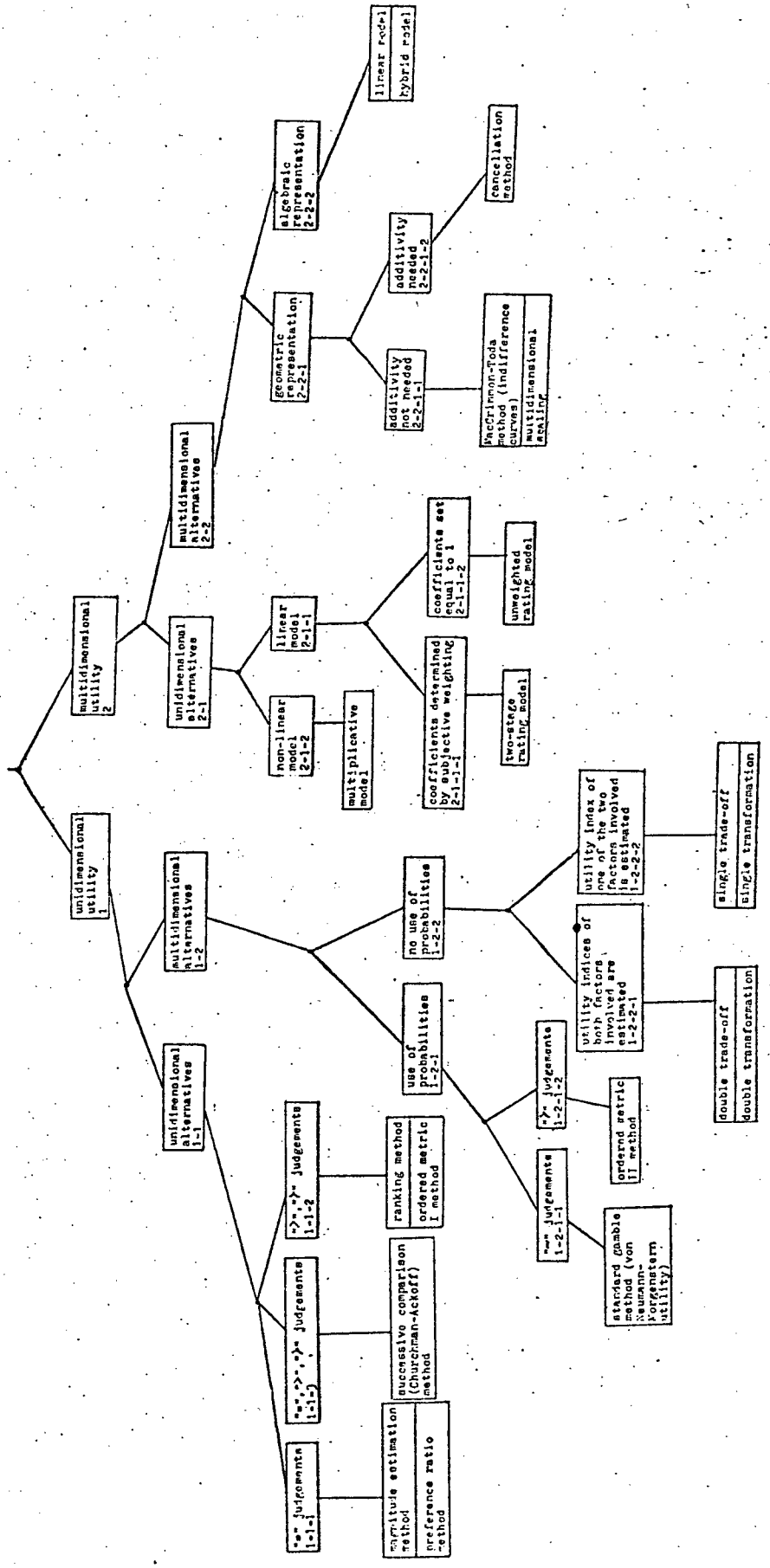


Figure 1 - Decision tree for Classification of methods

contribution of many factors. They are represented

algebraically by U(X1,...,Xn) and geometrically by a hypersurface in n+1-dimensional space. In general, multidimensional utilities are more realistic but less manageable for analysis. It will be shown later that certain assumptions can be made which will separate multidimensional utilities into several unidimensional utilities, Ui(Xi).

(1.2-1) Judgements Of Unidimensional Alternatives

The subject is required to evaluate unidimensional alternatives which can be represented geometrically as points on a line.

(1.2-2) Judgements Of Multidimensional Alternatives

The subject is required to evaluate multidimensional alternatives which can be represented geometrically as points in n-dimensional space. Evaluation consists of taking into account the contributions of several factors. Examples of multidimensional alternatives are company profiles ( (profit, prestige, number of personnel, number of branch offices) for example) and automobile performance ( (maximum speed, acceleration, stability on corners, engine size) for example). On the other hand, judgements between engine sizes of different automobiles or profits of different companies would involve only one factor; this is the case with unidimensional alternatives.

One of the difficulties expected to be encountered by presenting multidimensional alternatives to the subject is his inability to take into account simultaneously all the component factors.

Miller (1956) showed that there are limitations on the number of conceptual units that can be handled at any one time. Shepard et al (1961) has shown that learning became slower and more difficult as the number of attributes to be simultaneously judged increased. The results of DeSoto (1961) and Osgood et al (1956) revealed that subjects are unable to adequately grasp the notion of multidimensionality in making judgements. These pychological results seem to suggest that similar cognitive burden will arise in utility assessment of multidimensional alternatives.

(1-1-1) The Method Is Based Upon "Direct" Equality Judgements On Utilities

The subject is asked to assign numbers to factor levels according to his evaluation of their relative utilities.

(1-1-2) The Method Is Based Upon Preference Judgements Or "Direct" Inequality Judgements On Utilities

In the case of ranking, the subject is required to rank factor levels by preference while in the ordered metric method, the subject is asked to rank utility differences.

(1-1-3) The Method Is Based Upon "Direct" Equality Judgements On Utilities, Preference Judgements, And Direct Inequality Judgements On Utilities

(1-2-1) The Method Uses Probabilities

The subject is asked to compare between constructed alternatives which involve probability. The problems of confounding utility with probability were discussed in

Chapter 1. This confounding may not only confuse the subject but may also confuse the experimenter in his analysis. Methods using probabilities may present difficulties to subjects who do not adequately understand the probability concept. For these subjects, Gustafson et al (1972) suggest the use of two training devices, the probability wheel and the probability bar, which are both routinely used at the Stanford Research Institute.

(1-2-2) The Method Does Not Use Probabilities

(1-2-1-1) The Method Is Based Upon Indifference Judgements

(1-2-1-2) The Method Is Based Upon Preference Judgements

In general, preference judgements are less demanding than indifference judgements. Preference judgements do not generally require a precise perception of subjective magnitude while indifference judgements do. Indifference judgements assume the existence of infinite sensitivity of the subject, i.e. that the discrimination band (between preference of A over B and of B over A) is infinitesimally narrow. However this is not the case and psychophysicists were quite aware of this gradation of discrimination appearing in their empirical results (see Torgerson, 1958 for example). Mosteller and Nogee (1951) suggest that the width of the discrimination band in their utility experiment is a characteristic that distinguishes groups of people.

(1-2-2-1) The Utility Curves Of Both Factors Involved Are Estimated

(1-2-2-2) The Utility Curve Of One Of The Two Factors Involved Is Estimated

For the single trade-off and single transformation methods, the utility curve of one factor must be known beforehand.

(2-1-1) The Algebraic Expression For The Multidimensional Utility Is Based Upon A Linear Model

A type of additive utility is assumed.

(2-1-2) The Algebraic Expression For The Multidimensional Utility Is Based Upon A Non-Linear Model

The second order mixed partial derivatives do not vanish. The model allows for inclusion of interaction terms.

(2-1-1-1) The Coefficients Of The Linear Regression Equation Are Determined By Subjective Weighting

The coefficients are estimated by the subject according to his evaluation of the relative importance of each factor in determining overall utility.

(2-1-1-2) The Coefficients Are Set Equal To 1

The factors are given equal weight; i.e. it is assumed a priori that the factors are of equal importance in determining overall utility.

(2-2-1) The Multidimensional Utility Is Represented Geometrically

The utility is represented in terms of a surface or in terms of relative "distance" between points in a multidimensional space. Even though n-dimensional space for n>2 is difficult to represent on paper, the interpretation

lends itself more readily in terms of geometric rather than algebraic analogies.

(2-2-2) The Multidimensional Utility Is Represented Algebraically

The overall utility is described by an algebraic equation involving component utilities. Certain assumptions concerning aggregation of component utilities are needed.

(2-2-1-1) The Derivation Of A Utility Index Does Not Assume Additivity

It is often unrealistic to assume additivity.

(2-2-1-2) The Derivation Of A Utility Index Assumes Additivity Of Component Utilities

The additivity assumption asserts that the multidimensional utility equals the sum of its component utilities ( $U(X1,...,Xn)=U1(X1)+...+Un(Xn)$ ). This separability implies that the various components of the multidimensional alternative contribute independently to its overall utility. The additive utility model dominates the literature on riskless choice because it is a simple model. Using an additivity assumption, complex choices can be judged by judging simpler ones and analyzed by analyzing simpler ones. See Fishburn (1965a, 1965b, 1966) for further information regarding additivity.

## Section III - Discussion of methods [10]

Each of the methods which appear in the decision tree will be discussed here.

### Ranking Method

The actual application of the method requires the subject to either directly rank his preference among a list of factors or to indicate his preference between paired comparisons.

Eckenrode (1965) used various modifications of the ranking method as part of a study for comparison of utility assessment methods. In addition to the ranking method, the partial paired comparisons I, the partial paired comparisons II, and the complete paired comparisons method (the terminology is that used by Eckenrode) were used. In the partial paired comparisons I method, the factors are represented on the rows and columns of an upper triangular matrix. The subject is asked to indicate in each matrix entry, the more preferred of the pair of factors which are the coordinates of the matrix entry. Buel (1960) used this format for paired comparisons. In this format, each factor is paired only once with every other factor. In the partial paired comparisons II method, the subject is presented with

---

[10] The standard gamble method for derivation of von Neumann-Morgenstern utility and the MacCrimmon-Toda method for derivation of indifference curves will be discussed in Chapter 3.

a list of factor pairs. He is then asked to circle the more preferred factor of each pair. Again each factor is paired only once with every other factor. The complete paired comparisons method has the same format as the partial paired comparisons II method. However, each factor pair appears twice, once in the order A-B and once in the order B-A.

In presenting a list of paired comparisons, it seems common sense to arrange the pairs in such a fashion so that pairs involving a particular factor should be as far apart as possible and that no particular factor should appear preponderantly in one position. Phillips (1964) has generated tables of orders for stimulus pairs based upon n stimuli (n=3,...,15). His table of paired comparisons is based upon the works of Ross (1934) and Wherry (1938) and yields (1) the maximum possible spacing between pairs involving the same stimulus, and which ensure (2) that every stimulus appears an equal number of times in the first and second positions but there is to be the minimum possible number of pairs which have the same stimulus occurring either first in a pair or second in a pair, twice running (Wherry, 1938 described this undesirable feature as "space error").

A problem which may be encountered in the paired comparison method is the violation of the transitivity axiom. If the subject has firm belief in this axiom, he will consider the inconsistency as a slip in judgement and will accordingly rectify the rank. However, if the subject has

doubts about the validity of the axiom and refuses to rectify the inconsistency, the evaluated preferences may have limited value. In a study cited by Goode and Hatt (1952), 150 adults were asked to pairwise compare a list of five occupations. Out of this test sample, less than 9 percent expressed intransitive comparisons.

Suzuki (1957) obtained preference rankings of naval equipment models from naval officers. Beach (1972) obtained preference rankings of college courses from subjects.

## Characteristics

The ranking method is very simple to apply and demands very little from the subject.

The method of paired comparisons may be advantageous over direct ranking a list in reducing the number of factors per judgement. With direct ranking of a list of all factors, the large number of factors per judgement may be a strain upon the cognitive abilities of the subject. With a large number of factors, the subject may be unable to perceive adequately all the factors involved. However, for the method of paired comparisons, the process of evaluation may be too laborious and time-consuming for the impatient subject (using Phillips' table, 15 paired comparisons are needed for 6 factors, 21 for 7 factors, 28 for 8 factors, and so on).

The resulting scale is ordinal. Despite the weakness of the resulting scale, the ranking method is useful when

augmented with other methods which yield a measurement scale that is of higher order than ordinal. In fact, the ranking procedure is inherent, if not explicitly apparent, in other utility measurement methods.

Magnitude Estimation Method

The subject indicates his assessment of the relative worth of a factor by assigning its utility to a position on a linear subjective scale. The subjective scale is normally anchored at two points. The usual procedure for anchoring is to choose the least and the most preferred factors to be at the extreme ends of the scale. The unit and the origin for the subjective scale are arbitrary.

Torgerson (1958) employs a two-way classification of different empirical procedures used in obtaining measurements by the direct magnitude estimation technique. Four procedures result from pairwise combinations of single stimulus or multiple stimuli with limited categories or limited categories.

In a typical single stimulus presentation, several factors are presented to the subject one at a time in random order. The subject estimates the utility of a factor after the presentation of each factor. In the multiple stimuli presentation, all the factors are available and each factor is allowed to influence directly the judgement of each other factor. In the case of single stimulus, it is likely that the subject will change his reference points (which he bases

his judgements upon) as he progresses through the series of factors. However, in the case of multiple stimuli the subject is instructed to rearrange the utilities so that there is less possibility that different reference points will be used for different factors. However, there are the usual advantages of the single stimulus presentation over the multiple stimuli presentation as was discussed in regards to the ranking method.

Limited-category scales provide subjective scales of discrete categories while unlimited-category scales provide continuous scales. Limited-category scales with odd number of categories allow the possibility of indicating neutral preference while even number of categories do not. At selected points on either scale, annotations may be used to help the subject to quantify his subjective feelings (eg. very desirable, desirable, etc.). However this leaves open the danger of misinterpretation of the annotations. In psychophysics experiments, Stevens (1959) has shown that subjects apparently do not judge equal interval categories of limited category scales to be equal; the results indicated that categories at one end of the scale tended to include a greater interval of subjective magnitude than at the other end. Steven's explanation is that the subject's sensitivity is not uniform over the subjective scale. A given difference may seem less impressive at the other end. If the asymmetry in the subject's sensitivity is too significant, it is incorrect to assume an underlying linear subjective scale. This psychophysical phenomenon suggests

that similar nonlinearity may exist in utility scales.

Using the magnitude estimation method, the utilities of the following entities have been empirically assessed: job attributes (Vroom, 1971), hospital ward conditions (Huber,Sahney, and Ford,1969), college courses (Beach, 1972).

## Characteristics

The method demands greater sensitivity from the subject than the ranking method. In order to help the subject to quantitatively rate the factors, the subject may initially be asked to rank them.

The assumed linearity of the subjective scale may not be valid.

## Preference Ratio Method

In this method, it is assumed that the subject is capable of directly perceiving the ratio between two subjective magnitudes. Having assigned the utility of one factor level arbitrarily to specify the unit, the utilities of the remaining factor levels may be determined by the empirically derived ratios. Metfessel (1947) proposed the constant-sum method which was an alternative form of expressing fraction ratios. The subject is instructed to divide, say 100 units, between two selected factors according to their relative utilities. For example, the

subject may decide that one factor should be assigned 30 out of 100 units more than the other. In this case, he would distribute 35-65 and the corresponding ratio would be 35/65.

If the subject is not capable of reporting ratios in general, he may still be capable of making valid bisection (i.e. half or double ratios or Metfessel's 50-50) judgements. In the case of bisection judgements, the requirements placed upon the subject are considerably less than in the case of general ratio judgements.

The equisection method is a method employing the repeated application of bisection judgements. The task of the subject is to select n-1 of the remaining factors (after the least and most preferred factors have been fixed to the endpoints) so that the n-1 associated utilities divide the utility interval between the two endpoints into n equal intervals.

For the preference ratio method, Torgerson (1958) further classifies the method into direct-estimate and prescribed-ratio methods. The essential difference is that in the direct-estimate method, the subject is presented with two factor levels for which he is to provide a subjective estimate of the ratio of their utilities while in the prescribed-ratio method, the subject is asked to report the factor level for which the ratio of its utility to that of the utility of the standard is equal to a ratio prescribed by the experimenter.

In the prescribed-ratio method, the subject must be able to choose any factor level between the two endpoints; therefore the factor levels must be continuous. Utility values may be obtained only for certain factor levels. For the prescribed ratio of 1/m, only those factors whose utilities equal k•[m to the ath power] for a = 1,2,3,... (k is an arbitrarily assigned unit of utility) may be assessed.

On the other hand, the factor levels need not be continuous for evaluation by the direct-estimate method. In addition, all factor levels of interest to the assessor may be evaluated. With n factor levels, n-1 ratios (all compared to the standard) are obtained. However, the method may also be extended to pairwise comparison; Comrey (1950) proposed that all the factor levels (to be assessed) serve in turn as standards, thus giving n(n-1)/2 ratios from which a utility scale can be derived (see Comrey, 1950 and Torgerson, 1958 for calculation formulas). Comrey's procedure consists of reporting comparative judgements in the same manner as advocated by Metfessel (1947).

Klahr (1969) used Comrey's procedure in obtaining utility estimates of college admission attributes. Galanter (1962) used preference ratios to obtain a utility scale for money. Beach (1972) had subjects make preference ratio judgements in order to derive utility scales for college courses. Dudek and Baker (1956) obtained utility scales for neckties through preference ratio judgements of subjects.

## Characteristics

This method demands greater sensitivity from the subject than the magnitude estimation method. Some subjects may feel uncomfortable about conceptualizing in terms of ratios. Naive subjects may have to be trained in the concept of ratios. Reported judgements may be severely limited to certain easily interpretable or commonly used ratios.

The procedure for deriving a utility index from ratio judgements is very sensitive to errors in judgement. Errors will be multiplied by estimated ratios.

## Ordered Metric I Method

The term "ordered metric" is attributed to Coombs (1950). In deriving Coomb's direct ordered metric, the factor levels are first ranked and then the increments between adjacent utilities are ranked. For example, if the subject can not only state that his preferences are A to B and B to C but also state that his preferences of A to B is greater than his preference of B to C, then his utility index is represented by a triplet of numbers which satisfy:

$$U(A)>U(B)>U(C) \text{ and } U(A)-U(B)>U(B)-U(C).$$

Suppes and Winet (1955) suggest that from an alternative construction of choices and an assumption of additivity, an ordered metric along one dimension may also be obtained. For example, suppose A>B>C. The subject is then hypothetically given A and B. Next he is required to choose

between trading B for C, or A for D. If he trades B for C,

$$U(A \text{ and } C) > U(D \text{ and } B)$$

$$U1(A) + U1(C) > U1(D) + U1(B) \text{ by additivity}$$

$$U1(A) - U1(B) > U1(D) - U1(C)$$

Siegel (1956) extends Coomb's derivation of a direct ordered metric by devising a procedure for deriving a higher ordered metric scale in which all possible combinations of contiguous increments between adjacent utilities are ordered. Siegel proposes his maximin rule which provides an efficient means of eliciting judgements. His maximin rule is a rule which maximizes the amount of information (necessary to construct the scale) from a minimum number of reported judgements. The rule dictates the order in which the judgements should be examined in the derivation of the scale (see Hurst and Siegel, 1956).

Becker and Siegel (1962) obtained a utility scale of college grades from a higher ordered metric scale.

## Characteristics

The subject may find it too difficult to compare between differences in preferences. These choices are not commonly made and may seem artificial to the subject.

The ordered metric places bounds on the location of each factor on the utility scale. The bounds become more restrictive as more ordered metric relationships are obtained.

Ordered Metric II Method

Siegel (1956) has reported a method of obtaining a higher ordered metric from interrogation of a subject about his preference ordering of gambles. The subject is first required to rank the factors. The resulting ordinal scale gives us information about the preferences between factors but nothing about the relative magnitudes of factors. However, by further requiring the subject to rank his preferences between constructed 50-50 gambles, one may derive an ordered metric (or higher ordered metric) by use of the von Neumann-Morgenstern theorem on maximization of expected utility. For example[11],

$$(A,D;1/2) > (B,C;1/2) \Rightarrow$$

$$1/2\,(U(A)+U(D)) > 1/2\,(U(B)+U(C)) \Rightarrow U(A)-U(B) > U(C)-U(D).$$

The initial ranking of certain outcomes automatically gives us information about the ordering of some gambles (called the orderable pairs) but it is the ordinal relations between non-orderable pairs of gambles which contain the information necessary to change an ordinal scale to a higher-ordered scale. However, not all the non-orderable relations must be found in order to obtain the information necessary for higher-ordered metric scaling (maximin rule).

Farrer (1964) found the predictive ability of higher ordered metric scales (constructed for cigarettes) to be

---

[11] (A,B;p) represents a wager whose outcome is A with probability p or B with probability 1-p.

quite stable over a one month period.

Using the von Neumann-Morgenstern theorem, ordered metric scales of utility have been obtained for money (Coombs and Komorita, 1958) and various appliances (Coombs and Beardslee, 1954) while higher ordered metric scales have been obtained for books (Siegel, 1956) and cigarettes (Hurst and Siegel, 1956; Farrer, 1964).

## Characteristics

The use of gambles instead of certain outcomes in constructing a higher ordered metric presents an additional cue to be considered in the subject's judgement, namely that of probability. However, most subjects should have no problem in conceptualizing probabilities of 1/2 (the analogy of heads and tails is usually used).

The ordered metric places bounds on the location of each factor on the utility scale. The bounds become more restrictive as more ordinal relationships between gambles are found (i.e. the more ordered metric relationships that are obtained).

Successive Comparisons Method (Churchman-Ackoff Method)

In this method, it is assumed that the factors involved are utility independent and unidimensional. The method consists of successively comparing a factor (or level) with a group of factors (or factor levels) . At each comparison, the subject is required to assign tentative numbers to utilities of factors such that they are compatible with the currently expressed preferences of the subject. For each of the remaining comparisons, the previously assigned utilities are refined according to the currently expressed preference. Hopefully, the assigned utilities will converge during the latter comparisons. Churchman and Ackoff (1953) give one possible way of using the successive comparisons method. Briefly stated,

(1) Rank the factors. Let X1>X2>X3>...>Xn.

(2) Assign U(X1)=1.00 and numbers to U(X2),...,U(Xn) according to their relative preferences.

(3) Compare X1 with X2+X3+...Xn. [12]

(3i) If X1 >,=,< X2+X2+...+Xn, adjust U(X2),U(X3),...,U(Xn) so that U(X1)=1.00 >,=,< U(X2)+U(X3)+...+U(Xn).

(4) Compare X1 with X2+X3+...+Xn-1.

Adjust U(X2),...,U(Xn-1) according to expressed preference.

(5) Compare X1 with X2+X3+...+Xi where i=n-1(-1)3. Adjust U(X2),U(X3),...,U(Xi) accordingly.

(6) Continue until Xn-2 is compared with Xn-1+Xn.

The final refined values of U(X1),U(X2),...,U(Xn) form the

---

[12] The "+" is a logical "and" and not an algebraic "plus".

utility index for the factors.

Churchman and Ackoff (1953) cite various actual applications of this method in quality control and corporate planning.

## Characteristics

The method consists of successive refinement of estimates. The assigned utilities may not converge in which case the assignment of utilities is too erratic to be useful.

Furthermore, the assumption of utility independence among the factor levels involved may be too artificial to satisfy any real situation.

As an alternative (to the one previously cited) variation of the method, the comparisons for judgement may be randomly chosen as suggested by Churchman and Ackoff (1953).

## Single Trade-off

An indifference curve is estimated for two factors by finding points X and Y such that $(X,Y) \sim (Xr,Yr)$ where Xr and Yr are the reference points. From a knowledge of one of the two utility curves, one can estimate the other utility curve. Suppose that we have an estimate of U1(X). Then, from the additivity assumption,

$U(X,Y)=U1(X)+U2(Y)=U1(Xr)+U2(Yr)=c$ say, for (X,Y) on the

indifference curve.

Therefore U2(Y)=c-U1(X) where c is an arbitrary constant. c may be assigned arbitrarily while preserving interval properties.

## Characteristics

A utility curve of one of the two factors must be estimated previously by another method. From this given utility curve, the other factor's utility curve may be estimated from the trade-off or indifference curve.

The additivity assumption is needed.

## Double Trade-off

Two indifference curves, relatively close to each other, are estimated by finding points X and Y such that $(X,Y)\sim(Xr1,Yr1)$ and $(X,Y)\sim(Xr2,Yr2)$ for the reference points Xr1,Yr1,Xr2,Yr2. Figure 2 illustrates the two curves. A "flight of stairs" is then drawn between the two curves by a connected series of horizontal and vertical line segments

Indifference curve 1:

(1) U(X1,Y1)=d say

(2) U(X2,Y2)=d

(3) U(X3,Y3)=d

Indifference curve 2:

(4) U(X2,Y1)=e say

(5) U(X3,Y2)=e

(6) U(X4,Y3)=e

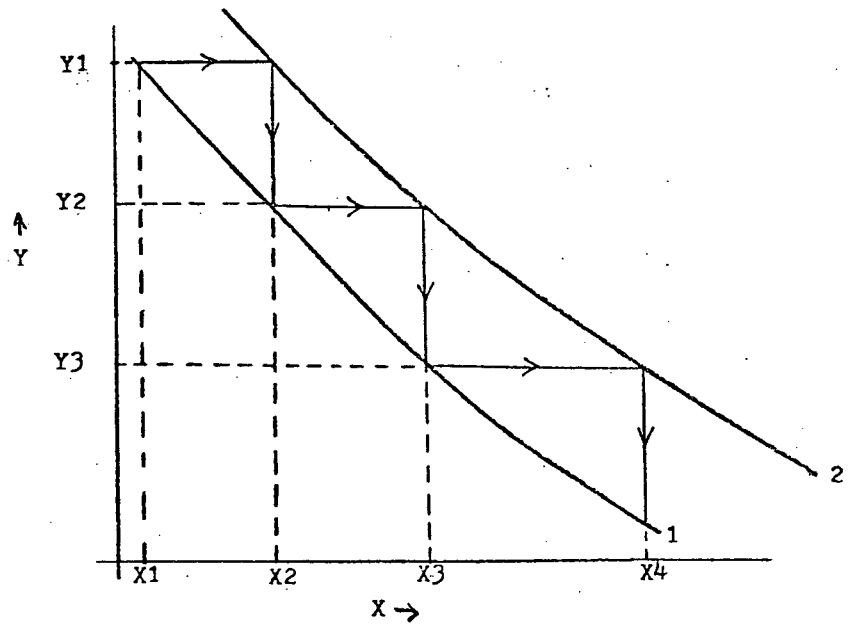


Figure 2 - Figure for Double Trade-off Method

Subtracting (4) from (1):

$U(X1,Y1)-U(X2,Y1)=d-e=c$ say

Then $U1(X1)-U1(X2)=c$ by additivity

Similarly: $U1(X2)-U1(X3)=c$ and $U1(X3)-U1(X4)=c$

Subtracting (5) from (1):

$U(X1,Y1)-U(X3,Y2)=d-e=c$

Then $U2(Y1)-U2(Y2)=c-(U1(X1)-U1(X3))$ by additivity

$=c-(2c)=-c$

Similarly: $U2(Y2)-U2(Y3)=-c$

Thus, the successive points on the indifference curves touched by the stairs define equal increments of utility for

each factor. By assigning an arbitrary number to the utility of a certain level of one factor and also assigning an arbitrary number to the increment of utility (c) for that factor (these arbitrary assignments will preserve interval properties) the utility curve for that factor may be estimated. The same procedure is followed for obtaining the other factor's utility curve.

Characteristics

From two indifference curves, certain levels of both factors which define constant adjacent increments of utility may be estimated. From these identified factor levels, piecewise linear representation of each factor's utility curve may be used as an approximation to the utility curve.

The closer together the two indifference curves are, the more the number of points on each utility curve that may be identified. However, the closer together the two indifference curves which the subject tries to estimate, the greater the chance of inconsistency (i.e. crossing of two indifference curves) occurring.

For the case of indifference curves that are double valued (this may occur for utility functions that are not monotonic with respect to the factor level) as in Figure 3 , the procedure of sketching a flight of stairs may be tricky or inappropriate.

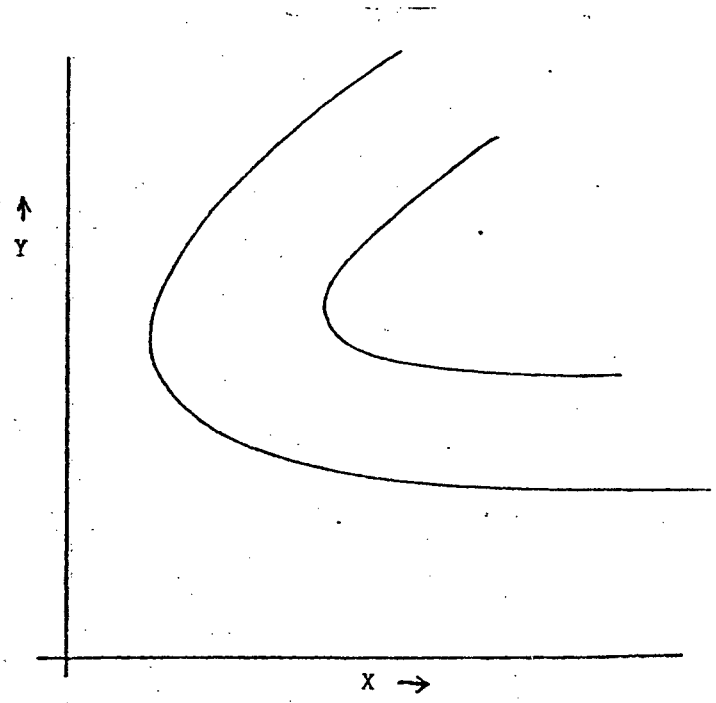


Figure 3 - Double Valued Indifference Curves

## Single Transformation

A transformation curve is constructed by estimating points (X,Y) for which (X,Yr)~(Xr,Y) where Xr and Yr are reference points. One possible procedure is to estimate Y for successive arbitrarily fixed X's, and then to estimate X for successive arbitrarily fixed Y's. From a given utility curve for one of the two factors, one can estimate the other factor's utility curve. Suppose that we have an estimate of U1(X). Then, from the additivity assumption,

$$U(X,Yr)=U1(X)+U2(Yr)$$

$U(X,Yr)=U(Xr,Y)=U1(Xr)+U2(Y)$ for points (X,Y) on the transformation curve.

Then, $U2(Y)=U1(X)+(U2(Yr)-U1(Xr))=U1(X)-c$ where c is an arbitrary constant. c may be assigned arbitrarily while preserving interval properties.

## Characteristics

A utility curve of one of the two factors must be estimated previously by another method. From this given utility curve, the other factor's utility curve may be estimated from the transformation curve.

The additivity assumption is needed.

The transformation curve is not as readily interpretable as the indifference curve.

Double Transformation

Two transformation curves, relatively close to each other, are estimated by finding X and Y such that $(X,Yr1)\sim(Xr1,Y)$ and $(X,Yr1)\sim(Xr2,Y)$ for the reference points Xr1, Yr1, Xr2. Figure 4 illustrates the two curves. A

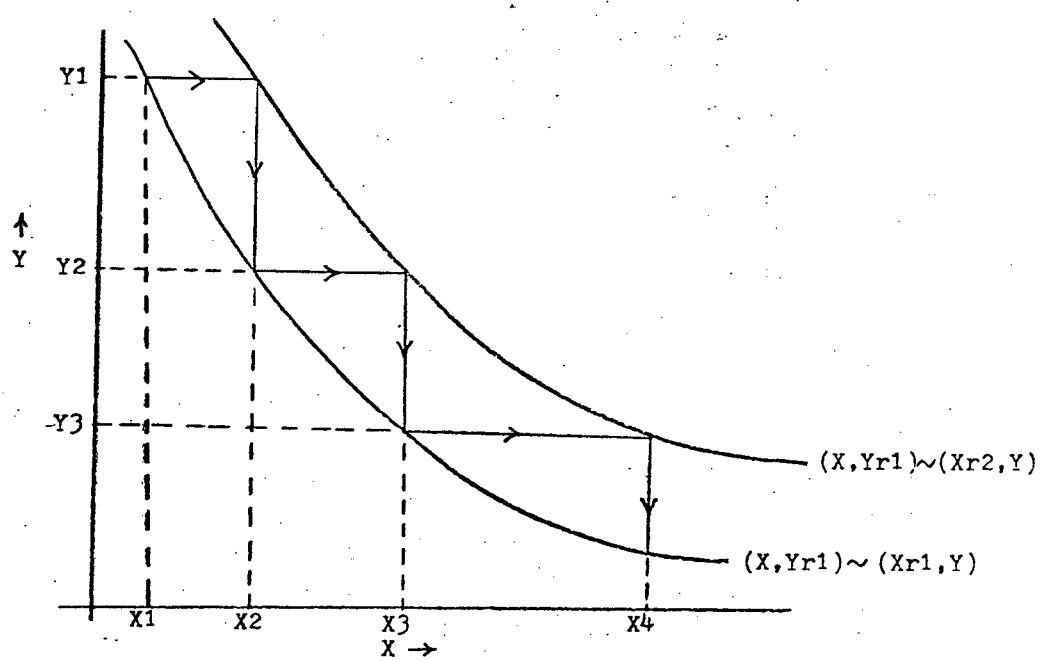


Figure 4 - Figure for Double Transformation Method

"flight of stairs" is then drawn between the two curves by a connected series of horizontal and vertical line segments. For the $(X,Yr1)\sim(Xr1,Y)$ transformation curve:

(1) $U(X1,Yr1)=U(Xr1,Y1)$

(2) $U(X2,Yr1)=U(Xr1,Y2)$

(3) $U(X3,Yr1)=U(Xr1,Y3)$

For the $(X,Yr1)\sim(Xr2,Y)$ transformation curve:

(4) $U(X2,Yr1)=U(Xr2,Y1)$

(5) $U(X3,Yr1)=U(Xr2,Y2)$

(6) $U(X4,Yr1)=U(Xr2,Y3)$

By additivity,

(1) $U(X1,Yr1)=U1(X1)+U2(Yr1)=U(Xr1,Y1)=U1(Xr1)+U2(Y1)$

or $U1(X1)+U2(Yr1)=U1(Xr1)+U2(Y1)$

By additivity,

(4) $U(X2,Yr1)=U1(X2)+U2(Yr1)=U(Xr2,Y1)=U1(Xr2)+U2(Y1)$

or $U1(X2)+U2(Yr1)=U1(Xr2)+U2(Y1)$

Subtracting (4) from (1):

$U1(X1)-U1(X2)=U1(Xr1)-U1(Xr2)=c$ say

Similarly: $U1(X2)-U1(X3)=c$ and $U1(X3)-U1(X4)=c$

By additivity, (5) $U1(X3)+U2(Yr1)=U1(Xr2)+U2(Y2)$

Subtracting (5) from (1):

$$U1(X1)-U1(X3)=U1(Xr1)-U1(Xr2)+U2(Y1)-U2(Y2)$$

$$2c=c+U2(Y1)-U2(Y2)$$

$$U2(Y1)-U2(Y2)=c$$

Similarly: $U2(Y2)-U2(Y3)=c$

Thus, the successive points on the transformation curves touched by the stairs define constant adjacent increments of utility for each factor. By assigning an arbitrary number to the utility of a certain level of one factor and also assigning an arbitrary number to the increment of utility for that factor (these arbitrary assignments will preserve interval properties), the utility

curve for that factor may be estimated. The same procedure is followed to obtain the other factor's utility curve.

Characteristics

From two transformation curves, certain levels of both factors which define constant adjacent increments of utility may be estimated. From these identified factor levels, a piecewise linear representation of each factor's utility curve may be used to approximate the utility curve.

For the case of transformation curves that are double valued as in Figure 5 , the procedure of sketching a flight of stairs may be tricky or inappropriate.

The additivity assumption is needed.

The transformation curve is not as readily interpretable as the indifference curve.

Cancellation Method

The method consists of deriving two indifference curves from which a third one may be estimated. The method assumes additivity of component utilities and is illustrated in Figure 6 .

Two "flight of stairs" are sketched between two derived indifference curves.

From indifference curve 1:

$$U(X1,Y2)=U(X2,Y3)$$

$$U1(X1)+U2(Y2)=U1(X2)+U2(Y3) \text{ by additivity } (1)$$

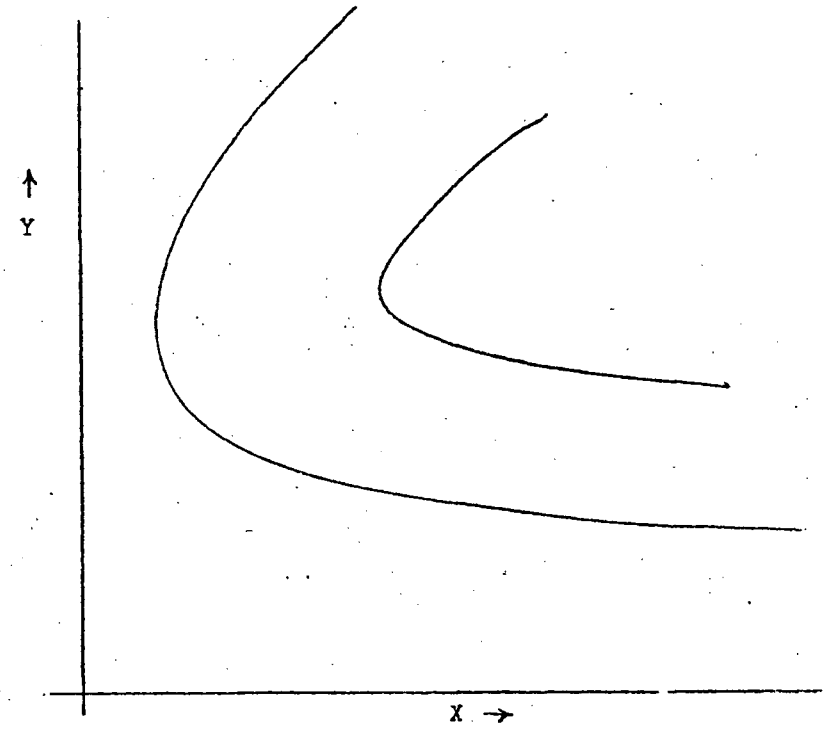


Figure 5 - Double Valued Transformation Curves

From indifference curve 2:

$$U(X2,Y1)=U(X3,Y2)$$

$$U1(X2)+U2(Y1)=U1(X3)+U2(Y2) \text{ by additivity } (2)$$

Adding together (1) and (2):

$$U1(X1)+U2(Y2)+U1(X2)+U2(Y1)=U1(X2)+U2(Y3)+U1(X3)+U2(Y2)$$

The deletion of U1(X2) and U2(Y2) from both sides of the equality sign (this is the double cancellation property) results in:

$$U1(X1)+U2(Y1)=U1(X3)+U2(Y3)$$

$$U(X1,Y1)=U(X3,Y3)$$

$$(X1,Y1)\sim(X3,Y3)$$

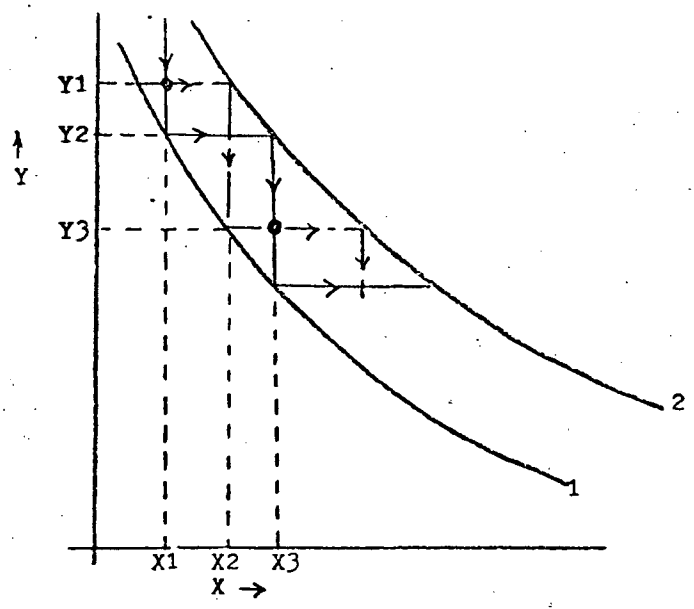


Figure 6 - Figure for Cancellation Method

It was shown from the above arguments that if (X1,Y2) and (X2,Y3) lie on the same indifference curve and (X2,Y1) and (X3,Y2) lie on the same indifference curve, then (X1,Y1) and (X3,Y3) lie on the same indifference curve. This relationship is formally given as the Thomsen condition (see Krantz et al, 1971, page 250).

By parallel arguments, it can be shown that a triple cancellation property exists (Reidmeister condition) for which three pairs of given related points on three indifference curves generate additional indifference points.

Higher order cancellation properties are also possible but involve complicated formulas for derivation of additional points.

## Characteristics

The closer together the two derived indifference curves are, the more number of points on the third curve that may be estimated.

The number of generated indifference points may be too few for adequately curve-fitting an additional indifference curve.

## Two-stage Rating Model

A linear model for overall utility is developed from a two-stage process of collecting data. The linear model is:

$$U(X1,\ldots,Xn)=W1\cdot U1(X1)+\ldots+Wi\cdot Ui(Xi)+\ldots Wn\cdot Un(Xn)$$

(1) The subject is required to estimate Ui(Xi).

(2) The subject is required to assign numbers to Wi according to his evaluation of each factor's relative importance in determining the overall utility. The procedure usually used for assigning numbers to Wi is the magnitude estimation method.

Studies (Sarbin, 1942; Smedslund, 1955) have shown that there are noticeable differences between subjective weights Wi and those determined "optimally" by fitting the proposed

linear regression model to the data. Shepard (1962) offers a plausible explanation by suggesting that subjective weighting involves comparing factors which occupy different dimensions. His analogy concerning the ease of comparing two still-life paintings relative to the difficulty of comparing a painting with an abstract sculpture is most appropriate.

In Hoepfl and Huber's study (1970), evaluation of teaching ability of hypothetical professors each described by six factors (i.e. n=6) were obtained.

Characteristics

A multidimensional utility is decomposed into its component utilities. A benefit of this decomposition is that deriving utilities along one dimension demands less from the subject than along several dimensions simultaneously.

Validation of the model may be accomplished by collecting subjective estimates of the overall utility and comparing these estimates with those predicted by the two-stage rating model (degree of agreement is indicated by the correlation coefficient).

This a a self-explicated model (a term used by Hoepfl and Huber, 1970) because the parameters Wi and Ui(Xi) are explicitly estimated by the decision-maker.

If n is large, the number of evaluations needed from the subject may be too large for practical application of the model.

## Unweighted Rating Model

A linear model for overall utility is used. The model equation is identical to the two-stage rating model equation except that the subjective weights Wi are set equal to 1. The model equation is:

$$U(X1,...Xn)=U1(X1)+...+Un(Xn)$$

The subject is required to estimate Ui(Xi). The method aviods the difficulty of assigning weights to the factors.

## Characteristics

A multidimensional utility is decomposed into its component utilities. The weights for the factors are assumed a priori to be equal in magnitude, i.e. it is asssumed that the component factors have equal importance in determining overall utility.

Validation of the model may be accomplished by collecting subjective estimates of the overall utility and comparing these estimates with those predicted by the two-stage rating model (degree of agreement may be indicated by the correlation coefficient).

If n is large, the number of evaluations needed from the subject may be too large for practical application of the method.

## Linear Model

The method consists of collecting multidimensional utility judgements, deriving model parameters from these judgements, and then predicting future multidimensional utility judgements from the model. The model used is a linear regression model with dummy variables (see Suits, 1957 regarding dummy variables).

$$U(X1,\ldots,Xn)=Uo+U11\bullet X11+\ldots+Uij\bullet Xij+\ldots+Unl\bullet Xnl$$

where $Xij=0$ if the multidimensional attibute in question does not possess the ith factor at the jth level, and $Xij=1$ otherwise.

$U(X1,\ldots,Xn)$ are estimated by the subject for various levels of the component factors. The usual procedure is to require the subject to rate multidimensional alternatives on a magnitude estimation scale. A linear regression is applied to the model equation to obtain "optimal" (in a least squares sense) estimates of Uo and Uij. Future predictions may be made from this model with the derived parameters.

## Characteristics

The subject is required to make utility estimates on multidimensional alternatives. The linear model proposes a decomposition of multidimensional utility into component parameters which are estimated by regression techniques.

The subject may be required to make many

multidimensional judgements before the regression results can be considered statistically significant.

The model differs from the two-stage rating model and the unweighted rating model in that the application of regression reveals which components in the model do not make a significant contribution to the overall utility. In the case of multiple regression, the F statistic associated with every estimated parameter gives an indication as to whether one should reject the null hypothesis that Uij=0 at the specified significance level. In the case of stepwise regression, the final equation of the stepwise iterations will contain only the statistically significant component parameters.

Validation of the model is indicated by $R^2$, the multiple correlation coefficient.

Multiplicative Model

A non-linear configural model for overall utility is used. The model used is:

$$U_o = U_o \prod_{i=1}^{n} U_i(X_i)^{A_i}$$

The model equation may be transformed to:

$$\ln(U(X1,\ldots,Xn) = \ln Uo + A1 \cdot \ln U1(X1) + \ldots + An \cdot \ln Un(Xn)$$ [13]

Huber, Sahney, and Ford (1969) have suggested that such a model might more nearly represent the form of the subject's

---

[13] lnx=natural logarithm of x

actual utility model than does the form of the previously discussed additive models especially in those cases in which some factors essentially act as screening factors.

The subject is required to estimate Ui(Xi) and U(X1,...,Xn) for selected component levels. The usual procedure is to require the subject to rate multidimensional alternatives on a magnitude estimation scale. The parameters Uo and Ai are estimated by the application of regression techniques to the transformed model equation.

Characteristics

A multidimensional utility is decomposed non-linearly into its component utilities. The subject is required to make multidimensional as well as unidimensional utility judgements. The parameters are estimated by regression techniques.

The subject may be required to make many multidimensional judgements before the regression results can be considered to be of statistical significance.

This model differs from the two-stage rating model and the unweighted rating model in that the application of regression can reveal which components in the model do not make a significant contribution to the overall utility. In the case of multiple linear regression, the F statistic associated with every estimated parameter gives an indication as to whether one should reject the null

hypothesis that Aij=0 at the specified significance level. In the case of stepwise regression, the final equation of the stepwise iterations will contain only the statistically significant component parameters.

## Hybrid Model

The method consists of collecting multidimensional and unidimensional judgements, deriving model parameters from these judgements, and then predicting future multidimensional utility judgements from the model. The model used is a linear regression model:

$$U(X1,...,Xn)=Uo+W1 \cdot U1(X1)+...+Wn \cdot Un(Xn)$$

The subject is required to estimate Ui(Xi) as well as U(X1,...,Xn) for various levels of the component factors.

A linear regression is applied to the model equation to obtain "optimal" (in a least squares sense) estimates of Uo and Wi. Future predictions may be made from this model with the desired parameters.

In Hoepfl and Huber's study (1970), the hybrid model was used to describe the evaluation of teaching ability of hypothetical professors, each described by six factors (i.e. n=6).

## Characteristics

The subject may be required to make many multidimensional judgements before the regression results

can be considered statistically significant.

The model equation differs from that of the two-stage rating model equation in that the coeficients for the hybrid model are determined objectively, i.e. by regression techniques.

In the case of multiple regression, the F statistic associated with every estimated coefficient gives an indication as to whether one should reject the null hypothesis that Wi=0 at the specified significance level. In the case of stepwise regression, the final equation of the stepwise iterations will contain only the statistically significant component parameters.

Validation of the model is indicated by $R^2$, the multiple correlation coefficient.

Multidimensional Scaling Method

The subject's preference ordering is represented in terms of "distance" in a multidimensional space.

The subject is required to make judgements about the similarities of pairs of multidimensional alternatives (for n distinct alternatives, n(n-1)/2 similarity judgements can be obtained). Messick (1956) suggests an empirical procedure for obtaining similarity judgements. In addition to similarity judgements, the subject is required to designate his most preferred alternative (ideal point). The method postulates that from a set of similarity judgements, a

spatial configuration can be constructed in which the alternatives are arranged such that the inverse rank order of interpoint Euclidean distances in the space corresponds to the rank order of similarities given in the input data (i.e. pairs of more similar alternatives are "closer" together). From this postulation, the preference ordering of the alternatives is directly related to the ordering of Euclidean distances in the space from the ideal point to each alternative, i.e. A is preferred to B if the Euclidean distance between A and the ideal point is less than that between B and the ideal point. The search for a spatial representation involves a tradeoff between: (1) maximizing the inverse correlation of interpoint distances rank and similarity measures rank by increasing the spatial dimension, and (2) achieving a more parismonious representation of the data by decreasing the spatial dimension.

Klahr (1969) has shown that a spatial configuration constructed from similarity judgements obtained from college admission officers was accurate in predicting the officers' preferences among college apllicants.

## Characteristics

If many multidimensional alternatives are to be used in constructing a spatial configuration, the resulting number of similarity judgements may be too large for practical application.

A preference ordering is easily interpretable in geometric terms.

The location of the ideal point is crucial in accurately determining the preference ordering.

The search for a spatial representation requires very long and complicated calculations. The experimenter must have access to an appropriate computer program for performing the calculations or be able to write such a program himself.

# Chapter 3 - Cardinal Utility Curves & Indifference Curves

## Section I - von Neumann-Morgenstern Cardinal Utility

### Introduction

The method for deriving a von Neumann-Morgenstern cardinal utility index (the method is sometimes called the standard gamble method) consists of presenting hypothetically constructed gambles to the subject from which a cardinal utility may be derived from his expressed preferences.

The method is based upon the maximization of expected utility which is the crucial theoretical result arising from the von Neumann-Morgenstern assumptions. However, previous to von Neumann and Morgenstern's work, traditional mathematical treatment of risky situations had proposed the notion of maximization of expected value[14]. Such a notion does not seem plausible in light of the fact that people buy insurance and lottery tickets. Premiums are more than the expected monetary gains from an insurance policy; ticket prices are more than the expected monetary wins from lottery draws. Furthermore,a faulty fundamental assumption inherent in this notion was recognized by Daniel Bernoulli (1738).

---

[14] The expected value of a money gamble is the sum of the products of monetary value of outcomes with their associated probabilities.

One of the implications of expected value maximization is that two persons will have identical preference towards the same risky situation since the expected value of the risky situation is identical for both persons. This implication is contradictory to our everyday experience. In order to circumvent this impasse and to resolve the insurance paradox, Bernoulli proposed that people act so as to maximize expected utility rather than expected value.

Although Bernoulli is credited with being the first to advocate the notion of expected utility maximization, it was not until 1944 when von Neumann and Morgenstern's book, Theory of Games and Economic Behavior , was published that a formal basis was provided for such a concept. In fact, they asserted that "it can be shown that under the conditions on which the indifference curve analysis is based very little extra effort is needed to reach a numerical utility" (1944, page 17). While the ability of a rational person to order preferences for certain outcomes is needed for the ordinal theory, the extra effort required for the cardinal theory is that he be able to also order probability combinations of outcomes. For example, suppose that a person expresses indifference between the certain outcome of $8.00 (which may be thought of as a gamble which will result in the outcome of $8.00 with probability 1 or any other outcome with probability 0) and a gamble in which he is offered a 40-60 chance of winning $10.00 or $0.00. From his expressed indifference, the von Neumann-Morgenstern theory implies that the utility of $8.00 and the expected utility of the

40-60 gamble are identical,i.e.

$$U(\$8.00)=0.40\bullet u(\$10.00)+0.60\bullet u(\$0.00)$$

Assigning $u(\$10.00)=1$ and $u(\$0.00)=0$ (two arbitrary assignments will preserve the properties of the cardinal utility scale since it is measurable up to a linear transformation; for details see Chapter 1),

$$u(\$8.00)=0.40(1)+0.60(0)=0.40$$

By varying the probabilities in the gamble, the utility for various monetary levels between \$0.00 and \$10.00 may be determined.

Such a concept as expected utility maximization is certainly appealing since it is not only intuitively plausible but also simple to comprehend as was demonstrated by the previous example. The essential difference between the von Neumann-Morgenstern presentation of the proposition and the previous suggestions of such a maximization concept is that their presentation demonstrated that the notion of expected utility maximization was logically equivalent to the acceptance of certain basic axioms, which in themselves seemed to be reasonable assumptions about human behavior. The importance of developing the theory axiomatically is that if the axioms have empirical validity, the empirical meaning of the theory is much more significant than if the theory's results are stated without justification. Therefore, the axiomatic system upon which the von Neumann-Morgenstern cardinal utility theory rests is crucially important in construct validation of the propositions which the theory claims to enunciate. In an attempt to gain

insight into the basic foundations of the von Neumann-Morgenstern cardinal utility theory, the following discussion will give a critical appraisal of each axiom underlying the theory and the results which are derived from the axioms. The axioms which will be discussed are: (1) the transitivity axiom, (2) the continuity of preferences axiom, (3) the sure-thing axiom, (4) the independence of ordering axiom, and (5) the compound-gamble axiom.

The transitivity axiom states that[15]: A>B and B>C => A>C. At the level of introspection, this axiom seems to be a reasonable assumption about human behavior; some experiments have supported this view while the results of others have claimed otherwise.

Marschak (1964) performed a casual experiment on his own graduate students to test for transitivity. Pairs of objects were presented in the following order:

(A1,B1);(A2,B2);...;(Am,Bm);

(B1,C1);(B2,C2);...;(Bm,Cm);

(C1,A1);(C2,A2);...;(Cm,Am).

The objects considered were those relevant to graduate students: jobs, trips, apartments, medical care, etc. The subject is considered to be not consistent if he prefers Ai to Bi, and Bi to Ci yet prefers Ci to Ai for any i. The results indicated that students satisfied the transitivity axiom when m was small; eg. m=5.

---

[15] A>B is defined to mean that A is preferred to B.

Michalos (1967) has argued that transitivity is an inaccurate empirical generalization and unacceptable normative principle although his arguments do not seem convincing.

In May's (1954) experiment, the results showed that subjects made intransitive preferences. May's explanation was that the subjects' preference orderings ,in fact, are intransitive. However, Rose (1957) claimed that the intransitive effect is an artifact arising in the course of the experiment. For one thing, at the moment of evaluation, a person's preferences may be transitive but during the course of the experiment, his preferences may remain transitive but change. Thus, the apparent intransitivity, reflects the change in preferences rather than "irrationality". For example,a person whose preference ordering is A>B>C will state that he prefers A to B and B to C but 30 minutes later he may change his preferences to C>A>B in which case he will say that he prefers C to A, which would give rise to an apparent intransitivity.

In experiments, the inevitable unintentional slips in judgement, hastiness in responding, and carelessness may contribute to misleading conclusions regarding transitivity. Rose (1957) found that the number of apparent intransitivities in the judgements of each of his subjects was inversely correlated with the time they took to complete the experiment, a result which suggests that hastiness in responding may have contributed to apparent intransitivity.

In contrast to May's results, Papandreou (1953) offered experimental evidence to support the hypothesis that individual preference systems satisfy the axiom of transitivity. Two different experiments were conducted involving a total of 24 subjects and 17,604 responses (pairwise choices). Both experiments gave results which strongly support the transitivity hypothesis.

Tullock (1964) has argued that the assumption of transitivity is not particularly dubious. Furthermore, he postulated that any apparent intransitivity may be tested by presenting the subject with a choice among all elements of the intransitive loop simultaneously. If the subject's preference ordering is actually intransitive, he will be unable to choose among the elements because for any element of the intransitive loop there is always a more preferred element. The ability of May's subjects to rank apparently intransitive elements is contradictory to the hypothesis that they had intransitive preference orderings.

The axiom regarding <u>continuity of preferences</u> states that: $A>B>C \Rightarrow$ there exists a probability p $(1>p>0)$ for which $(A,C;p)>B$ and also there exists a probability q $(1>q>0)$ for which $B>(A,C;q)$.[16]

The axiom asserts that there exists an intermediate value of p between 0 and 1 for which the probability

---

[16] $(X,Y;p)$ represents the gamble whose outcome is X with probability p or Y with probability 1-p.

"mixture" (i.e. (A,C;p)) of the less preferred outcome C and the more preferred outcome A will be preferred to the intermediately preferred outcome B.[17]

This axiom implies that the probability p plays a critical role in determining relative utility magnitudes of outcomes. The probability p for which (A,C;p)>B reflects the relative positions of A,B, and C on the utility scale. The lower limit for p (for which (A,C;p)>B) will increase as the "distance" on the utility scale between A and B increases.

Introspection would indicate that such an axiom be invalid in cases where A and B are highly disparate in utility value from C. For example, if alternative A is receiving 2 pennies, alternative B is receiving 1 penny, and alternative C is being tortured to death, it seems unlikely that a person will prefer (2¢,being tortured to death;p) to receiving 1¢ no matter how much (1-p) is lowered (for 1>p>0). One may dismiss the contradictory nature of this example by claiming that such disparity in values never occur in practical situations or by asserting that a sufficiently small probability (1-p) does not exist since people cannot relate to very small probabilities (eg. what is the psychological significance of a probability of one-millionth or one-trillionth?). Most scholars have accepted

---

[17] The discussion of the axiom will be restricted to the case for which (A,C;p)>B. From parallel considerations, the discussion can be easily extended to the case for which B>(A,C;q).

this axiom as a good enough approximation to actual behavior to be useful while others (such as Hausner and Wendel, 1952;Hausner, 1954;Thrall, 1954) have modified the cardinal theory of utility in which this axiom is not assumed to hold.

The sure-thing axiom states that: $A>B \Rightarrow A>(A,B;p)$. Conversely, $B>A \Rightarrow (A,B;p)>A$. The preferences (>) are changed to indifferences ($\sim$) for $p=1$.[18]

The gamble (A,B;p) will result in one of either two outcomes. If outcome A occurs, the person will be indifferent between this outcome and the sure-thing A; if outcome B occurs, the person will prefer the sure-thing A to this outcome. Thus, the gamble (A,B;p) has two possible outcomes A and B for which the sure-thing A is at least as preferred as either and is definitely preferred to one of them (namely the outcome B). Discussed along these lines, the axiom certainly seems like a reasonable description of human behavior as well as a norm by which human behavior should abide.

Despite the intuitive plausibility of this axiom, Marschak (1950) has given mountain climbing and Russian Roulette as two examples for which the axiom apparently does not hold. Marschak claims that in mountain climbing, (living safely,serious injury;p)>living safely even though living

---

[18] The following discussion will apply to the former case of A>(A,B;p) from which a parallel discussion is implicit for the latter case of (A,B;p)>A.

safely>serious injury and similarly for Russian Roulette, (staying alive,being killed;5/6)>staying alive even though staying alive>being killed; both instances seem to illustrate behavior contradictory to the axiom. However, Adams (1960) suggested that mountain climbing and Russian Roulette cannot be formulated in such simple terms. He pointed out that Marschak had assumed that "living safely" at home (i.e. the alternative to not mountain climb) is identical to "living safely" in the alternative to mountain climb, an assumption which he felt was incorrect. Similarly, Adams felt that in the case of Russian Roulette, the term "staying alive" involved in not participating and in participating does not assume the same meaning in both cases. If the person participates, he has prestige to gain and if he does not participate, he has his courage to lose.

Even if Adam's explanation is sufficient, it would still seem that mountain climbing and Russian Roulette would contradict the sure-thing axiom because the undertaking of each activity biases the utilities of the outcomes involved in the gamble (the act of mountain climbing and playing Russian Roulette enhances the utility of staying alive). Thus, if under certainty, A>B and the "love of gambling" is significant enough so that the risk involved in (A,B;p) is sufficient to enhance the utility of B, the gamble (A,B;p) may be preferred to A, which is behavior that is incompatible with the axiom. Thus, love of risk or gambling

is contradictory to the sure-thing axiom[19]. The axiom may be rescued by assuming that risk-taking in situations of practical interest will influence the utilities of the outcomes to a negligible extent.

The axiom, regarding independence of outcome ordering, states that: (A,B;p) (B,A;1-p). The psychological premise underlying this axiom is that the order of presentation of the gamble outcomes should not affect a person's subjective value of that gamble. For example, suppose that a gamble is described by two cards, each written with an outcome and its associated probability. The axiom asserts that in presenting the gamble to the subject for judgement, the order in which the cards are presented to the subject is irrelevant. The intuitive appeal of this axiom is obvious.

The compound-gamble axiom states that a person is indifferent between (1) a gamble which offers a further gamble as outcome and (2) a gamble which is a reduced form of the first gamble, i.e. the probabilities of outcomes are the statistical equivalent to the first gamble. In notational form,

$$((A,B;p1),B;p2)\sim(A,B;p1\cdot p2)$$

The axiom is confronted by three objections of

---

[19] The love of risk used in this context is not to be confused with a person's preference for unfair gambles, a preference behavior indicated by a concave-upward utility curve (Mosteller and Nogee, 1954 dubbed the term 'extravagant behavior' to describe this latter behavior.). Extravagant behavior is compatible with the axiom while love of risk is not.

psychological significance.

(1) The axiom asserts that subjective estimates of probabilities coincide with the statistically specified probabilities. Chapter 1 points out that such coincidence is not apparent in experimental results (eg. Edwards and Phillips, 1966).

(2) The "love of gambling" is incompatible with the dictates of the axiom. It may be argued that those people who react negatively to risk taking (i.e. detest gambling) will prefer (A,B;p1•p2) to ((A,B;p1),B;p2) since the latter gamble involves the possible participation in two gambles while people who react positively to taking of risks (i.e. love gambling) will prefer ((A,B;p1),B;p2) to (A,B;p1•p2) since the former gamble involves the possible participation in two gambles. This latter observation of human behavior is capitalized upon by casino owners. Slot machines have three revolving wheels (representing three risky situations) instead of one wheel (representing one risky situation) with equivalent probabilities of winning.

(3) The axiom asserts that people do not have probability preferences. Although the reduced form of the more complex gamble ((A,B;p1),B;p2) is statistically equivalent to the simpler gamble (A,B;p1•p2), a person may prefer ((A,B;p1),B;p2) to (A,B;p2) because he prefers the probabilities p1 and p2 to p1•p2 (or vice versa). For example, suppose a person is required to state his preference or indifference between ((A,B;1/2),B;1/2) and

(A,B;1/4). Although the axiom claims that a person will be indifferent between the two gambles, it is quite conceivable that he will prefer ((A,B;1/2),B;1/2) to (A,B;1/4) because he can more readily relate to a probability of 1/2 (head or tail, girl or boy newborn, etc.) than to probabilities of 1/4 and 3/4. In a series of experiments, Edwards (1953, 1954a, 1954b, 1954c) provided experimental evidence to show that people have probability preferences that cannot be accounted by utility considerations alone.

This axiom is perhaps the most controversial of the axioms not only because of the three previously mentioned objections but also because of its implicit assumption that humans are infinitesimally sensitive creatures. This assumption is implicit in the other axioms but is of crucial importance here because of the indifference relation (as contrasted to a preference relation) stated in the compound-gamble axiom. If subject X strictly prefers ((A,B;0.58),B;0.37) to (A,B;0.2146), can one conclude that subject X violated the axiom since $(0.58)\bullet(0.37)=0.2146$?

From these previously stated axioms, von Neumann and Morgenstern (1947) were able to deduce that:

(1) utility of z may be represented by numbers, u(z).

(2) $x>y \Rightarrow u(x)>u(y)$

(3) $u(x,y;p) = p\bullet u(x)+(1-p)\bullet u(y)$

(4) The utility u(z) is unique up to a linear transformation; i.e. the utility scale is cardinal. The cardinality result implies that unit and origin for this

utility scale are arbitrary.

Discussion of Previous Experiments

The following discusses previous experimental attempts to measure cardinal utility using the von Neumann-Morgenstern model. Other experiments employing elaborate modifications of the basic model (eg. stochastic model of choice, subjective expected utility model) are not discussed here.

Mosteller and Nogee's (1951) experiment represented the first attempt to measure cardinal utility using the von Neumann-Morgenstern model. National Guardsmen and Harvard undergraduates served as subjects and a game called poker dice served as the betting task. The subject was repeatedly required to accept or refuse bets stated in terms of rolls at poker dice. Subjects played with $1.00, which they were given at the beginning of each experimental session. Subjects were provided with a table which informed them whether a given bet was fair, or better or worse than fair. The subjects were presented with bets of the following form:

> "You can bet or not bet five cents against ten dollars that you can roll five dice once and get a better poker hand than 44441."
>
> or
>
> "You have the opportunity of betting or not betting five cents against this double

offer: if you beat 22263 you will receive 20 cents; if you do not beat 22263 but do beat 66431, you will receive three cents. If you do not beat either, you will lose the five cents you must risk to play. You will roll the dice only once."

The indifference offer was operationally defined as the amount of money for which the subject would accept the bet 1/2 of the time[20]. The experiment required approximately 30 hours from each subject.

The results indicated that for Harvard undergraduates utility for money was approximately proportional to money up to a point, after which marginal utility decreased. However, for the Guardsmen, their utility curves showed increasing marginal utility for money, which indicated that they were willing to play unfair gambles.

Realizing that interpretations of probabilities may be biased, Davidson, Suppes and Siegel (1957) set out to identify an event for which the subjective probability of occurrence is equal to its subjective probability of non-occurrence so that this event, rather than a stated probability, could be used in a gamble. A coin toss (head or tail), a die roll (odd number or even number), a two coin

---

[20] By providing this operational definition of indifference, Mosteller and Nogee have tacitly given a stochastic interpretation to the preference relations.

toss (match or non-match), and some other simple games were tried, but in each case most subjects showed a definite preference for one of the "equal-chance" outcomes. After much trial and error, a satisfactory event was found. The event was the toss of a specially designed die. On three faces of the die, 'ZOJ' was engraved while on the other three faces, 'ZEJ' was engraved. Two more die were similarly engraved with 'WNH' and 'XEQ', 'QUG' and 'QUJ'. Glaze (1928) had shown that these nonsense syllables had practically no association value. With all three die, their preliminary experimentation showed that the subjective probability of occurrence is equal to its subjective probability of non-occurrence; i.e., the subject associated a probability of 1/2 with the event. The subject was required to make pairwise comparisons between gambles. The instructions took on the following form:

> "I'll present you with two alternatives. You are to choose one of them. For example, if you want to bet on ZEJ against the field, you will win 18¢ if ZEJ comes up when you toss the die, but you will lose 4¢ if the field comes up (that is, if any side but the ZEJ comes up). On the other hand, if you want to bet on the field, you will win 6¢ if the field comes up, but you will lose 12¢ if ZEJ comes up."

From expressed preferences between these type of choices,

the experimenters were able to determine upper and lower bounds on their subjects' utility functions. The bounds were generally close together. The results showed that 15 out of the 19 subjects made choices as if they were attempting to maximize expected utility.

In Swalm's experiment (1966), the utility curves for money were derived from business executives. Realizing the problem of confounding subjective value of outcome with subjective probability, Swalm attempted to minimize this effect by using only probabilities of 1/2. Although Swalm expected that the probability 1/2 could be easily related (eg. to the concept of a coin flip), many subjects felt that the hypothetically constructed situations involving 50-50 odds were unrealistic because most of the situations encountered by them in their individual businesses were not 50-50 gambles. Subjects were repeatedly required to state the certainty equivalents of 50-50 gambles whose two outcomes are given by the experimenter and whose utilities were previously derived. The questions took on the following form:

> "Suppose that you planned to purchase a general-purpose machine but a colleague proposed, instead, to buy a more efficient special-purpose machine. Both cost the same; the difficulty is that the contract for which the special-purpose machine would be required has only a 50-50 probability of being

> received. If it is received, the special-purpose machine will yield a profit of $250,000. If not, your net income will be zero. On the other hand, the general-purpose machine will produce a certain savings of, say, $100,000. Which would you recommend?"

The money amounts used in the gambles were chosen to be meaningful to the subject. Losses as well as gains were involved.

The results of the experiment indicated that: (1) the subjects did not maximize expected monetary value, (2) the utility curves provide a basis for identifying risk-aversion (concave downward utility curve) and extravagant behavior (concave upward utility curve), (3) the utility curves revealed different attitudes toward risk decisions among executives of the same company, and (4) most of the utility curves indicated risk-aversion.

## Discussion of our Experiment

The standard gamble method was used to determine von Neuman-Morgenstern cardinal utility for days in bed. The application of the method was repeated in the second experimental session. The application of the method involved finding certainty equivalents to 50-50 gambles stated in the context of a scenario. The essence of the scenario is given here.

> "You are suffering from a case of the flu. Your family doctor informs you that in your particular case there are two possible treatments available. The first treatment will result in either a days of rest in bed or b days of rest in bed with equal chances. The second possible treatment will result in a fixed number of days in bed for certain. State the number of days in bed for certain for which you will be indifferent between this treatment and the first one."

Initially, a and b were chosen to be 0 and 15 days since our domain of interest was from 0 to 15 days. Then the subject was repeatedly asked to state certainty equivalents x to 50-50 gambles with outcomes a and b (whose utilities are known) from which the utility of x was calculated according to the following formula:

$$u(x)=0.50 \cdot u(a)+0.50 \cdot u(b)$$

The utilities of 0 and 15 days were arbitrarily set to 1 and 0 respectively. The gambles were presented in the form of the previously mentioned scenario. Approximately 10 points were identified on each subject's utility curve. The utility curve was approximated by piecewise linear interpolation. The procedure for determination of the utility curve was repeated in the next session.

## Section II - Ordinal Utility Represented by Indifference Curves

### Introduction

Classical utility theory was used by theorists (eg. Jevons, Menger, and Marshall) to establish consumer demand for commodities. When relations between utilities of different commodities were considered, Jevons, Menger, and Marshall assumed the additivity of utilities. However, such an additivity notion is inconceivable for commodities which are not independent. This objection was resolved through recourse to Edgeworth's (1881) notion of indifference curves for which the restrictive additivity assumption is not needed and from which substitutability and complementary relationships between commodities may be interpreted. Furthermore, Pareto (1906) seriously doubted whether preferences could be measured on a numerical scale as implied by the classical cardinal theory. Rejecting classical cardinal utility in favor of ordinal utility, he asserted that the same conclusions about consumer demand that had been drawn from classical cardinal measures could be drawn from analysis of indifference curves. In fact, Hicks and Allen (1934) and Samuelson (1938) derived all the usual conclusions about consumer behavior from analysis of indifference curves. Samuelson asserted that the structure of the theory of consumer choice could be derived from observation of choices among alternatives available to a consumer (the concept of revealed preference). The essence

of this approach is that each choice defines a point and a slope in the choice space from which a family of slopes constitute an indifference hyperplane.

Ordinalists argued for abandoning cardinal utility because analysis of indifference curves could deduce the same results in the area of riskless choice as cardinal utility could with its stronger assumptions. In general, the only required assumption for the derivation of indifference curves is that concerning weak ordering of commodity bundles, i.e. that the subject is able to express preference or indifference between commodity bundle pairs and that his preference and indifference relations are transitive[21]. In terms of the indifference map, the intersection of two indifference curves implies the violation of transitivity of the indifference relation. This transitivity assumption is also required in the cardinal utility theory but the other axioms required for cardinal utility are not essential to the ordinal theory. Thus a stronger utility scale requires a stronger axiomatic system (this point is discussed in Chapter 1).

---

[21] Transitivity was discussed in Section I of this chapter.

Discussion of Previous Experiments

The scarcity of studies devoted to experimental derivation of indifference curves is a particular striking feature of the literature.

There were two notable early experimental attempts to derive indifference curves. In 1931, Thurstone performed a simple experiment to elicit preferences from a subject whereupon indifference curves were derived. Thurstone's research assistant served as the subject. Preference judgements between hat and overcoat combinations were required from the subjects. From these preference judgements, Thurstone was able to locate an indifference curve between two regions of the choice space, namely the region in which the reference combination is preferred to the combinations in the region and the region in which the hat and overcoat combinations in the region are preferred to the reference combination. The same procedure was repeated for hats and shoes, and for shoes and overcoats. From five psychophysical laws, Thurstone chose Fechner's logarithmic law to fit an indifference curve between the two regions. Four indifference curves were fitted in the hat X shoe space and 4 indifference curves were fitted in the hat X overcoat space.

From these indifference maps, indifference curves in the shoe X overcoat space were predicted using the additivity assumption. The predictions were surprisingly accurate but Thurstone (1953) hypothesized that the

consistencies "were the result of careful instructions to assume a uniform motivational attitude."

The second experimental determination of indifference curves was attempted by Rousseas and Hart (1951). The experiment required 67 students to rank sets of three bacon-egg combinations. Vectors in the choice space were constructed based upon directions of preference between choices. Those vectors which were consistent with the experimenters' assumed saturation levels and convexity properties were used for constructing the indifference curves. By making a dubious assumption regarding homogeneity of preferences among the students, a composite indifference map was derived by curve fitting a set of vectors. A major difference between this experiment and Thurstone's is that Rousseas and Hart provided a motivation for careful consideration by forcing the subject to consume the top ranked egg-bacon combination.

From the economist's point of view, the previously mentioned experiment (and generally any experimental derivations of indifference curves or any other forms of utility for that matter) face objections of which the essence is that hypothetically constructed choices made in a controlled laboratory situation do not reflect "actual" preferences. Characterizing the economist's standpoint, Wallis and Friedman (1942) expressed doubts about the applicability of deriving indifference curves on laboratory data for which MacCrimmon and Toda (1969) offered rebuttals.

In 1969, MacCrimmon and Toda presented an efficient method for determining indifference curves based upon dominance concepts. Briefly, if both commodities are monotonic in preference (i.e. more is preferred to less or vice versa) expressed preference between a chosen point in the choice space and the reference point leads to the identification of a rectangular subset of the acceptance region and a rectangular subset of the rejection region[22]. The acceptance region consists of all points which are accepted in lieu of the reference point while the rejection region consists of all points which are rejected in lieu of the reference point. Expressed choice between other points in the choice space and the reference point leads to identification of a greater portion of each region until the band between the identified portions of the acceptance and the rejection region is sufficiently narrow to locate an indifference curve. For the case of only one monotonic valued commodity, lines instead of rectangular subsets in each region are identified.

In a training session lasting about 2 hours, the 7 subjects were taught to derive their own indifference curves using the previously mentioned method. In order to provide an incentive for revealing true preferences, payoffs were made according to the derived indifference curves. The domain of choice included ball point pens, money, and French

---

[22] The procedure will be more fully discussed later in this chapter.

pastries. In the first session, 7 indifference curves were derived in the money X pen space for which it was assummed that more is preferred to less for both commodities (rectangular method). In the second session, 4 indifference curves were derived in the money X pastry space for which it was assumed that more is preferred to less for money but not necessarily for all levels of pastries (line method). Finally subjects were given a series of pairwise comparison bundles on either side of the indifference curves drawn. Of 21 of these consistency checks given to each subject, an average of less than one choice per subject was actually inconsistent with the appropriate indifference curve.

Discussion of our Experiment

The subject was required to choose among alternatives stated in the context of a scenario. The essence of the scenario is presented here.

> "You are in a bed recovering from a case of the flu. The treatment prescribed to you by your physician is complete rest in bed. However, your physician informs you that after you leave the bed, the flu has some chance of immediately recurring and hence there will be a probability that you will have to return to bed for an additional five days.
>
> Choose between resting for :
>
> (i) a days in bed with p probability of an

additional 5 days in bed

or

(ii) b days in bed with q probability of an additional 5 days in bed."

Indifference as well as strict preference judgements were allowed. The days presented were whole numbers and the probabilities were given to the first decimal place (i.e. 0.0(0.1)1.0).

The method we used is adopted from the MacCrimmon-Toda method for derivation of indifference curves. Because of the special nature of our alternatives (i.e. outcome X risk instead of the usual outcome X outcome choice space), we have added certain assumptions in addition to the monotonicity and transitivity assumptions which MacCrimmon and Toda used (1969). These added assumptions will be effective in identifying additional portions of the reject and accept region. Assumptions we used are presented here.

(1) monotonicity of preferences: An alternative is preferred to all other alternatives which have more number of days in bed and more probability of an additional 5 days in bed. An alternative is rejected in favor of all other alternatives which have less number of days in bed and less probability of an additional 5 days in bed.

(2) transitivity: Strict preference and indifference relations are transitive.

(3) Let (a,b) represent an alternative where the first coordinate represents number of days in bed and the second coordinate represents the associated probability of an

additional 5 days in bed.

(a) (d1,p1) is preferred to all other alternatives (d2,p2) where d2>(d1+5).

(b) (d1,1) (d1+5,0)

To illustrate the method, suppose that we wish to determine the indifference curve that passes through the reference point Ro=(10,0.50) . Suppose that our domain of interest is 0≤days in bed≤20 and 0≤probability of an additional 5 days≤1.0. The method is equivalent to determining subsets of (1) an accept region A and (2) a reject region R, whose common boundary is the indifference curve. The accept region consists of all points which are preferred to Ro, while the reject region consists of all the points which are rejected in favor of Ro. Of course, the experimental application of the method cannot ever hope to identify the total accept and reject regions since this would require an infinite number of judgements from the subject. However the application of the method endeavors to identify as much of the accept and reject regions as practically possible from which an indifference curve may be approximately located.

(i) Using assumption (1), subsets of the A and R regions are shaded as shown in Figure 7 .

(ii) Using assumption 3(a), the unshaded region is further reduced as shown in Figure 8 .

(iii) The subject is then presented with a point in the

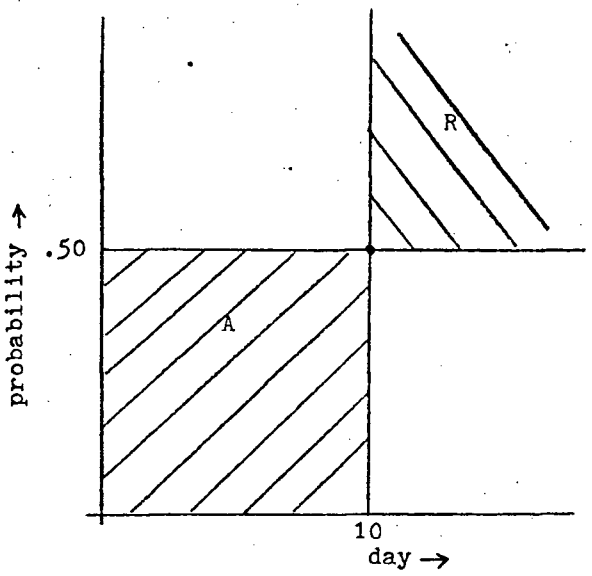


Figure 7 - Identifying Accept and Reject Regions

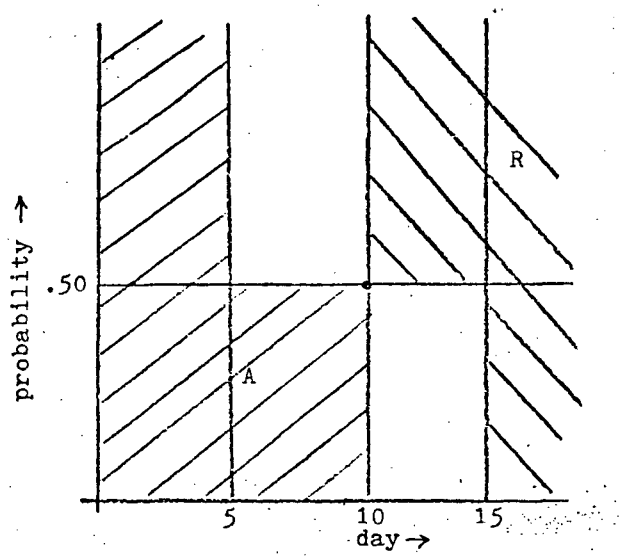


Figure 8 - Identifying Accept and Reject Regions

unshaded area and asked to compare this point with Ro. Suppose the subject prefers (12,0.20) to Ro. Using assumptions (1) and (2), a subset of the A region is shaded as shown in Figure 9 .

(iv) Furthermore, using assumptions 3(a) and (2), a subset of the A region is shaded as shown in Figure 10 .

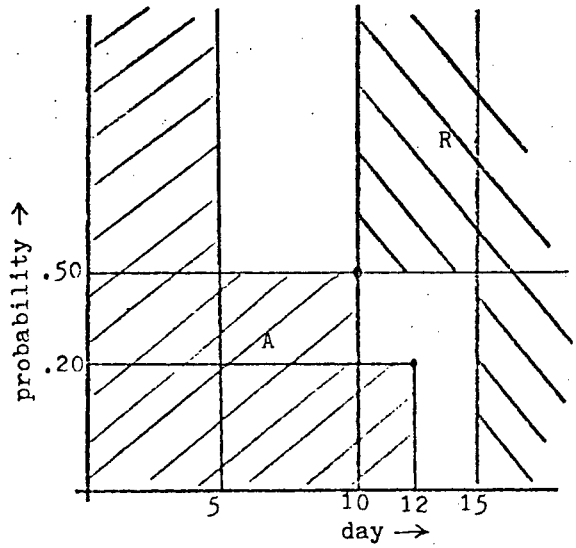


Figure 9 - Identifying Accept and Reject Regions

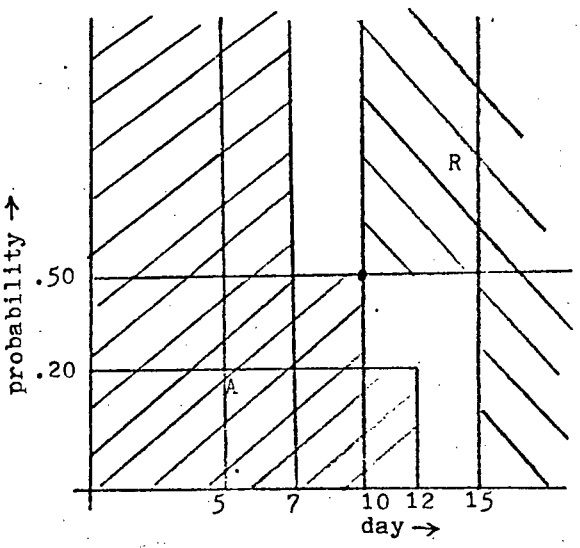


Figure 10 - Identifying Accept and Reject Regions

Thus, we see that for each choice between presented alternatives, the unshaded area is further reduced.

The choices were sequentially presented to the subject until the unshaded area was reduced as much as possible (using whole number of days and first place decimal probabilities). A piecewise linear interpolation was used to approximate the indifference curve within the region

unshaded area.

# Chapter 4 - Results, Analysis, and Conclusions

## Section I - Introduction

The study focuses upon the application of two methods for eliciting preferences from subjects, which have received significant attention in the literature. The methods are: (1) the MacCrimmon-Toda method for constructing indifference curves, and (2) the standard gamble certainty equivalence method for constructing von Neumann-Morgenstern cardinal utility functions.

While individual discussion of each method was completed in Chapter 3, the aim of this chapter is to report on an experiment which was conducted for comparing the two methods in terms of the following criteria:

(1) Test-retest correspondence of preference judgements

(2) The existence of "personal attitudes" which affect the test-retest correspondence

(3) Linearity of derived indifference points in the day X probability space.

(4) Goodness of prediction

In addition, the study attempts to identify inter-method correspondence of predictions of preferences.

Finally, the study inquires into some possible relationships between test-retest and inter-method correspondences and attitudes (eg. concerning rationality,

interpretation of probabilities, and discrimination bands).

The sample consisted of 23 commerce students, of which 14 were undergraduates and 9 were graduates. All have had some previous exposure to the concept of utility, but none had ever participated in experiments for eliciting preferences. The domain of choice concerned decisions to stay in hospital bed and take a risk of readmission for an additional period. All subjects have stayed in hospital at least once, but only 4 have stayed in hospital for 3 or more days in the year preceding the experiment.

The study consisted of experimental derivations of indifference curves and utility functions in repeat tests with the same sample of subjects, as well as the administration of a questionnaire in the last experimental session. The experimental sessions were conducted separately with each subject. In each session the subject was presented with two scenarios relating to choices involving hospital days in bed and probability of additional five hospital days in bed. The first scenario presented the subject with three reference points consisting of days in bed for sure and probability of additional fixed number of hospital days. The reference points in the day X probability space were respectively (4,0.50), (7,0.50), and (10,0.50). A method developed by MacCrimmon and Toda (1969) was employed to elicit preferences from the subject in relation to a given reference point and consequently deriving an indifference curve, i.e. a curve reflecting trade-offs which maintain

the welfare of the subject at an equal level. Following the application of this method, a von Neumann-Morgenstern utility function for hospital days in bed was derived for the subject using certainty equivalences for gambles involving chances of days in bed. The order of application of each on the two methods of eliciting preferences was reversed for the second session.

This procedure was repeated twice for each subject with a minimum of one week delay and a maximum of two weeks delay between sessions for each subject.

In the third session the subject was asked to fill a questionnaire consisting of questions related to the following themes:

(1) Attitudes concerning the acceptance of particular fundamental rationality axioms.

(2) Propensities for certain judgemental modes of evaluation.

(3) Evaluation of various components of the experiment (eg. the scenario introducing the choice space for each method).

(4) Identification of the discrimination band in the choice space (probabilities and days).

(5) The subjective interpretation of certain colloquial probability expressions.

The first theme of the questionnaire attempted to provide an indication of the subject's agreement with the appropriateness of some of the fundamental axioms underlying the methods used in the experiment. We have used Savage's

defence of rationality axioms (1954) in designing the form of the questionnaire items. Savage argued that rationality provides the rules for reasonable behavior and that when a subject is aware of violating these rules he will tend to revise his decisions. Our questionnaire items presented the subject with examples illustrating violations of some axioms. Then he was asked to rate his agreement with the need to revise the decision, on a Lickert scale (ranging from 1=strongly disagree through 4=neutral to 7=strongly agree).

The following are examples from the questionnaire of cases where the transitivity axiom and the compound-gamble axiom are violated:

> "George prefers driving a Ford Pinto to a Toyota MK II. Furthermore, he prefers driving a Toyota MK II to a Datsun 1600. Yet, from a rent-a-car which offers a Datsun 1600 or a Ford Pinto at the same rental rate, George rents a Datsun 1600 instead of a Ford Pinto. Realizing this "inconsistency", George should change his choice to Ford Pinto."
>
> 7=strongly agree-6-5-4-3-2-1=strongly disagree

> "A sweepstake ticket entitles the holder to either a prize of $1.00 or a chance in the grand final draw. The grand final draw prize will be either $100.00 or $1.00.

Another sweepstake ticket entitles the holder to a prize of $1.00 or $100.00.

Both sweepstake tickets sell for the same price. Taking the chances of winning into account, Dan calculates that the probabilities of winning each prize are the same for both sweepstakes. In spite of this information, Dan insists upon buying the second sweepstake ticket and is even willing to pay slightly more for this ticket. Dan should stop favoring the second sweepstake."

7=strongly agree-6-5-4-3-2-1=strongly disagree

Questions relating to the second theme tapped attitudes concerning the general mode felt appropriate for this domain of decision making, eg. to what extent one prefers careful logical judgement to spontaneous response to the problem situation in health matters. Again the subject was asked to rate his agreement on a Lickert scale. The following are examples of these questions:

"In health matters, people ought to carefully evaluate their preferences among alternatives without being influenced by their mood or emotion at the moment of evaluation."

7=strongly agree-6-5-4-3-2-1=strongly disagree

"In matters concerning illness, people ought

to evaluate their preferences among alternatives before the illness actually occurs because under pain and discomfort thay may not be clearly aware of their preferences."

7=strongly agree-6-5-4-3-2-1=strongly disagree

The third theme of the questionnaire focused on evaluation of the experiment and the methods used. The subject was asked to: (1) rate the realism of the presentation of the scenarios introducing the domain of choice for each method, and (2) compare the difficulty of judgements required by each method. Finally questions were directed to identify the confidence the subject has in the methods. To this end two questions about each method were presented to the subject; one question was concerned with his willingness to have the elicited preferences used in lieu of his personal judgements when a situation of choice arises, while the other was concerned with the question of which preferences should dominate in a decision making situation- those which were obtained prior to the health situation or judgements spontaneously made in the face of the situation. Examples of the questions are presented below:

"Suppose that with reference to a particular health matter, a trained health personnel derives your indifference curves. If a situation resembling the scenario arises in

. real life, would you let a physician determine the decision for you from a careful consideration of your indifference curves?"

7=without any doubt-6-5-4-3-2-1=with complete doubt

"suppose that you were to actually encounter a situation where you had to compare two alternatives each involving days of rest in bed and associated probability of additional days in bed (as in our experiment). Furthermore suppose that the decision you actually make does not conform with your responses using method ... in this experiment. In light of this information, how important do you feel that it is for you to change your decision?"

7=extremely important-6-5-4-3-2-1=extremely unimportant

The fourth theme in the questionnaire focuses upon measurement of "discrimination "bands", i.e. to what extent changes in stimuli such as "day" or "probabilities" are perceived significant by the subject and significantly affect his judgements. For example the subject was asked to rate the significance of changes in probabilities from 0.5 to 0.55 as opposed to 0.6 to 0.8. We have selected values which reflect both the levels of probabilities and the degree of change in probabilities. Similar comparisons were obtained for changes in the number of days in hospital.

The fourth theme in the questionnaire consisted of one question aiming at providing insight into the possible biasing effects of using probabilities in eliciting preferences. We have attempted to associate colloquial expressions of risk with the objective scale of probability. The form of the question presented below is based on a method which was used by Lichtenstein and Newman (1967) for a similar purpose.

> "What probabilities do you associate with the following words (or phrases):
> (a) certain
> (b) unlikely
> (c) highly probable
> (e) uncertain
> (e) probable
> (f) impossible
> (g) extremely likely" [23]

The final procedure used in our experiments presented subjects with re-evaluation of "gambles" for which diammetrically opposed choices were indicated for the subject using the alternative methods. The subject was requested to make an additional judgement as to his preferences among these gambles.

---

[23] The responses are presented in Appendix A.

## Section II - Discussion of Experimental Design

The central indices chosen for evaluating and comparing the methods were: (a) test-retest squared deviations for each method (intramethod inconsistency), and (b) intermethod squared deviations of corresponding indifference predictions obtained by each method (intermethod inconsistency).

In order to provide a base by which meaningful comparisons between the two methods of eliciting preferences could be made, the utility curve for hospital days derived by the standard gamble method was transformed into the day X probability space using the following procedure:

(1) The utility of the first reference point (4,0.50) was calculated from the utility curve. The point (4,0.50) represents a "gamble" for which the outcome will be 4 days with probability 0.50 or 4+5=9 days with probability 0.50. Therefore, the utility of the reference point is

$$Uref=0.50 \cdot u(4)+0.50 \cdot u(9)$$

which may be calculated since u(4) and u(9) can be found from the subject's empirically derived utility curve.

(2) From this reference utility, the (day,probability) trade-off equivalents to the reference point was calculated for probabilities from 0 to 1 at increments of 0.05. The problem becomes that of finding X for a given p such that:

$$u(X,p)=Uref$$

$$\text{or } (1-p) \cdot u(X)+p \cdot u(X+5)=Uref$$

The equation takes on the same form as f(X)=0 for which the bisection search method was used to find X for a given p

(for p=0(0.05)1).

Thus, from following these two procedures, we have obtained from the von Neumann-Morgenstern utility curve, an indifference curve corresponding to the one we have derived using the MacCrimmon-Toda method. The MacCrimmon-Toda derived trade-off curve and the von Neumann-Morgenstern derived trade-off curve intersect at the reference point since the points on each trade-off curve are trade-off equivalents to the same reference point. The two procedures were similarly applied to the same utility curve in order to derive a second trade-off curve with points which are trade-off equivalents to the second reference point of (7,0.50). The entire procedure applied to the two reference points was repeated for the utility curve derived in the second experimental session.

Once the two trade-off curves derived by each method for each session were identified, the two measures of inconsistency, test-retest and intermethod squared deviations could be calculated.

An indication of the test-retest inconsistency for each method was provided by the squared deviations between the two trade-off curves derived for both sessions for a given method. For the MacCrimmon-Toda method, the squared deviation between the two derived trade-off curves was denoted by $(MT1-MT2)^2$ and defined as:

$$(MT1-MT2)^2 = \sum_{i=1}^{21} \sum_{j=1}^{2} (MT_j^1(i)-MT_j^2(i))^2$$

Where $MT_j^1(i)$= the value of the first argument of the (day,probability) point on the MacCrimmon-Toda derived trade-off curve for the probability of (21-i)/20, the first experimental session, and the jth reference point,

and $MT_j^2(i)$=the value of the first argument of the (day,probability) point on the MacCrimmon-Toda derived trade-off curve for the probability of (21-i)/20, the second experimental session, and the jth reference point.

Similarly, the test-retest squared deviation for the standard gamble method was denoted by $(VM1-VM2)^2$ and defined as:

$$(VM1-VM2)^2 = \sum_{i=1}^{21} \sum_{j=1}^{2} (VM_j^1(i)-VM_j^2(i))^2$$

Where $VM_j^1(i)$= the value of the first argument of the (day,probability) point on the standard gamble derived trade-off curve for the probability of (21-i)/20, the first experimental session, and the jth reference point,

and $VM_j^2(i)$=the value of the first argument of the (day,probability) point on the standard gamble derived trade-off curve for the probability of (21-i)/20, the second

experimental session, and the jth reference point.

An indication of the intermethod inconsistency was provided by the squared deviation between the trade-off curves derived from both methods for a given experimental session. For the first experimental session, the intermethod deviation was denoted by $(MT1-VM1)^2$ and defined as:

$$(MT1-VM1)^2 = \sum_{i=1}^{21} \sum_{j=1}^{2} (MT_j^1(i)-VM_j^1(i))^2$$

For the second experimental session, the intermethod deviation was denoted by $(MT2-VM2)^2$ and defined as:

$$(MT2-VM2)^2 = \sum_{i=1}^{21} \sum_{j=1}^{2} (MT_j^2(i)-VM_j^2(i))^2$$

Finally, an indication of the inconsistency between the trade-off curves for each method averaged over the two sessions was denoted by $(av(MT)-av(VM))^2$ and defined as:

$$(av(MT)-av(VM))^2 = \sum_{i=1}^{21} \sum_{j=1}^{2} \left( \left( \frac{MT_j^1(i)+MT_j^2(i)}{2} \right) - \left( \frac{VM_j^1(i)+VM_j^2(i)}{2} \right) \right)^2$$

The conceptual models underlying our experiment are based upon a number of propositions as to possible variables which affect test-retest, intermethod, and averaged-method deviations.

## Test-retest Correspondence of Preferences

The first model which is presented in Figure 11 hypothesizes that attitudes toward assumptions which underly each method, confidence in the method (as a measure of motivation), propensities for certain judgemental modes of evaluation, realism of method scenario, the width of the discrimination band for probability and day stimuli, bias in interpreting probability stimuli, and the use of simple (linear) rules for judging preferences, would all tend to affect correspondence between responses in repeat tests and original responses. As to the discrimination band, we have hypothesized that there is an optimal level of sensitivity to stimuli magnitudes,i.e. there exists a threshold sensitivity level, deviation from which will lead to more pronounced differences in test-retest responses.

## Intermethod Correspondence of Preferences

The second model (presented in Figure 12 ) differs from the first only in the definition of the dependent variables: $(MT1-VM1)^2$ for the first session and $(MT2-VM2)^2$ for the second session. It would seem that most of the independent variables which are hypothesized to affect test-retest consistency will also affect intermethod consistency. However, the degree of association may differ between the dependent and the independent variables in the two cases.

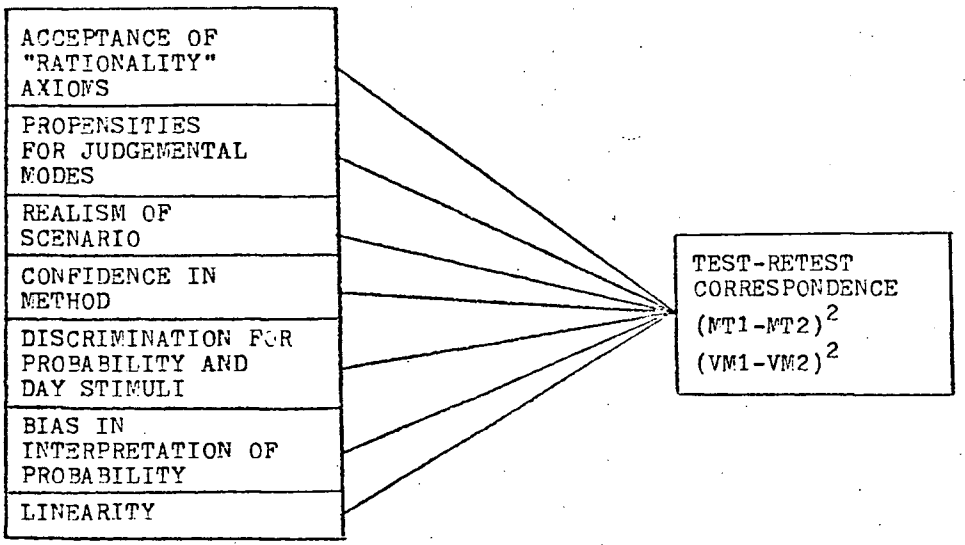


**Figure 11 - Test-retest Model**

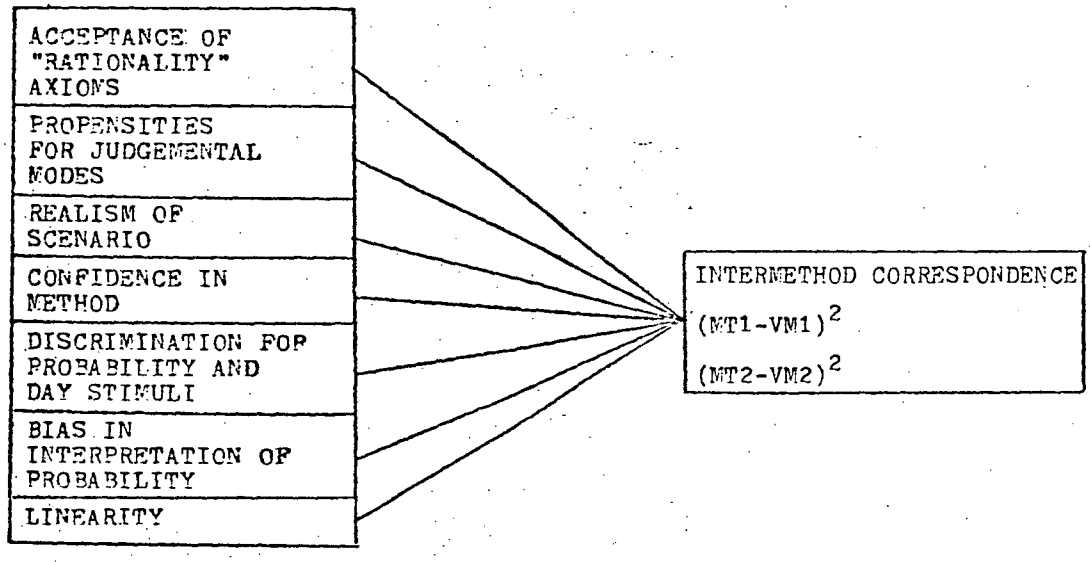


Figure 12 - Intermethod Model

## Section III - Results and Analysis

Dependent Variables (Inconsistency Measures) - Profile of Results

Figure 13 displays the profile of inconsistency measures for each of three groups: undergraduates, graduates, and the total sample. The tabled responses bring to attention two striking patterns: (1) the median for each of the inconsistency measures for the graduate group is lower than that for the undergraduate group, and (2) some inconsistency measures are higher in value than others across all groups.

One possible explanation for the first observation is that the graduates, because of a longer socialization process and a greater skill in mental computing, tended to make a much more conscientious effort in responding consistently and make fewer errors in computing their responses. The second observation is that the median for $(MT1-MT2)^2$ is higher than for $(VM1-VM2)^2$ in both sample groups. The observation that the median for $(MT2-VM2)^2$ is lower than that for $(MT1-VM1)^2$ for both sample groups seems to suggest that intermethod consistency may improve with experience (learning). The median for $(av(MT)-av(VM))^2$ is lower than that for $(MT1-MT2)^2$ and $(VM1-VM2)^2$ for all three samples; thus, averaging indifference curves derived by each method seems to improve intermethod consistency.

| | Undergraduates N=14 | | | | Graduates N=9 | | | | Total N=23 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Median | σ | Median | Mean | Median | σ | Range | Mean | Median | σ | Range |
| $(MT1-MT2)^2$ | 37.5 | 30.7 | 27.8 | 4.5-95.2 | 8.5 | 6.6 | 8.2 | 0.0-22.2 | 26.1 | 20.8 | 26.2 | 0.0-95.2 |
| $(VM1-VM2)^2$ | 10.2 | 4.7 | 12.0 | 0.7-41.5 | 4.1 | 3.3 | 2.7 | 0.0- 9.3 | 7.8 | 3.3 | 9.8 | 0.0-41.5 |
| $(MT1-VM1)^2$ | 38.7 | 42.1 | 21.4 | 5.1-69.2 | 19.0 | 14.8 | 16.1 | 0.0-42.3 | 31.0 | 31.6 | 21.4 | 0.0-69.2 |
| $(MT2-VM2)^2$ | 28.1 | 25.5 | 9.9 | 10.5-52.0 | 16.9 | 12.2 | 15.2 | 0.0-49.1 | 23.7 | 23.6 | 13.1 | 0.0-52.0 |
| $(av(MT)-av(VM))^2$ | 22.7 | 23.6 | 8.0 | 9.4-34.8 | 15.0 | 8.4 | 14.0 | 0.0-41.5 | 19.7 | 20,9 | 11.1 | 0.0-41.5 |

All entries are in units of $(day)^2$

Figure 13 - Inconsistency Measures - Profile of Results

## The Independent Variables

The independent variables, which were proposed in our test-retest and intermethod models, are operationally defined in terms of our indices based upon questionnaire items. Responses to those questionnaire items which are relevant to a particular concept are aggregated to provide an index for that concept. The following discussion indicates the concepts and their respective definitions.

Refer to Appendix B for a list of questionnaire items.

(a) acceptance of "rationality" axioms

The independent variable, characterizing acceptance of "rationality" axioms, is defined in terms of an index which is a simple linear combination of the Lickert scales for the three questionnaire items that measure attitudes concerning acceptance of the transitivity axiom, the sure-thing axiom, and the compound-gamble axiom. The index is normalized in such a way that a value of 1 indicates strong disagreement with all three axioms, while a value of 7 indicates strong agreement with all three axioms.

(b) propensities for judgemental modes of evaluation

The index for this independent variable is a linear combination of the Lickert scales for the three questionnaire items that measure attitudes concerning three modes of evaluation: (1) careful unemotional evaluation, (2) prior-to-situation evaluation versus in-situation evaluation, and (3) logical systematic evaluation. The first two modes refer to health matters specifically. The index is normalized so that a value of 1 indicates an extreme negative attitude towards all three modes, while a value of 7 indicates an extreme positive attitude towards all three judgemental modes.

(c) bias in interpretation of colloquial probability statements

The index for this independent variable is a linear combination of the responses for two particular

questionnaire items. One item requires the subject to assign a probability to the colloquial term "certain" while the other requires the subject to assign a probability to the colloquial term "impossible". The index measures the deviation of assigned probabilities from the probabilities of 100 percent and zero percent which are conventionally associated with the terms "certain" and "impossible". The index is normalized so that a value of 0 indicates no deviation while a value of 100 indicates extreme deviation.

(d) discrimination band for probability and day stimuli

Two indices were developed to measure sensitivity towards day stimuli and probability stimuli. The index for day stimuli is a linear combination of Lickert scales for the questionnaire items that measure sensitivity towards various day stimuli. The index for probability stimuli is defined similarly. The two indices are normalized so that a value of 1 indicates extreme insensitivity (wide discrimination band) while a value of 7 indicates extreme sensitivity (narrow discrimination band).

(e) confidence in method

The index for this independent variable is a linear combination of the Lickert scales for the two questionnaire items that provide some measure of confidence in methods: (1) willingness to use elicited preferences in lieu of personal judgements when a choice situation arises, and (2) willingness to revise spontaneous judgements which are contradictory to elicited references. One index representing

confidence in each method was developed. The indices are normalized so that a value of 1 indicates complete non-confidence while a value of 7 indicates complete confidence.

(f) realism of method scenario

The value of the index for this independent variable is identical to the response of the questionnaire item which requires the subject to rate the realism on a Lickert scale (7=extremely realistic to 1=extremely unrealistic). An index representing realism of method scenario was developed for each of the two methods.

In addition to the questionnaire items, four additional independent variables were defined to represent linearity of trade-off curves[24]:

(1) RMT1=($R^2$ for MT curve of first session and first reference point+$R^2$ for MT curve of first session and second reference point)/2.

(2) RVM1=($R^2$ for VM curve of first session and first reference point+$R^2$ for VM curve of first session and second reference point)/2.

(3) RMT2=($R^2$ for MT curve of second session and first reference point+$R^2$ for MT curve of second session and second reference point)/2.

(4) RVM2=($R^2$ for VM curve of second session and first reference point+$R^2$ for VM curve of second session and second

---

[24] These abbreviations will be used in the following definitions and in any later discussions: MT=MacCrimmon-Toda, and VM=von Neumann-Morgenstern.

reference point)/2.

Independent Variables - Profile of Results

Figure 14 summarizes the responses for each of the independent variables for all three groups. The independent variable, characterizing acceptance of "rationality" axioms, indicates a central tendency towards acceptance (median is greater than 4). The medians for the independent variable representing propensities for certain judgemental modes of evaluation, show that at least 50% of the subjects do not display a strong positive attitude. Furthermore, the medians for the independent variables representing confidence in method, show that at least 50% express a neutral or non-confidence in both methods of eliciting preferences; the medians of the independent variables representing realism of scenario for each method, show that at least 50% display neutral belief.

The mean $R^2$ values for the derived indifference curves are close to 1.00 for both undergraduate and graduate groups. It is quite conceivable that: (1) the subjects conscientiously follow a linear rule, or (2) the subjects do not make a conscientious effort to follow a linear rule but the methods of eliciting preferences induce them to provide judgements which relect linearity. The mean $R^2$ for the von Neumann-Morgenstern derived curves are higher than for the corresponding MacCrimmon-Toda derived curves. If the subjects do follow a simple linear rule for response

judgements (suggested by a high $R^2$), the difference in $R^2$ values could be due to the fact that it is computationally more demanding to apply this rule to a choice between pairs of gambles (MacCrimmon-Toda method) than it is to apply this rule to choosing a sure-thing which is judged to be indifferent to a gamble (standard gamble method).

Linear Associations among Independent Variables

Figure 15 displays the Spearman correlation matrix of the independent variables. The correlation (0.50) between acceptance of "rationality" axioms and propensities for judgemental modes of evaluation is significantly high. In other words, acceptance of particular "rationality" axioms in simple preference situations correlates positively with agreement that evaluation of preferences should follow a logical, consistent, and unemotional path of reasoning.

The correlations between confidence in each method of eliciting preferences and the propensities for judgemental modes of evaluation seem to suggest that subjects who agree with the appropriateness of the general "rational" modes of evaluation, also agree with the appropriateness of two particular modes of evaluation, namely the standard gamble method and the MacCrimmon-Toda method. The hypothesis that subjects who are confident in one particular method for eliciting preferences will also be confident in the other method is supported by the significantly positive correlation between confidence in the standard gamble method

| | Undergraduates N=14 | | | | Graduates N=9 | | | | Total N=23 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Median | σ | Range | Mean | Median | σ | Range | Mean | Median | σ | Range |
| ACCEPTANCE OF "RATIONALITY" AXIOMS | 5.2 | 5.6 | 1.4 | 3.0-7.0 | 5.8 | 5.5 | 1.0 | 4.0-7.0 | 5.4 | 5.6 | 1.2 | 3.0-7.0 |
| PROPENSITIES FOR JUDGEMENTAL MODES | 4.8 | 4.5 | 0.9 | 3.6-6.6 | 4.5 | 4.4 | 0.6 | 3.6-5.6 | 4.7 | 4.4 | 0.8 | 3.6-6.6 |
| BIAS IN PROBABILITY INTERPRETATION | 4.4 | 0.4 | 7.4 | 0.0-22.5 | 2.3 | 1.0 | 3.4 | 0.0-10.0 | 3.6 | 0.4 | 6.1 | 0.0-22.5 |
| DISCRIMINATION FOR PROBABILITY | 4.2 | 4.5 | 1.0 | 1.8-6.0 | 3.9 | 3.5 | 0.9 | 2.6-5.8 | 4.1 | 4.2 | 1.0 | 1.8-6.0 |
| DISCRIMINATION FOR DAY | 2.8 | 2.9 | 0.8 | 1.6-4.6 | 3.3 | 3.2 | 1.1 | 2.0-5.0 | 3.0 | 2.9 | 0.9 | 1.6-5.0 |
| CONFIDENCE IN MACCRIMMON-TODA METHOD | 3.3 | 3.2 | 1.1 | 2.0-5.5 | 3.5 | 3.3 | 1.5 | 1.0-6.0 | 3.4 | 3.3 | 1.2 | 1.0-6.0 |
| CONFIDENCE IN STANDARD GAMBLE METHOD | 3.5 | 3.5 | 1.0 | 2.0-6.0 | 3.6 | 3.7 | 1.5 | 1.0-6.0 | 3.5 | 3.6 | 1.2 | 1.0-6.0 |
| REALISM OF MACCRIMMON-TODA METHOD | 4.8 | 5.3 | 1.5 | 1.0-6.0 | 4.5 | 4.6 | 1.5 | 2.0-7.0 | 4.7 | 5.0 | 1.5 | 1.0-7.0 |
| REALISM OF STANDARD GAMBLE METHOD | 3.0 | 3.0 | 1.2 | 1.0-5.0 | 4.6 | 4.5 | 1.4 | 3.0-7.0 | 3.6 | 3.4 | 1.4 | 1.0-7.0 |
| RMT1- LINEARITY OF 1ST SESSION MT CURVES | 0.91 | 0.92 | 0.03 | 0.86-0.98 | 0.94 | 0.96 | 0.07 | 0.76-1.00 | 0.92 | 0.93 | 0.05 | 0.76-1.00 |
| RVM1- LINEARITY OF 1ST SESSION VM CURVES | 0.96 | 0.99 | 0.04 | 0.88-0.99 | 0.99 | 0.99 | 0.00 | 0.97-1.00 | 0.97 | 0.99 | 0.03 | 0.88-1.00 |
| RMT2- LINEARITY OF 2ND SESSION MT CURVES | 0.91 | 0.92 | 0.06 | 0.70-0.97 | 0.96 | 0.96 | 0.03 | 0.91-1.00 | 0.93 | 0.93 | 0.05 | 0.70-1.00 |
| RVM2- LINEARITY OF 2ND SESSION VM CURVES | 0.98 | 0.99 | 0.02 | 0.91-0.99 | 0.98 | 0.98 | 0.01 | 0.96-1.00 | 0.98 | 0.99 | 0.02 | 0.91-1.00 |

Figure 14 - Independent Variables - Profile of Results and confidence in the MacCrimmon-Toda method.

There is a significant linear correlation of acceptance of particular "rationality" axioms with confidence in the MacCrimmon-Toda method but not with confidence in the standard gamble method. It is quite conceivable that those who accept the appropriateness of rational behavior in nonpersonal situations, express a similar attitude in personal situations only to a certain extent, i.e. are confident in the predictions made by the MacCrimmon-Toda method (based upon weak assumptions of behavior) but are not confident in the predictions made on the basis of the standard gamble method (based upon stronger assumptions of behavior).

## Linear Associations between Independent and Dependent Variables

Figure 16 displays the Spearman correlations between the independent and dependent variables.

The significant positive correlation between $(MT2-VM2)^2$ and propensities for rational modes of evaluation is incompatible with our hypothesis since it implies that for those who accept "rational" modes of evaluation, predictions based on alternative methods tend to show less correspondence.

The significant correlation between the discrimination for probability stimuli with $(VM1-VM2)^2$ but not with $(MT1-MT2)^2$ is surprising since the MacCrimmon-Toda method

| N=23 | ACCEPTANCE OF "RATIONALITY" AXIOMS | PROPENSITIES FOR JUDGEMENTAL MODES | BIAS IN INTERPRETATION OF PROBABILITY | DISCRIMINATION FOR PROBABILITY | DISCRIMINATION FOR DAYS | CONFIDENCE IN MACCRIMMON-TODA METHOD | CONFIDENCE IN STANDARD GAMBLE METHOD | REALISM OF MACCRIMMON-TODA METHOD SCENARIO | REALISM OF STANDARD GAMBLE METHOD SCENARIO |
|---|---|---|---|---|---|---|---|---|---|
| ACCEPTANCE OF "RATIONALITY" AXIOMS | | +.50 | | | +.30 | +.38 | | | +.52 |
| PROPENSITIES FOR JUDGEMENTAL MODES | | | | | +.51 | +.69 | +.50 | | |
| BIAS IN INTERPRETATION OF PROBABILITY | | | | | | | | | |
| DISCRIMINATION FOR PROBABILITY | | | | | | | | | |
| DISCRIMINATION FOR DAYS | | | | | | +.33 | | | +.29 |
| CONFIDENCE IN MACCRIMMON-TODA METHOD | | | | | | | +.83 | | +.31 |
| CONFIDENCE IN STANDARD GAMBLE METHOD | | | | | | | | | |
| REALISM OF MACCRIMMON-TODA METHOD SCENARIO | | | | | | | | | |
| REALISM OF STANDARD GAMBLE METHOD SCENARIO | | | | | | | | | |

Only correlations significantly nonzero at 0.10 confidence level are shown.

Figure 15 - Spearman Correlation Matrix for Independent Variables

requires subjects to make judgements upon choices involving

various levels of probability stimuli, while the standard gamble method requires subjects to make judgements upon choices involving only the probability 1/2.

The significant correlations of the inconsistency measures with the linearity measures seem to suggest that linearity is the dominant feature affecting consistency. The adoption of a simple rule for combining attributes of outcomes reduces computational errors in evaluating preferences.

## The Questionnaire Items

Through the use of independent variables, questionnaire items of conceptual similarity, were aggregated to form one composite score or index to represent the common concept. However, the use of an aggregated score to define an independent variable may disguise: (1) responses to individual questionnaire items which comprise that independent index (or variable), (2) associations between responses of questionnaire items comprising that index, (3) associations between responses of individual questionnaire items comprising that index with responses of individual questionnaire items comprising other indices, and/or (4) associations between individual questionnaire items and inconsistency measures (dependent variables). To circumvent these objections, the following discussions will present (1) a profile of questionnaire item responses, (2) a correlation analysis performed between questionnaire item responses, and

| N=23 | $(MT1-MT2)^2$ | $(VM1-VM2)^2$ | $(MT1-VM1)^2$ | $(MT2-VM2)^2$ | $(av(MT)-av(VM))^2$ |
|---|---|---|---|---|---|
| ACCEPTANCE OF "RATIONALITY" AXIOMS | | | | | |
| PROPENSITIES FOR JUDGEMENTAL MODES | | | | +.35 | |
| BIAS IN INTERPRETATION OF PROBABILITY | | | -.33 | | -.31 |
| DISCRIMINATION FOR PROBABILITY | | -.33 | | | |
| DISCRIMINATION FOR DAYS | | | -.38 | | -.47 |
| CONFIDENCE IN MACCRIMMON-TODA METHOD | | | | | |
| CONFIDENCE IN STANDARD GAMBLE METHOD | | | | +.28 | +.27 |
| REALISM OF MACCRIMMON-TODA METHOD SCENARIO | | | | | |
| REALISM OF STANDARD GAMBLE METHOD SCENARIO | | | | -.46 | |
| RMT1- LINEARITY OF 1ST SESSION MT CURVES | -.50 | | -.83 | | -.68 |
| RVM1- LINEARITY OF 1ST SESSION VM CURVES | -.45 | -.40 | -.80 | -.31 | -.66 |
| RMT2- LINEARITY OF 2ND SESSION MT CURVES | -.54 | -.40 | | -.78 | -.41 |
| RVM2-LINEARITY OF 2ND SESSION VM CURVES | | -.83 | | | |

Only correlations significantly nonzero at the 0.10 confidence level are shown.

Figure 16 - Spearman Correlations between Independent and Dependent Variables

(3) a correlation analysis performed between the

questionnaire item responses and the inconsistency measures (dependent variables).

Questionnaire Responses - Profile of Results

Figure 17 displays the profile of questionnaire responses for each of three groups. The most striking feature is the large range of responses for most questionnaire items. Thus, although the students of the undergraduate and the graduate groups have educational backgrounds similar to members of their group, attitudes are not homogeneous within any group.

The median responses for the confidence in method items fall on the negative attitude half of the Lickert scale.

The profile for the probability discrimination band indicates that the median responses (7=extremely significant to 1=extremely insignificant) increases as the probability difference increases. However, this relationship is not compatible with the profile for the day discrimination band. Thus, the median perceived significance of a probability stimuli increases with increasing stimuli differences while the same does not hold for day stimuli.

The median interpretations of "certain" and "impossible" are close to 100% and 0% respectively. However, the profile indicates that there were extreme responses of 70% and 35% for "certain" and "impossible" respectively.

| | N=23 | Undergraduates N=14 | | | | Graduates N=9 | | | | Total N=23 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | mean | median | σ | range | mean | median | σ | range | mean | median | σ | range |
| ACCEPTANCE OF "RATIONALITY" AXIOMS | transitivity | 5.0 | 5.5 | 1.7 | 2-7 | 5.8 | 6.1 | 1.2 | 3-7 | 5.3 | 5.8 | 1.6 | 2-7 |
| | sure-thing | 5.2 | 6.0 | 2.0 | 1-7 | 5.8 | 5.8 | 1.0 | 5-7 | 5.4 | 5.7 | 1.7 | 1-7 |
| | compound-gamble | 5.5 | 6.0 | 1.7 | 1-7 | 5.0 | 5.0 | 1.8 | 2-7 | 5.3 | 5.6 | 1.7 | 1-7 |
| PROPENSITIES FOR JUDGEMENTAL MODES | unemotional evaluation | 4.8 | 5.1 | 1.5 | 2-7 | 5.0 | 5.2 | 1.4 | 3-7 | 4.9 | 5.1 | 1.3 | 2-7 |
| | prior-to-situation evaluation | 4.0 | 4.0 | 1.7 | 2-7 | 3.5 | 4.0 | 1.9 | 1-6 | 3.8 | 4.0 | 1.8 | 1-7 |
| | logical systematic evaluation | 4.5 | 4.8 | 1.3 | 2-7 | 5.1 | 5.2 | 1.6 | 3-7 | 5.4 | 5.4 | 1.4 | 3-7 |
| REALISM OF SCENARIOS | realism of MT scenario | 4.8 | 5.3 | 1.5 | 1-6 | 4.5 | 4.6 | 1.5 | 2-7 | 4.7 | 5.0 | 1.5 | 1-7 |
| | realism of VM scenario | 3.0 | 3.0 | 1.2 | 1-5 | 4.6 | 4.5 | 1.4 | 3-7 | 3.6 | 3.4 | 1.4 | 1-7 |
| CONFIDENCE IN METHOD | elicited in lieu of personal judgements (MT) | 3.4 | 3.1 | 1.3 | 2-7 | 3.3 | 2.7 | 2.0 | 1-7 | 3.4 | 3.0 | 1.6 | 1-7 |
| | revision of "inconsistenties" (MT) | 3.3 | 3.1 | 1.4 | 1-6 | 3.6 | 3.7 | 1.6 | 1-6 | 3.4 | 3.2 | 1.5 | 1-6 |
| | elicited in lieu of personal judgements (VM) | 3.7 | 3.5 | 1.6 | 1-3 | 3.3 | 2.7 | 2.0 | 1-7 | 3.5 | 3.1 | 1.7 | 1-7 |
| | revision of "inconsistencies" (VM) | 3.4 | 3.1 | 1.6 | 1-6 | 3.8 | 4.0 | 1.7 | 2-7 | 3.6 | 3.4 | 1.6 | 1-6 |
| DISCRIMINATION FOR PROBABILITY | 0.05 vs 0.1 | 3.2 | 2.9 | 1.9 | 1-7 | 2.7 | 2.0 | 2.1 | 1-7 | 3.0 | 2.7 | 1.9 | 1-7 |
| | 0.5 vs 0.55 | 3.1 | 3.2 | 1.2 | 1-5 | 3.3 | 3.3 | 1.0 | 2-5 | 3.2 | 3.2 | 1.1 | 1-5 |
| | 0.9 vs 0.95 | 3.7 | 3.5 | 1.8 | 1-6 | 3.8 | 4.0 | 1.7 | 1-6 | 3.8 | 3.7 | 1.8 | 1-6 |
| | 0.1 vs 0.2 | 4.6 | 4.8 | 1.1 | 2-6 | 3.7 | 3.2 | 1.8 | 2-7 | 4.3 | 4.5 | 1.4 | 2-7 |
| | 0.3 vs 0.5 | 5.3 | 5.6 | 1.3 | 2-7 | 4.6 | 4.7 | 1.5 | 3-7 | 5.0 | 5.4 | 1.4 | 2-7 |
| | 0.6 vs 0.8 | 5.5 | 5.7 | 1.3 | 3-7 | 5.0 | 5.0 | 1.1 | 3-7 | 5.3 | 5.3 | 1.2 | 3-7 |
| DISCRIMINATION FOR DAY | 1/4 vs 1/2 | 1.9 | 1.3 | 1.6 | 1-7 | 2.8 | 2.2 | 2.2 | 1-7 | 2.3 | 1.6 | 1.8 | 1-7 |
| | 2 vs 2¼ | 2.0 | 2.1 | 0.9 | 1-4 | 3.1 | 2.3 | 2.0 | 1-6 | 2.4 | 2.1 | 1.5 | 1-6 |
| | 15 vs 14½ | 1.8 | 1.2 | 1.3 | 1-5 | 2.2 | 2.0 | 1.2 | 1-4 | 2.0 | 1.4 | 1.2 | 1-5 |
| | 4 vs 5 | 4.3 | 4.7 | 1.1 | 2-6 | 4.5 | 4.8 | 1.0 | 3-6 | 4.4 | 4.7 | 1.0 | 2-6 |
| | 10 vs 9 | 4.0 | 4.0 | 1.4 | 2-6 | 3.8 | 4.0 | 1.0 | 2-5 | 4.0 | 4.0 | 1.3 | 2-6 |
| BIAS IN INTERPRETATION OF PROBABILITY | assign probability to "certain" | 94.5 | 99.4 | 10.7 | 70-100 | 96.1 | 98.7 | 6.9 | 80-100 | 95.1 | 99.4 | 1.9 | 70-100 |
| | assign probability to "impossible" | 3.2 | 0.1 | 9.5 | 0-35 | 0.5 | 0.3 | 1.6 | 0-5 | 2.2 | 0.1 | 7.5 | 0-35 |

Figure 17 - Questionnaire Responses - Profile of Results

Linear Associations among the Questionnaire Responses

Figure 18 displays the correlation matrix of the questionnaire item responses.

The two significant correlations among the questionnaire items characterizing acceptance of "rationality" axioms indicates that those who expressed strong agreement with the axiom of riskless behavior (transitivity) also expressed strong agreement with the axioms of risky behavior (sure-thing and compound-gamble).

There is no significant linear correlation between acceptance of the sure-thing axiom and acceptance of the compound-gamble axiom. Although the existence of a significant non-linear relationship is possible, it is also quite conceivable that persons who related to the simplicity of the sure-thing axiom may not necessarily have been able to relate to the relative complexity of the compound-gamble axiom.

It is interesting to note a significant positive correlation between acceptance of the transitivity axiom and the need for revision of spontaneous choice (which is found to be contradictory to elicited preferences) for the case of the MacCrimmon-Toda method which is based primarily upon the assumed transitivity of preferences. This association lends credence to Savage's (1954) definition of rationality that people accept rationality if they are willing to revise choices which are "irrational".

The abundance of significant positive correlations among sensitivities towards stimuli of various probability differences may suggest that subjects maintain their relative (to other subjects) sensitivity over all stimuli levels. The intercorrelations among sensitivities towards stimuli of various day differences suggests a similar result for day as for probability stimuli.

This figure also shows significant entries of the correlation matrix for sensitivities toward day stimuli (column of matrix) with sensitivities toward probability stimuli (row of matrix) which were previously hidden by the apparent nonassociation between the independent variables representing each stimulus. These correlations may have significance for the MacCrimmon-Toda method in which subjects were required to judge pairs of day and probability stimulus.

Linear Associations - Questionnaire Responses & Dependent Variables

The results displayed in Figure 19 do not indicate any significant linear correlations between any particular "rationality" axioms with any inconsistency measures. Although these results do not prelude the lack of any significant relationship (a non-linear association is possible), it is conceivable that subjects who agree with "rationality" may not necessarily make judgements conforming to "rationality". If, in fact, this explanation is the

N=23

Only correlations significantly nonzero at the 0.10 confidence level are shown.

| | | ACCEPTANCE OF "RATIONALITY" AXIOMS | | | PROPENSITIES FOR JUDGEMENTAL MODES | | | REALISM OF SCENARIOS | | CONFIDENCE IN METHOD | | | | DISCRIMINATION FOR PROBABILITY | | | | | | DISCRIMINATION FOR DAY | | | | | | BIAS IN INTERPRETATION OF PROBABILITY | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | transitivity | sure-thing | compound-gamble | unemotional evaluation | prior-to-situation evaluation | logical systematic evaluation | realism of MacCrimmon-Toda scenario | realism of standard gamble method scenario | elicited in lieu of personal judgements (MT) | revision of "inconsistencies" (MT) | elicited in lieu of personal judgements (VM) | revision of "inconsistencies" (VM) | 0.05 vs 0.1 | 0.5 vs 0.55 | 0.9 vs 0.95 | 0.1 vs 0.2 | 0.3 vs 0.5 | 0.6 0.8 | 1/2 vs 1/4 | 2 vs 2½ | 15 vs 14½ | 4 vs 5 | 10 vs 9 | assignment of probability to "certain" | assignment of probability to "impossible" |
| ACCEPTANCE OF "RATIONALITY" AXIOMS | transitivity | | +.69 | +.39 | +.48 | | | +.54 | | +.49 | | +.38 | | | +.30 | | | | +.31 | +.30 | | +.32 | | | | |
| | sure-thing | | | | | | | | | | | | | | +.48 | | | | | | | | | | | |
| | compound-gamble | | | | +.54 | | +.28 | | | -.62 | | | | | | | | | +.48 | | | | | +.34 | | |
| PROPENSITIES FOR JUDGEMENTAL MODES | unemotional evaluation | | | | | | +.33 | | +.32 | | | | | | | | | | +.40 | | | | | | | |
| | prior-to-situation evaluation | | | | | | | | | +.49 | -.33 | | | | -.34 | | | | | | | | | +.33 | | |
| | logical systematic evaluation | | | | | | | +.33 | | -.31 | +.32 | -.40 | +.28 | | +.16 | +.28 | | | | +.46 | +.58 | | +.50 | | | -.55 |
| REALISM OF SCENARIOS | realism of MacCrimmon-Toda scenario | | | | | | | | | | | | | | | | | | | | | | | | | -.29 |
| | realism of standard gamble method scenario | | | | | | | | | | | | | -.37 | | | -.30 | | | +.32 | +.31 | +.33 | -.30 | | | |
| CONFIDENCE IN METHOD | elicited in lieu of personal judgements (MT) | | | | | | | | | | | +.59 | | | | | | | | -.41 | -.28 | | | | | |
| | revision of "inconsistencies" (MT) | | | | | | | | | | | | +.93 | -.35 | | +.21 | -.35 | | | +.31 | +.34 | +.33 | +.35 | | | |
| | elicited in lieu of personal judgements (VM) | | | | | | | | | | | | | | | | | | | -.42 | -.31 | | | | -.33 | |
| | revision of "inconsistencies" (VM) | | | | | | | | | | | | | | | | | | | | | | | | | |
| DISCRIMINATION FOR PROBABILITY | 0.05 vs 0.1 | | | | | | | | | | | | | +.70 | +.32 | | | +.69 | +.30 | +.32 | | | | | | |
| | 0.5 vs 0.55 | | | | | | | | | | | | | | | +.58 | +.38 | | +.30 | +.34 | +.19 | | +.36 | | | |
| | 0.9 vs 0.95 | | | | | | | | | | | | | | | | | | +.32 | | | | +.39 | +.23 | | |
| | 0.1 vs 0.2 | | | | | | | | | | | | | | | | | +.74 | +.33 | | | | | | | |
| | 0.3 vs 0.5 | | | | | | | | | | | | | | | | | | +.11 | | | | | | | |
| | 0.6 0.8 | | | | | | | | | | | | | | | | | | | | -.29 | | | | | |
| DISCRIMINATION FOR DAY | 1/2 vs 1/4 | | | | | | | | | | | | | | | | | | | | +.76 | +.48 | +.45 | | | |
| | 2 vs 2½ | | | | | | | | | | | | | | | | | | | | | +.45 | +.67 | | | -.32 |
| | 15 vs 14½ | | | | | | | | | | | | | | | | | | | | | | +.39 | +.33 | | |
| | 4 vs 5 | | | | | | | | | | | | | | | | | | | | | | | +.68 | | -.47 |
| | 10 vs 9 | | | | | | | | | | | | | | | | | | | | | | | | +.97 | -.35 |
| BIAS IN INTERPRETATION OF PROBABILITY | assignment of probability to "certain" | | | | | | | | | | | | | | | | | | | | | | | | | |
| | assignment of probability to "impossible" | | | | | | | | | | | | | | | | | | | | | | | | | |

Figure 18 - Spearman Correlation Matrix for Questionnaire Responses

actual case, the result demonstrates that persons who want

to act rationally may not actually act rationally. This result lends credence to Raiffa's (1961) rationale for a normative utility theory. Briefly, Raiffa asserts that utility theory should provide norms of behavior for those people who want to act rationally but do not.

The density of significant entries in the correlation matrix for sensitivities toward day stimuli versus the relative sparsity of the corresponding matrix for probability stimuli may suggest that subjects allow consideration of the day stimuli to dominate their choice when required to make judgements upon choices involving probability and day, i.e. the outcome hospital days has more effect upon choice than does the probability of outcome.

Comparison of Scores for Inconsistency Measures

To compare the methods used, we have tested for significance of differences between scores for inconsistency measures (test-retest, intermethod, and averaged intermethod). In order to compare inconsistency scores, each subject was used as his own control; the Wilcoxon matched-pairs signed-ranks test was employed to test for differences between: (1) MacCrimmon-Toda test-retest inconsistency and standard gamble test-retest inconsistency, (2) test-retest inconsistency (for each method) and averaged intermethod inconsistency, and (3) intermethod inconsistency (for each session) and averaged intermethod inconsistency.

To test for differences in inconsistency measures given

| | | $(MT1-MT2)^2$ | $(VM1-VM2)^2$ | $(MT1-VM1)^2$ | $(MT2-VM2)^2$ | $(av(MT)-av(VM))^2$ |
|---|---|---|---|---|---|---|
| ACCEPTANCE OF "RATIONALITY" AXIOMS | transitivity | | | | | |
| | sure-thing | | | | | |
| | compound-gamble | | | | | |
| PROPENSITIES FOR JUDGEMENTAL MODES | unemotional evaluation | | -.44 | -.28 | | |
| | prior-to-situation evaluation | | | +.31 | +.35 | +.45 |
| | logical systematic evaluation | | | -.36 | | -.45 |
| REALISM OF SCENARIOS | realism of MT scenario | | | | | |
| | realism of VM scenario | | | | -.46 | |
| CONFIDENCE IN METHOD | elicited in lieu of personal judgements (MT) | | | | -.28 | -.38 |
| | revision of "inconsistenties" (MT) | | | | | |
| | elicited in lieu of personal judgements (VM) | | | -.25 | -.23 | -.46 |
| | revision of "inconsistencies" (VM) | | | | | |
| DISCRIMINATION FOR PROBABILITY | 0.05 vs 0.1 | | | | | |
| | 0.5 vs 0.55 | | | | | |
| | 0.9 vs 0.95 | | | | | -.38 |
| | 0.1 vs 0.2 | | | | | |
| | 0.3 vs 0.5 | | -.47 | | | |
| | 0.6 vs 0.8 | | -.60 | | | |
| DISCRIMINATION FOR DAY | 1/4 vs 1/2 | -.29 | | -.34 | -.29 | -.43 |
| | 2 vs $2\frac{1}{2}$ | | | -.39 | -.40 | -.63 |
| | 15 vs $14\frac{1}{3}$ | | | | | -.30 |
| | 4 vs 5 | | +.31 | -.38 | | -.38 |
| | 10 vs 9 | | | | | |
| BIAS IN INTERPRETATION OF PROBABILITY | assign probability to "certain" | | | +.49 | | +.42 |
| | assign probability to "impossible" | | | +.49 | | +.42 |
| LINEARITY | RMT1 | -.50 | | -.82 | | -.68 |
| | RVM1 | -.45 | -.40 | -.80 | -.31 | -.66 |
| | RMT2 | -.54 | -.40 | | -.78 | -.41 |
| | RVM2 | | -.83 | | | |

Only correlations significantly nonzero at the 0.10 confidence level are shown.

Figure 19 - Spearman Correlations - Questionnaire Responses & Dependent Variables

particular probability ranges and reference points, we have

sampled points from each of four probability ranges: (1) 0.00-0.25, (2) 0.25-0.50, (3) 0.50-0.75, and (4) 0.75-1.00, and evaluated respective differences between indifference curves associated with particular reference points. The terms $(MT1-MT2)^2ij$, $(VM1-VM2)^2ij$, $(MT1-VM1)^2ij$, $(MT2-VM2)^2ij$, and $(av(MT)-av(VM))^2ij$ are defined as sum of squared deviations between values of days associated to the two indifference curves with reference point j [ j=1 for (4,0.50) reference point and j=2 for (7,0.50) ] and to sampled points from probability region i [i=1,2,3,4 correspond to probability regions 0.00-0.25, 0.25-0.50, 0.50-0.75, and 0.75-1.00 respectively].

MacCrimmon-Toda vs Standard Gamble Test-retest

Figure 20 displays the statistical results of the one-tailed Wilcoxon test of the null hypothesis that $(VM1-VM2)^2ij=(MT1-MT2)^2ij$. In 5 of the 8 sample classes ij, the test indicates that the standard gamble method test-retest predicted preferences are more consistent than the test-retest preferences predicted by the MacCrimmon-Toda method. One suspects that differences in consistency could be due to differences in difficulty of judgements required by each method. In fact, 57% of the total sample felt that judgements required from them in the standard gamble method were easier to make, 30% felt that the judgements in the MacCrimmon-Toda method were easier to make while 13% felt that the judgements in both methods were of equal difficulty. Furthermore, the results discussed previously

show that the average $R^2$, a measure of linearity, of the standard gamble curves is higher than that of the MacCrimmon-Toda curves. These two results seem to suggest that ease of judgements and linearity contribute to consistency.

N=23
Only entries which indicate a significant difference at the .005 confidence level are shown.
Each entry is a Wilcoxon sum of similar-signed ranks.

| Probability region i | Reference Point j | $(MT1-MT2)^2$ vs $(VM1-VM2)^2$ |
|---|---|---|
| 1 | 1 | |
| 2 | 1 | 52 |
| 3 | -1 | 22 |
| 4 | 1 | 61 |
| 1 | 2 | |
| 2 | 2 | 40 |
| 3 | 2 | 45 |
| 4 | 2 | |

$(VM1-VM2)^2 < (MT1-MT2)^2$ for significant entries in the column

Figure 20 - MacCrimmon-Toda vs Standard Gamble Test-retest

Test-retest (for each method) vs Averaged Intermethod

Figure 21 displays the statistical results of the one-tailed Wilcoxon test of the hypothesis that $(MT1-MT2)^2ij=(av(MT)-av(VM))^2ij$. In the sample classes ij from the probability ranges of 0.50-0.75 and 0.75-1.00, and the indifference curve associated with the first reference point, test-retest differences in preferences predicted from the MacCrimmon-Toda method are significantly smaller than the corresponding differences between indifference curves obtained for the same reference point using differences between average curves (each average curve is obtained by applying one of the methods over two experimental sessions and averaging the two curves).

The following figure also displays the statistical results of the one-tailed Wilcoxon test of the null hypothesis that $(VM1-VM2)^2ij=(av(MT)-av(VM))^2ij$. In 6 of the 8 sample classes ij, test-retest preferences predicted from the standard gamble method are significantly more consistent than the averaged intermethod correspondence. One may conclude that the correspondence of predictions obtained from repeat application of each method is generally higher than correspondence between predictions obtained using different methods.

N=23
Only entries which indicate a significant difference at the .005 confidence level are shown.
Each entry is a Wilcoxon sum of similar-signed ranks.

| Probability Region i | Reference Point j | $(av(MT)-av(VM))^2$ vs $(VM1-VM2)^2$ | $(av(MT)-av(VM))^2$ vs $(MT1-MT2)^2$ |
|---|---|---|---|
| 1 | 1 | | |
| 2 | 1 | 34 | |
| 3 | 1 | 32 | 67 |
| 4 | 1 | | 67 |
| 1 | 2 | 66 | |
| 2 | 2 | 24 | |
| 3 | 2 | 28 | |
| 4 | 2 | 58 | |

$(VM1-VM2)^2 < (av(MT)-av(VM))^2$ for significant entries in first column
$(MT1-MT2)^2 < (av(MT)-av(VM))^2$ for significant entries in second column

Figure 21 - Test-retest (for each method) vs Averaged Intermethod

Intermethod (for each session) vs Averaged Intermethod

Figure 22 displays the statistical results of the one-tailed Wilcoxon test of the null hypothesis that $(MT1-VM1)^2ij=(av(MT)-av(VM))^2ij$. Only in two sample classes is averaged intermethod correspondence higher than intermethod preferences predicted from the first experimental session.

The following figure also displays the statistical results of the one-tailed Wilcoxon test of the null hypothesis that $(MT^2-VM2)^2ij=(av(MT)-av(VM))^2ij$. Only in one sample class is averaged intermethod correspondence higher than intermethod preferences predicted from the second

experimental session. One may conclude that averaging of repeat results for each method generally contribute marginally to the correspondence of predictions obtained from each method.

N=23
Only entries which indicate a significant difference at the .005 confidence level are shown.
Each entry is a Wilcoxon sum of similar-signed ranks.

| Probability region i | Reference Point j | $(av(MT)-av(VM))^2$ vs $(MT1-VM1)^2$ | $(av(MT)-av(VM))^2$ vs $(MT2-VM2)^2$ |
|---|---|---|---|
| 1 | 1 | | |
| 2 | 1 | | |
| 3 | 1 | 47 | |
| 4 | 1 | 73 | 68 |
| 1 | 2 | | |
| 2 | 2 | | |
| 3 | 2 | | |
| 4 | 2 | | |

$(av(MT)-av(VM))^2 < (MT1-VM1)^2$ for significant entries in first column
$(av(MT)-av(VM))^2 < (MT2-VM2)^2$ for significant entries in second column

Figure 22 - Intermethod (for each session) vs Averaged Intermethod

Goodness of Prediction

An average of 4 pairs of diammetrically opposed gambles were presented to each subject[25]. The results showed that an average of 72% of the preferences expressed by each subject were in accord with predictions made by the MacCrimmon-Toda method.

---

[25] From the day X probability choice space, we sampled points which lie between the indifference area derived by the MacCrimmon-Toda method in the second session and the indifference curve derived from the standard gamble method in the second session. Pairs of gambles consisting of these points and the associated reference points were presented to the subject in the third session.

Bibliography

Ackoff, R.L., S.K. Gupta, and J.S. Minas. Scientific Method: Optimizing Applied Research Decisions John Wiley & Sons, 1962.

Adams, E.W. Survey of Bernoullian utility theory. In: Mathematical Thinking in the Measurement of Behavior editor: Solomon, H. The Free Press of Glencoe, Illinois. 155-268, 1959.

Alberoni, F. Contribution to the study of subjective probability I. The Journal of General Psychology 66:241-264, 1962.

Alchian, A.A. The meaning of utility measurement. American Economic Review 43:26-50, 1953.

Beach, B. Direct and indirect methods for measuring utility. University of Washington, Department of Psychology, Research report, unpublished, 1972.

Becker, S.W. and S. Siegel. Utility and level of aspiration. American Journal of Psychology 75:115-120, 1962.

Bentham, J. An introduction to the Principles of Morals and Legislation Oxford: The Clarendon Press, 1907.

Bernoulli, D. Speciman theoriae novae de mensura sortis. Commentarii Academiae Scientiarum Imperiales Petropolitanae 5:175-192, 1738. (Translated by L. Somer in Econometrica 22:23-36, 1954.)

Bridgman, P.W. The Logic of Modern Physics New York, 1927.

Buel, W.D. A simplification of Hay's method of recording paired comparisons. Journal of Applied Psychology 44:347-348, 1960.

Churchman, C.W. and R.L. Ackoff. An approximate measure of value. Operations Research 2:172-187, 1954.

Churchman, C.W. Prediction and Optimal Decision: Philosophical Issues of a Science of Values Prentice-Hall, Englewood Cliffs, N.J., 1961.

Comrey, A.L. A proposed method for absolute ratio scaling. Psychometrika 15:317-325, 1950.

Coombs, C.H. Psychological scaling without a unit of measurement. Psychological Review 57:145-158, 1950.

Coombs, C.H. and D.C. Beardslee. Decision making under uncertainty. In: Decision Processes editors: Thrall, R.M., C.H. Coombs, and R.L. Davis. John Wiley & Sons, New York, 1954.

Coombs, C.H. and S.C. Komorita. Measuring utility of money through decisions. American Journal of Psychology 71:383-389, 1958.

Davidson, D., J.C.C. McKinsey, and P. Suppes. Outlines of a formal theory of value I Philosophy of Science 22:140-160, 1955.

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

DeSoto, C.B. The predilection for single orderings. Journal of Abnormal Social Pschology 62:16-23, 1961.

Dudek, F.J. and K.E. Baker. The constant-sum method applied to scaling subjectivve dimensions. American Journal of Psychology 69:616-624, 1956.

Eckenrode, R.T. Weighting multiple criteria. Management Science 12:180-192, 1965.

Edgeworth, F.Y. Mathematical Psychics London: Kegan Paul, 1881.

Edwards, W. Probability-preferences in gambling. American Journal of Psychology 66:349-364, 1953.

Edwards, W. Variance preferences in gambling American

Journal of Psychology 67:441-452, 1954a.

Edwards, W. The reliability of probability-preferences. American Journal of Pyschology 67:68-95, 1954b.

Edwards, W. Probability preferences among bets with differing expected values. American Journal of Psychology 67:56-67, 1954c.

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

Edwards, W., H. Lindman, and L.D. Phillips. Emerging technologies in making decisions. In: New Directions in Psychology II New York: Holt, Rinehart, and Winston, 265-325, 1965.

Edwards, W. and L. Phillips. Conservatism in a simple probability inference task. Journal of Experimental Psychology 72:346-354, 1966.

Ellsberg, D. Classic and current notions of measurable utility. Economic Journal 64:528-556, 1954.

Farrer, D.N. Predictive stability of the higher ordered metric scale. American Journal of Psychology 77:120-122, 1964.

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

Fishburn, P.C. Independence, trade-offs, and transformations in bivariate utility functions. Management Science 11:792-801, 1965b.

Fishburn, P.C. A note on recent developments in additive utility theories for multiple-factor situations. Operations Research 14, 1966.

Fishburn, P.C. Methods of estimating additive utilities. Management Science 13(7):435-453, 1967.

Fishburn, P.C. Utility theory. Management Science 14(5):335-378, 1968.

Galanter, E. The direct measurement of utility and subjective probability. American Journal of Psychology 75:208-220, 1962.

Glaze, J.A. The association value of nonsense syllables. Journal of Genetic Psychology 35:255-267, 1928.

Goode, W.J. and P.K. Hatt. Methods in Social Research McGraw-Hill, 1952.

Gustafson, D.H., N.P. Kneppreth, and R.P. Leifer. The

assessment of worth: techniques (draft B). Paper submitted for presentation at HSR Conference on Health Status Index, 1972.

Hausner, M. Multidimensional utilities. In: Decision Processes editors: Coombs, C.H., R.L. Davis, and R.M. Thrall. John Wiley & Sons, New York, 167-180, 1954.

Hausner, M. and J.G. Wendel. Ordered vector spaces. Proceedings of the American Mathematical Society 3:977-982, 1952.

Hicks, J.R. and R.G.D. Allen. A reconsideration of the theory of value. Economica 1:52-75,196-219, 1934.

Hoepfl, R.I. and G. Huber. A study of self-explicated utility models. Behavioral Science 15:408-416, 1970.

Huber, G., V. Sahney, and D. Ford A study of subjective evaluation models. Behavioral Science 14:483-489, 1969.

Hurst, P.M. and S. Siegel. Predictions of decisions from a higher ordered metric scale of utility. Journal of Experimental Psychology 52:138-144, 1956.

Klahr, D. Decision making in a complex environment: the use of similarity judgements to predict preferences. Management Science 15(11):595-617, 1969.

Krantz, D.H., R.D. Luce, P. Suppes, and A. Tversky Foundations of Measurement Volume I New York: Academic Press, 1972.

Lewis, C.I. Analysis of Knowledge and Valuation LaSalle, Illinois, 1946.

Lichtenstein, S. and J.R. Newman. Empirical scaling of common verbal phrases associated with numerical probabilities. Psychonomic Science 9(10):563-564, 1967.

MacCrimmon, K.R. and M. Toda. The experimental determination of indifference curves. Review of Economic Studies 36:433-451, 1969.

Marschak, J. Rational behavior, uncertain prospects, and measurable utility. Econometrica 18:111-141, 1950.

Marschak, J. Actual versus consistent decision behavior. Behavioral Science 9(2):103-110, 1964.

May, K.O. The intransitivity of individual preferences. Econometrica 21:476, 1953.

May, K.O. Intransitivity, utility, and the aggregation of preference patterns. Econometrica 22:1-13, 1954.

Messick, S.J. An empirical method of multidimensional

successive intervals. Psychometrika 21:367-376, 1956.

Metfessel, M. A proposal for quantitative reporting of comparative judgements. Journal of Psychology 24:229-235, 1947.

Michalos, A.C. Postulates of rational preference. Philosophy of Science 34:18-22, 1967.

Miller, G.A. The magical number seven, plus or minus two. Psychological Review 63:81-97, 1956.

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

Osgood, C.S., G.J. Suci, and P.H. Tannenbaum. The Measurement of Meaning Urbana, Illinois: University of Illinois Press, 1957.

Papandreou, A.G. An experimental test of an axiom in the theory of choice. Econometrica 21:477(abstract), 1953.

Pareto, V. Manualedi Economia Politica 1906.

Perry, R.B. General Theory of Value Cambridge, Massachusetts, 1950.

Philipps, J.P.N. On the presentation of stimulus-objects in

the method of paired comparison. American Journal of Psychology 77:660-664, 1964.

Preston, M.G. And P. Baratta An experimental study of the auction value of an uncertain outcome. American Journal of Psychology 61:183-193, 1948.

Raiffa, H. Risk, ambiguity, and the Savage axioms: comment. Quarterly Journal of Economics 75:690-694, 1961.

Ramsey, F.P. Truth and probability. In: The Foundations of Mathematics and Logical Essays by F.P. Ramsey. New York: Harcourt Brace, 1931.

Rose, A.M. A study of irrational judgements. The Journal of Political Economy 394-402, 1957.

Ross, R.T. Optimum order for the presentation of pairs in the method of paired comparisons. Journal of Educational Psychology 25:375-382, 1934.

Rousseas, S.W. and A.G. Hart. Experimental verification of a composite indifference map. Journal of Political Economy 59:288-318, 1951.

Samuelson, P.A. A note on the pure theory of consumer's behavior. Economica 5:61-71,353-354, 1938.

Sarbin, J.R. A contribution to the study of actuarial and individual methods of prediction. American Journal of Sociology 48:593-602, 1942.

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

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

Siegel, S. A method for obtaining an ordered metric scale. Psychometrika 21:207-216, 1956.

Siegel, S. Non-parametric Statistics: For the Behavioral Sciences McGraw-Hill, 1956.

Smedslund, J. Multiple-probability Learning Oslo: Akademisk Forlag, 1955.

Stevens, S.S. Measurement, psychophysics, and utility. In: Measurement: Definitions and Theories editors: Churchman, C.W. and P. Ratoosh. John Wiley & Sons, New York, 1959.

Suits, D.B. Use of dummy variables in regression equations. Journal of the American Statistical Association 52:548-551, 1957.

Suppes, P. and W. Winet. An axiomization of utility based upon the notion of utility differences. Management Science 1:259-270, 1955.

Suzuki, G. Procurement and allocation of naval electronic equipments. Naval Research Logistics Quarterly 4:1-7, 1957.

Swalm, R.O. Utility theory-insights into risk taking. Harvard Business Review 123-136, Nov-Dec 1966.

Thrall, R.M. Applications of multi-dimensional utility theory. In: Decision Processes editors: Coombs, C.H., R.L. Davis, and R.M. Thrall. John Wiley & Sons, New York, 1954.

Thurstone, L.L. The indifference function. Journal of Social Psychology 2:139-167, 1931.

Thurstone, L.L. Personal communication with W. Edwards, 1953.

Torgerson, W.S. Theory and Methods of Scaling John Wiley & Sons, 1958.

Torgerson, W.S. Multidimensional scaling of similarity. Psychometrika 30(4):379-393, 1965.

Tullock, G. The irrationality of intransitivity. Oxford Economics Papers 16:401-406, 1964.

Tyszynski, H. Comparisons between increments of "utility". Economic Journal 258-263, 1954.

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

Vroom, V.H. and E. Deci. The stability of post-decision dissonance: a follow-up study of the job attitudes of business school graduates. Organizational Behavior and Human Performance 6:36-49, 1971.

Wallis, W.A. and M. Friedman. The empirical derivation of indifference functions. In: Studies in Mathematical Economics and Econometrics editors: Lange, O., F. McIntyre, and T. Yntema; in memory of Henry Schultz ,University of Chicago Press, 1942.

Weldon, J.C. A note on measures of utility. Canadian Journal of Economics and Political Science 16:227-233, 1950.

Whellwright, P. A Critical Introduction to Ethics revised edition, New York, 1949.

Wherry, R.J. Orders for the presentation of pairs in the method of paired comparisons. Journal of Experimental Psychology 23:651-660, 1938.

## Appendix A

The subjects were required to assign nummerical estimates to expressions of probability. Figure 23 displays the results and where applicable compares them to the results obtained by Lichtenstein and Newman (1967).

| Our results N=23 colloquial phrase | mean | median | σ | range | | Results of Lichtenstein and Newman (1967) N=180 | | | |
|---|---|---|---|---|---|---|---|---|---|
| certain | 95 | 99 | 9.3 | 70 | -100 | | | | |
| extremely likely | 86 | 88 | 5.8 | 70 | -95 | | | | |
| highly probable | 79 | 80 | 9.7 | 57 | -99 | 89 | 90 | 0.0 | 60 -90 |
| probable | 65 | 65 | 11.8 | 50 | -90 | 71 | 75 | 0.1 | 1 -99 |
| uncertain | 37 | 42 | 14.9 | 5 | -55 | 40 | 50 | 0.1 | 8 -90 |
| unlikely | 23 | 22 | 12.9 | 5 | -60 | 18 | 16 | 0.1 | 1 -45 |
| impossible | 2 | 0 | 7.5 | 0 | -35 | | | | |

Figure 23 - Numerical Estimates of Probability Expressions

## Appendix B

This Appendix groups together questionnaire items which are relevant to a particular independent variable and within each group provides a description for each item. The questionnaire items are denoted by brief descriptive expressions.

### Acceptance of "Rationality" Axioms

#### (1) transitivity

George prefers driving a Ford Pinto to a Toyota MK II. Furthermore, he prefers driving a Toyota MK II to a Datsun 1600. Yet, from a rent-a-car which offers a Datsun 1600 or a Ford Pinto at the same rental rate, George rents a Datsun 1600 instead of a Ford Pinto. Realizing this "inconsistency", George should change his choice to Ford Pinto.

7=strongly agree-6-5-4-3-2-1=strongly disagree

#### (2) sure-thing

Upon entering his local confectionery store, Bill decides to spend a dime on either a bag of jelly beans or a grab bag which contains either jelly beans or chocolate. Although he prefers eating jelly beans to eating chocolate, Bill spends the dime on a grab bag rather than on a bag of jelly beans. In light

of his actual preferences, Bill ought to reverse his decision and spend the dime on a bag of jelly beans instead.

7=strongly agree-6-5-4-3-2-1=strongly disagree

(3) compound-gamble

A sweepstake ticket entitles the holder to either a prize of $1.oo or a chance in the grand final draw. The grand final draw prize will be either $100.00 or $1.00.

Another sweepstake ticket entitles the holder to a prize of $1.00 or $100.00.

Both sweepstake tickets sell for the same price. Taking the chances of winning into account, Dan calculates that the probabilities of winning each prize are the same for both sweepstakes. In spite of this information, Dan insists upon buying the second sweepstake ticket and is even willing to pay slightly more for this ticket. Dan should stop favoring the second sweepstake.

7=strongly agree-6-5-4-3-2-1=strongly disagree

## Propensities for Judgemental Modes of Evaluation

### (1) unemotional evaluation

In health matters, people ought to carefully evaluate their preferences among alternatives without being influenced by their mood or emotion at the moment of evaluation.

7=strongly agree-6-5-4-3-2-1=strongly disagree

### (2) prior-to-situation evaluation

In matter concerning illness, people ought to evaluate their preferences among alternatives before the illness actually occurs because under pain and discomfort they may not be clearly aware of their preferences.

7=strongly agree-6-5-4-3-2-1=strongly disagree

### (3) logical systematic evaluation

Suppose that in real life you had to make a decision in a choice situation involving several alternatives. How important do you feel that it is for you to analyze your preferences in a logical systematic manner (as was done in the experimental sessions) before making a decision?

7=extremely important-6-5-4-3-2-1=extremely unimportant

Realism of Scenario

How realistic do you feel that the ... scenario is?

7=extremely realistic-6-5-4-3-2-1=extremely unrealistic

Confidence in Method

(1) elicited in lieu of personal judgements

Suppose that with reference to a particular hypothetical scenario concerning a health matter, a trained health personnel derives your indifference (or utility) curve. If a situation resembling the scenario arises in real life, would you let a physician determine the decision for you from a careful consideration of your indifference (or cardinal utility) curve?

7=without any doubt-6-5-4-3-2-1=with complete doubt

(2) revision of "inconsistencies"

Suppose that you were to actually encounter a situation where you had to compare two alternatives each involving a domain of choices as presented in the ... scenario? Furthermore suppose that the decision you actually make does not conform with your elicited preferences. In light of this information, how important do you feel that it

is for you to change your decision?

7=extremely important-6-5-4-3-2-1=extremely unimportant

Discrimination For Probabability

Rate the significance of the following differences between probabilities on a 1 to 7 scale:

(1) occurrence with probability 0.5 as opposed to 0.55

(2) occurrence with probability 0.9 as opposed to 0.95

(3) occurrence with probability 0.1 as opposed to 0.2

(4) occurrence with probability 0.6 as opposed to 0.8

(5) occurrence with probability 0.3 as opposed to 0.5

(6) occurrence with probability 0.05 as opposed to 0.1

7=extremely significant-6-5-4-3-2-1=extremely insignificant

Discrimination for Day

Rate the significance of the following differences between number of days of rest in bed on a 1 to 7 scale:

(1) 1/2 day of rest in bed as opposed to 1/4 day

(2) 15 days of rest in bed as opposed to 14

1/2 days

(3) 4 days of rest in bed as opposed to 5 days

(4) 2 days of rest in bed as opposed to 2 1/4 days

(5) 10 days of rest in bed as opposed to 9 days

7=extremely significant-6-5-4-3-2-1=extremely insignificant

Bias in Interpretation of Probability

What probabilities do you associate with the following words (or phrases):

(1) certain

(2) unlikely

(3) highly probable

(4) uncertain

(5) probable

(6) impossible

(7) extremely likely