# Assessment of Attribute Importances and Consumer Utility Functions: von Neumann-Morgenstern Theory Applied to Consumer Behavior

JOHN R. HAUSER
GLEN L. URBAN*

Von Neumann-Morgenstern utility theory is a promising method to measure and model consumer preference. Its theoretical foundations provide explicit risk measures and testable behavioral conditions for alternative preference models. This paper summarizes selected results from the theory, addresses measurement and validity issues, and discusses the applicability of the theory to consumer research.

Modeling and measuring how consumers form preferences for products or services is critical to the understanding of consumer behavior. Considerable research has been applied to the task of determining how consumers combine perceptions of product attributes into preferences. For example, early work was directed at applying psychological concepts developed by Fishbein (1967). In many of these applications, a linear additive function of directly stated "importance weights" of product attributes and ratings of product attributes was used to predict a preference measure (see review by Wilkie and Pessemier 1973). Carroll (1972) used regression to fit a utility function to stated preferences by specifying the location of an "ideal point" based on the assumption of a utility function form. Work in conjoint analysis used monotonic analysis of variance to estimate "part worths" based on stated rank order preferences with respect to various prespecified product attributes (Tversky 1967; Green and Wind 1973; Green and Srinivasan 1978). In econometrics, stochastic modeling of observed choice with the logit form has been used to estimate the importance of attributes (McFadden 1970; 1975; 1978).

Another technique that shows promise for improvement in assessing importances is von Neumann-Morgenstern (vN-M) utility theory (von Neumann and Morgenstern 1947). The purpose of this paper is to assess the applicability of vN-M utility theory to modeling consumer choice behavior. We review some of the key concepts of utility theory that are important in consumer behavior, and emphasize the differences between vN-M utility theory and the other methods of modeling consumer preferences. The appropriateness of vN-M utility theory to consumer choice is discussed and an empirical application of modeling a consumer utility function for new health services is presented.

* John R. Hauser is Assistant Professor of Marketing and of Transportation, Graduate School of Management, Northwestern University, Evanston, IL 60204. Glen L. Urban is Professor of Management Science, Alfred E. Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139.

## KEY CONCEPTS OF VON NEUMANN-MORGENSTERN UTILITY THEORY

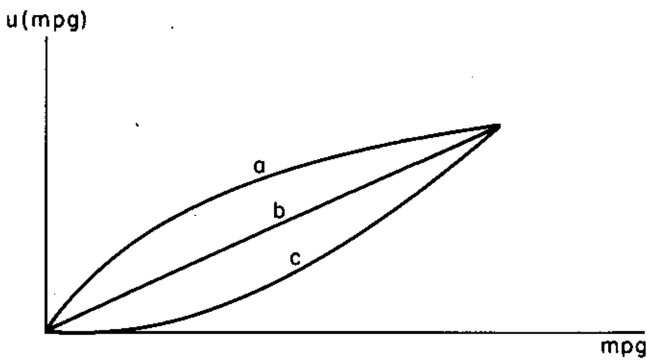
Since first proposed in 1947, von Neumann-Morgenstern utility theory has received much attention as a method of decision analysis. Keeney and Raiffa (1976) have written a comprehensive textbook on the subject and Farquhar (1977 a) has summarized recent research and applications. We will discuss a selected set of key concepts of vN-M utility theory relating to risk, measurement, functional form, and parameter estimation with emphasis on the issues in vN-M utility theory relevant to modeling consumers' choice.

### Risk

Consumer behavior under risky situations is an important problem, especially for major purchases or decisions such as those regarding consumer durables, major industrial products, or health services. One of the most unique and useful features of vN-M utility

**FIGURE A**

MODELING RISK




theory is its explicit modeling of risk. Unlike expectancy value, preference regression, conjoint analysis, or logit analysis, which often include risk as a variable, vN-M utility theory explicitly includes risk in its axiomatic foundations (von Neumann and Morgenstern 1947).

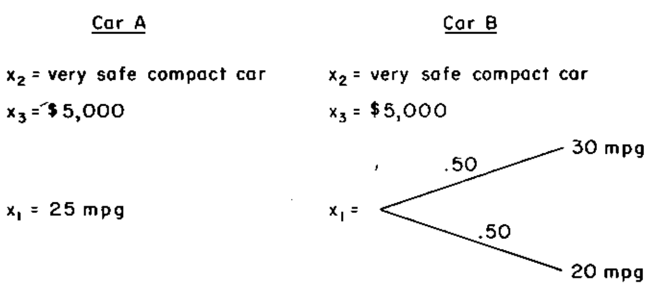
Risk is modeled by transforming the independent variable, e.g., miles per gallon (mpg) of a used car, by a function that reflects the decision-maker's response to uncertain outcomes. Figure A shows several functions ($a$, $b$, or $c$) that could be used to transform the independent variable, in this case mpg.

Von Neumann-Morgenstern utility theory bases the selection of the transformation function on the decision-maker's response to a choice between a risky situation and a riskless situation. A lottery is presented to the decision-maker. Figure B shows an example of a lottery for assessing a consumer's risk transformation for the attribute of gas mileage for a used car. (This is a diagram; the actual consumer question must be carefully worded.)

The consumer is told that Car A has a guaranteed mileage of 25 mpg. S/he is told that the mileage of Car B is uncertain and that it is equally likely to be 30 or 20 mpg. (For example, imagine that you buy a used car and cannot determine the true mileage until you drive it for over 2,000 miles, but once you buy it, you are stuck with the purchase. The expected values

**FIGURE B**

SCHEMATIC OF A LOTTERY AS A REPRESENTATION OF CONSUMER'S PERCEPTION OF mpg




for both cars are the same, but Car B's gas mileage is uncertain.) If s/he prefers Car A, then the consumer is called risk averse due to a reluctance to take the risk. If s/he prefers Car B, s/he is risk prone because s/he would rather take a chance even though the average outcome is no better than the certain outcome. Finally, if s/he has no preference, s/he is risk neutral. We have specified the safety and price of Car A and Car B because, in general, the consumer's risk aversion or risk proneness may depend on these characteristics.

In vN-M utility theory, $u$(mpg) is scaled to represent this behavior. For example, if our consumer were indifferent between Car A and Car B, the utility of 25 mpg would equal 0.50 times the utility of 30 mpg plus 0.50 times the utility of 20 mpg. In this case, the mpg of the car (or any linear rescaling of mpg) could serve as utility and the linear model, Curve b, would apply. If our consumer preferred Car A, we could ask further lottery questions until we found a value for the mpg of Car A so that our consumer was indifferent between Car A and Car B. If this occurs when the mileage of Car A equals 23.5 mpg, we must scale $u$(mpg) so that the utility of 23.5 equals 0.50 times the utility of 30 mpg plus 0.50 times the utility of 20 mpg. One such scaling is $u(20) = 0.0, u(23.5) = 0.50$, and $u(30) = 1.00$.

Such a utility function is termed "risk averse" because the consumer prefers a safe outcome to an uncertain outcome with the same expected value. Raiffa (1972) shows that such risk averse utility curves must be concave, as shown by curve a in Figure A. Note that the difference in expected values (25 mpg for Car B minus 23.5 mpg for Car A) can be thought of as a type of risk premium (Keeney and Raiffa 1976, p. 151). The opposite is true for those who are "risk prone." They will be indifferent at an expected lottery outcome less than the certain outcome. A risk prone function would be convex, as shown by Curve c in Fig. A.

The utility function in Figure A could be drawn based on lottery responses for all values of mpg, but if a special form of risk response called "constant risk aversion" can be justified, the process can be simplified. Raiffa (1972) shows that the functional form in this case is:

$$u(x) = a + b\, exp(-rx), \qquad (1)$$

where $a, b, r$ are parameters to be estimated, $x$ represents mpg, and $r$ is called the coefficient of risk aversion. If $r \to 0$, the function becomes linear; if $r > 0$, risk aversion is present.

Constant risk aversion means that the consumer is concerned only with the spread of outcomes, not their actual levels. For example, if we add a catalytic converter to both Cars A and B (reducing the mpg of both cars by 5 mpg), then constant risk aversion implies that the consumer's choice between Car A and Car B remains unchanged. Constant risk aversion can be tested as shown in Figure C.

If a person is constantly risk averse and indifferent in Question 1 Figure C, he must also be indifferent in Questions 2 and 3, or in any question where a fixed amount is added to all the outcomes.

When constant risk aversion is true, measurement and estimation are simplified. When it is not true, the transformation must be represented by another function or a series of points. For example, suppose the consumer becomes less risk averse if larger and larger fixed amounts are added to all outcomes. Such a decreasingly risk averse consumer might prefer Car C in Question 1, be indifferent in Question 2, and prefer Car H in Question 3. One utility function that could represent this behavior is $u(x) = a + b \log (x + c)$, where $a$, $b$, and $c$ are parameters to be estimated.[1]

The result of the risk modeling is a transformation of the independent variable. If the respondent is risk neutral, a linear function is appropriate. This would make utility theory equivalent to the existing linear approaches of expectancy value and the linear case of preference regression and conjoint analysis. If there is risk response, a nonlinear transformation will result. Preference regression and conjoint and logit analyses can use nonlinear forms, but utility theory has advantages. Its functional form is based on responses to risky situations and, in cases such as constant risk aversion, it provides functional forms based on testable conditions.

## Measurement

The previous section has demonstrated a measurement innovation associated with vN-M utility theory. Measuring risk aversion to particular attributes by lotteries is a feature of vN-M utility theory. The theory also provides a different method for making comparisons between attributes based on indifference questions and pairs of certain levels of attributes. Although the concept of indifference and indifference curves is frequently used in economics, the uniqueness of vN-M utility functions to a positive linear scaling (von Neumann and Morgenstern 1947, p. 25) makes indifference a practical measurement technique. Parameter derivation from indifference questions is illustrated in the Appendix.

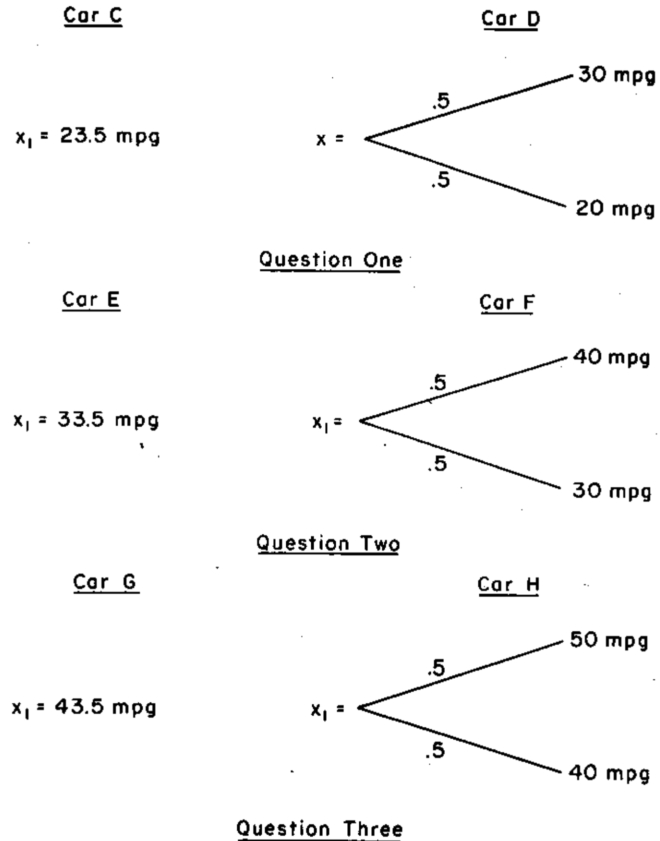
In Figure D, Car J is described by its attributes of price, safety, and mpg, and Car K, by price and safety only. The consumer is asked to specify the mpg of Car K such that s/he would neither prefer Car J nor Car K. For example, suppose the consumer set the mpg of Car K at 30 mpg. Then s/he is saying that a mpg of 31 would cause him or her to prefer Car K, while a mpg of 29 would cause him or her to prefer Car J. We have specified the safety of Car J and Car K because, in general, the consumer's response to the indifference question can depend upon this fixed and common value of safety.

This approach to utility assessment is different from the direct statement used in expectancy value models, the rank order procedure of conjoint analysis, or the preference scaling procedure of preference regression methods.

[1] It is interesting to note that the quadratic utility function, $u(x) = a + bx + cx^2$, popular in consumer behavior models such as PREMAP, is increasingly risk averse. Utility theory also provides an explicit measure of local risk aversion, $r(x) = -u''(x)/u'(x)$ where $u'$ and $u''$ represent the first and second derivatives of $u(x)$. Note that $r(x) = r$ in Equation 1. The properties of $r(x)$ are discussed in detail in Keeney and Raiffa (1976, p. 160).

**FIGURE C**

CONSTANT RISK AVERSION TESTING




## Functional Form

The uniqueness of vN-M utility functions makes indifference measurement theoretically feasible, but to be practical we must identify functional forms that are appropriate. With known functional forms, we can specify the vN-M utility function by relatively few pa-

**FIGURE D**

SCHEMATIC OF INDIFFERENCE MEASUREMENT

| Car J | Car K |
|---|---|
| $x_2$ = very safe compact car | $x_2$ = very safe compact car |
| $x_3$ = \$5,000 | $x_3$ = \$3,000 |
| $x_1$ = 40 mpg | $x_1$ = ______ mpg |

rameters rather than trying to measure every point on the utility surface.

One of the strengths of vN-M theory is that there exists a rich literature identifying which functional forms are appropriate under reasonable behavioral assumptions (Keeney and Raiffa 1976). We will illustrate a few of the functional forms most appropriate for modeling consumer behavior.

In Figure B we specified the safety and price of both Cars A and B, even though they were the same for both cars. Suppose that the consumer's response to the lottery question on mpg did not depend on these fixed and constant levels of safety and price. We would then say that mpg was "utility independent" of safety and price. Keeney (1972) shows that if each product attribute is utility independent of all other attributes, the utility function must be of the following form, called the "quasi-additive" form:

$$U(x_1,x_2, \ldots ,x_K) = \sum_k w_k u_k(x_k) + \sum_k \sum_{m>k} w_{k,m} u_k(x_k) u_m(x_m) + \ldots w_{1,2,3,\ldots K} u_1(x_1) u_2(x_2) \ldots u_K(x_K), \quad (2)$$

where $x_k$ is a measure of the $k^{\text{th}}$ attribute, $u_k(x_k)$ is the utility transformation for attribute $k$, $w_k$ is the importance weight for the $k^{\text{th}}$ attribute, and $w_{k,m}$, $w_{k,m,n}$, . . . $w_{1,2,3,\ldots K}$ are the importance weights for interactions of attributes.

Utility independence can be tested by repeating the lottery questions where the levels of price and safety of Cars A and B are on a common but higher (or lower) level.

Equation 2 simplifies measurement, as risk must only be assessed for one attribute at a time. But Equation 2 still requires $K$ uni-attributed functions and $2^K - 1$ parameters for $K$ attributes. Equation 2 can be simplified with an assumption known as "pairwise preferential independence."

In general, the consumer's tradeoffs (see Figure D) between mpg and price may depend on how safe the car is. But if these tradeoffs do not depend on safety, then we say that the attribute pair mpg-price is preferentially independent of safety. The assumption can be tested by repeated measures where the level of safety is changed to a new level in Figure D. Consumers' preferences may still depend on safety, but the relative tradeoffs between price and mpg must be independent of safety. Keeney (1974) shows that if each pair of attributes is preferentially independent of its complement set of all other attributes and if utility independence applies, then the quasi-additive form can be simplified by:

$$w_{k,m} = W w_k w_m,\ w_{k,m,n} = W^2 w_k w_m w_n,\ \ldots\ w_{1,2,\ldots K} = W^{K-1} w_1 w_2 \ldots w_K, \quad (3)$$

where $W$ is a measurable constant.

This special form is known as the multiplicative form because it can be readily factored. If there are $K$ attributes, it requires only $K + 1$ constants and, as a result, is much easier to measure. There are alternative functional forms if independence conditions can be justified for some rather than all the attributes (see Farquhar 1975).

To illustrate the importance of these simplifications, suppose we are measuring a utility function for $K = 10$ attributes and suppose that each attribute is represented by $L = 3$ levels. In general, the utility function has $L^K - 1 = 3^{10} - 1 = 59{,}048$ points. The quasi-additive form required $K = 10$ uni-attributed functions of $L - 1 = 2$ points each plus $2^K - 1$ parameters. This gives $K \cdot (L - 1) + 2^K - 1 = 1{,}043$ parameters. If the quasi-additive form is simplified to the multiplicative form, then only $K \cdot (L - 1) + K + 1 = 31$ parameters are required. Finally, if we can use the constantly risk averse form (Equation 1), this can be reduced to 21 parameters ($r_k$ and $w_k$ for $k = 1, 2, \ldots, 10$ plus $W$).

The multiplicative constant, $W$, can be given behavioral interpretation. Keeney (1974) shows that if $W > 0$, the attributes act as complements; if $W = 0$, they are independent, and if $W < 0$, they are substitutes. Richard (1975) later gave a multiattributed risk aversion interpretation. If $W < 0$, the consumer is risk averse with respect to the combined set of attributes (total purchase). Similarly, $W = 0$ implies risk neutrality and $W > 0$, risk proneness. Note that if $W = 0$, then all higher order terms in Equation 3 drop out and the functional form becomes additive.[2]

Utility theory models reduce to the commonly used linear additive model when consumers are risk neutral, utility and preferential independent, and the interactive coefficient ($W$ in Equation 3) is zero, all of which are testable conditions. It is evident that vN-M utility theory provides a richness in functional form. While other statistical procedures allow nonlinear and interaction effects, the vN-M utility theory functional form is based on specific response assumptions that can be measured and empirically tested.

## Estimation

Utility theory does not use a statistical estimation procedure such as those applied in conjoint, preference regression, or logit analyses. The parameters are uniquely calculated directly from the lottery and tradeoff questions. Only enough questions are asked to algebraically solve for each parameter.[3] The calculations are made efficient by scaling the utility functions, $u(x)$, so that they have a value of 0 at the minimum $x$ and 1.0 at the maximum $x$. The total utility $U(x_1,x_2 \ldots x_K)$ is similarly scaled so it is 0 when all $x_k$ are at their minimum and 1.0 when all $x_k$ are at their maximum.

[2] Fishburn (1971) gives necessary and sufficient testable conditions for the additive form.

[3] See Appendix for a detailed example of the calculation procedure.

With these scaling conventions, the number of lottery and tradeoff questions are reduced. If constant risk aversion is justified, only one lottery question is needed per attribute. If utility independence is to be checked, a second lottery is needed for each attribute. The number of tradeoffs required is the number of attributes less one. These can be repeated to test preferential independence. In addition, one combination lottery and tradeoff is required (see Appendix, Figure G).

The number of lotteries and tradeoffs can become large, but in many of the derived functional forms measurement is feasible. In this measurement, questions are repeated to check for consistency (Keeney and Raiffa 1976), but measurement error is not explicitly modeled. Although likely to be overcome in the future, this lack of an error theory prevents statistical statements about parameter estimates and prevents multiple observations to achieve minimum variance or maximum likelihood estimates. Furthermore, heuristic guidelines must be used to test independence and risk aversion properties as measurement error can confound any tests. In this aspect, vN-M utility theory is conceptually different from conjoint, logit, and preference regression approaches, which make specific distributional assumptions about errors and have statistical tests available. In modeling consumer behavior with vN-M utility theory, consistency must be assumed or checked by repeated assessment.

## APPROPRIATENESS IN MODELING CONSUMER BEHAVIOR

Von Neumann-Morgenstern utility theory has not achieved widespread use in modeling consumer behavior. To use vN-M utility theory for this purpose, a consumer researcher must address the issues of predictive validity, measurement feasibility, empirical assumption testing, descriptive validity, selection of the attributes, and measurement error.

Usually, vN-M utility theory is used to guide a single decision-maker or a small group of decision-makers. In these applications, vN-M utility theory helps the decision-maker "rationally" evaluate alternative decisions (such as the site of an airport, in Keeney and Raiffa 1972, Chapter 8) and quantitatively incorporate uncertainty about the outcome of any decision. There is no need for descriptive validity as the research goal is to modify the decision. Although utility is usually used prescriptively, it may be used descriptively (Krantz, Luce, Suppes, and Tversky 1971, p. 49). In fact, recent work in marketing (Hauser 1978) and psychometrics (Kahneman and Tversky 1978) indicates improved axiom systems for descriptive use of utility or utility-like functions.

In consumer behavior, the goal is to describe and predict how a consumer will behave when presented with alternative choices, such as the purchase of a new product. Alternatively, we may wish to use persuasive strategies to influence choice by modifying the consumer's utility function. Thus, in consumer behavior an application of utility theory must be tested for predictive validity. For example, one might estimate a consumer's utility function and then present the consumer with a choice of products. The measured utility function should predict his/her choice better than a random model and at least as well as alternative models, such as preference regression or logit analysis.

When, for example, a manager's career rests on the outcome of a major decision, such as what fleet of police vehicles to purchase, s/he will make available the necessary time (at least four to eight hours, over several weeks) to have his/her utility function assessed. This is not possible in consumer behavior. Often for representativeness, a large number of consumers must be measured. A four-hour interview for each of 100 consumers would be infeasible. Furthermore, most consumers will be unwilling to participate in an interview of much more than one hour in length. Thus, in a short 45–60 minute interview the consumer must be motivated and educated to the lottery and tradeoff questions necessary for assessment and must respond to the assessment and verification questions. Furthermore, the tasks cannot be too onerous or too complex, but must involve the consumer so that s/he gives thoughtful answers that reflect reality. Often this means that the consumer researcher must make careful tradeoffs among the generality (and hence appropriateness) of the utility function, the number of attributes, the degree of assumption testing, and redundancy for estimation. Depending on the goals of a particular study, the researcher may include more attributes and a simpler functional form or less attributes and a more complex functional form. Similarly, a researcher may concentrate on assumption testing or parameter estimation.

In decision analysis, one can fully test all assumptions. If an assumption is violated (say utility independence), the analyst can redefine the attributes (Farquhar 1977a) or switch to a utility function with a less restrictive assumption. This continues until an appropriate function is specified and assessed. The decision-maker then adopts this function, which, by definition, is the correct function. In consumer behavior the orientation is descriptive validity rather than definitional validity. This means that the responses of questions must be used to infer the appropriateness of a representation. As the vN-M utility theory functions are derived from assumptions of preferential and utility independence, which can be tested, we can indirectly infer descriptive adequacy if the assumptions are met.

The final consideration in the use of vN-M utility theory in consumer behavior is the appropriateness and the measurement of the attributes used in the utility function. In prescriptive utility theory, the preference measures are usually quantifiable (e.g., tons of hydrocarbon released into the atmosphere) although when necessary "softer" measures are used, such as "political flexibility" (Keeney and Raiffa 1976, Chapter 7).

In consumer choice, the attributes must represent the evaluative dimensions consumers actually use. Furthermore, the researcher must be careful not to mix physical cues (e.g., travel time of a bus) with psychological perceptions based on those physical cues (e.g., general performance of a bus) in the same utility function, as perceptions are intervening variables for the physical cues (Brunswik 1952). This often means that perceptual measures, such as the "quality" of a health care plan, must be psychologically scaled, and that the indifference questions be asked relative to the scaled variables. This introduces measurement error in the attributes and hence the utility function. Thus, the consumer researcher must make a decision whether to assess the utility function over physical cues or psychological perceptions. The former have less measurement error, but confound perception and preference, while the latter introduce measurement error.

In summary, vN-M utility theory has a number of advantages that make it particularly attractive for consumer research. Among these are: (1) explicit incorporation of risk preference, (2) the indifference task for consumer measurement, (3) identification of appropriate functional forms, and (4) axiomatic derivation from testable behavioral assumptions (preference and utility independence).[4] But before vN-M utility theory can be empirically applied to consumer behavior, a number of important issues must be addressed. Among those are: (1) the descriptive and predictive validity in consumer choice, (2) the feasibility of the consumer task used in measurement, (3) the use of psychologically scaled perceptual measures as attributes, and (4) how to make the research tradeoffs among complexity, assumption testing, and parameter estimation that are necessary to limit the measurement task.

The following is an empirical example that highlights the issues and suggests a set of practical solutions. We hope this example will facilitate discussion and encourage researchers to develop more and better techniques to address these issues in applying vN-M utility theory. The reader should recognize that this was a marketing research application to help an organization design their health care service, and a number of research tradeoffs were made. Application with differing goals, such as detailed assumption testing, would concentrate on different questions in the design of the consumer task.

## EMPIRICAL EXAMPLE: CONSUMER PREFERENCE FOR HEALTH CARE DELIVERY SYSTEMS

Health Maintenance Organizations (HMOs) have been proposed as a method of reducing costs and increasing availability and quality of health services. Although some HMOs have been successful, a major problem is how to gain sufficient enrollment. An HMO being developed by the Massachusetts Institute of Technology (MIT) provided the setting for this application. The MIT health service could better design their service and hence attract sufficient enrollment if they understood the consumer process. To address this problem, we developed a normative consumer response model (Hauser and Urban 1977) that included models of consumer awareness, perceptions, preferences, trial, and repeat. The consumer model led to managerial actions that were able to increase enrollment by roughly 50 percent.[5]

### Measurement

One hundred students were randomly selected from MIT registration rolls for utility assessment. Each student was requested (by telephone) to participate in a study that would require a 45-minute personal interview and completion of a written survey. The net response rate for both tasks was 79 percent.

Four perceptual attributes of health services were identified: (1) quality of care, (2) personalness, (3) convenience, and (4) value. These were deduced by factor analysis in a previous study of 450 MIT faculty, students, and staff.[6] Quality correlated with trust, preventive care, and availability of good doctors and hospitals. Personalness reflected a friendly atmosphere with privacy and no bureaucratic hassle. Value was not just price, but paying the right amount for the services. Convenience reflected location, waiting time, and hours of operation. In the utility assessment, these four attributes were rated on a seven-point scale (extremely poor, very poor, poor, satisfactory, good, very good, excellent). Linear regression was used to relate the seven-point scales to the factor scores.

A pretest of the lottery questions indicated that a constantly risk averse function was a reasonable approximation for most students. Thus, given a 45-minute constraint, we concentrated on a parameter estimation with testing of utility and preferential independence of the four attributes.

The interview included a series of warm-up questions to help respondents understand lottery questions. The lottery questions were asked to assess risk aversion and check for utility independence. Six tradeoff questions were asked to assess attribute importances and check for preferential independence. These responses plus one combined lottery and tradeoff question (see Appendix Figure G) supplied enough data to calculate each respondent's multiplicative utility function over the four health service attributes.

[4] The existence of a utility function is also based on testable axioms, e.g., see von Neumann and Morgenstern (1947), Herstein and Milnor (1953), Friedman and Savage (1952), Jensen (1967), or Marshak (1950).

[5] The managerial setting and the complete model are available in Hauser and Urban (1977).

[6] See Hauser and Urban (1977) for a detailed discussion of this study.

**TABLE 1**

COMPARISON OF AVERAGE IMPORTANCE ESTIMATES AND THE ABILITY OF THE MODEL TO PREDICT FIRST PREFERENCE

| Item | Normalized importance weights | | | | Proportion of correct first preference prediction |
|---|---|---|---|---|---|
| | Quality $w_1$ | Personalness $w_2$ | Value $w_3$ | Convenience $w_4$ | |
| Utility assessment | | | | | |
| Raw importance weights | .30 | .19 | .26 | .25 | .50 |
| Marginal weights | .31 | .25 | .25 | .19 | |
| Preference regression | | | | | |
| Least squares | .32 | .09 | .38 | .21 | .47 |
| Monotonic | .34 | .08 | .31 | .27 | .45 |
| Logit analysis | .34 | .16 | .34 | .16 | .43 |

Each respondent also filled out a written form so that alternative procedures could be implemented to provide empirical comparisons to vN-M utility theory. In this survey, students rated their existing care and three concept statements (MIT HMO, Harvard Community Health Plan, and Massachusetts Health Foundation) on 16 Likert-scaled (five-point agree/disagree) basic attributes. The attributes were selected as representative of the full range of consumer perceptions through qualitative research (group discussions) and used in the factor analysis in the previous large scale survey.

Rank order preference for the three concepts and existing care were obtained so that preference regressions and logit analysis could be carried out.

## Descriptive Validity

There are three indicators of descriptive validity: (1) assumption testing, (2) predictive accuracy, and (3) consistency with alternative models.

*Assumption testing.* We defined the criteria for the satisfaction of the assumption of utility or preferential independence to be less than a 10 percent difference in the measured response from the first and second lotteries or tradeoffs. By these criteria, utility independence held in 69 percent of the cases tested and preferential independence held in 88 percent of the cases tested. These numbers indicate that the assumptions are justified for a substantial majority of consumers. These numbers also imply that more complex functions may be useful for at least some of the consumers.

*Utility parameter estimates and predictive accuracy.* Table 1 shows the averaged raw and marginal weights for quality, personalness, convenience, and value in normalized form. The marginal weights are the first differential of the utility function at the point of the respondent's chosen health care system. Although our sample was reasonably homogenous (all MIT students), there was considerable individual variation. The interquartile ranges for importance weights were: quality +12.5 percent to −18 percent of the median, personalness +45 percent to −31 percent, value +14 percent to −29 percent, and convenience +17 percent to −23 percent.

In addition to "importance" ($w_k$), utility theory also provides explicit measures of risk aversion ($r_k$) and interaction ($W$). In our sample, consumers were definitely risk averse ($\bar{r}_1 = 0.693$ for quality, $\bar{r}_2 = 0.332$ for personalness, $\bar{r}_3 = 0.424$ for value, and $\bar{r}_4 = 0.310$ for convenience, where $\bar{r}_k$ indicates median). Variation was similar to that for importances, but the full interquartile range was risk averse for quality and value, and risk averse or risk neutral for personalness and convenience. Behaviorally, this questions the appropriateness of a linear model and suggests that consumers are reluctant to switch to new alternatives (HMOs) if they are uncertain of the plan's ability to deliver the promised quality, personalness, value, or convenience.

The full interquartile interval for the interaction coefficient ($W$) was between −0.99 and −0.93, indicating strong substitutability between attributes and multiattributed risk aversion for most consumers. Behaviorally, this questions the appropriateness of an additive model and suggests a strong interaction among the attributes.

Based on the utility parameters, predictions were made of the first choice among the four concepts given to students in the written questionnaire. The ratings of quality, personalness, convenience, and value for each concept were used as independent variables and the concept with the highest utility was designated the first choice. These predictions were true for 50 percent of the respondents. This can be compared to a random forecast of 25 percent (equal probability over the four alternatives).

*Comparison to alternative models.* Von Neumann-Morgenstern utility assessment does well compared to random assignment. We also compared the vN-M predictions to some representative models of consumer preference. Table 1 reports the results of these comparisons. Both preference regression and logit analysis use linear preference functions:

**TABLE 2**

COMPARISON OF PREDICTED AND ACTUAL MARKET SHARES FOR PREFERENCE PREDICTION

| Item | Existing care | Harvard Community Plan | MIT HMO | MA Health Foundation | $\chi^2$ ($df = 3$) |
|---|---|---|---|---|---|
| Actual | .34 | .11 | .42 | .13 | — |
| Utility | .30 | .08 | .42 | .20 | 4.1 |
| Preference regression | | | | | |
| Least squares | .19 | .19 | .45 | .18 | 13.4 |
| Monotonic | .20 | .24 | .41 | .15 | 13.2 |
| Logit | .22 | .23 | .35 | .20 | 14.0 |

$$u(x_1,x_2,x_3,x_4) = \sum_k w_k x_k, \qquad (4)$$

where $w_k$ are the importance weights. These should be compared to the marginal weights reported for vN-M utility assessment.

The models differ in the way that the importance weights are estimated. In preference regression the $w_k$ are estimated by both least squares and monotonic regression. The dependent measure is stated preference ($P_{ij}$) of health alternative $j$ for individual $i$. The explanatory variables were the factor scores ($x_{ijk}'$) of the perceptual attributes of quality, personalness, convenience, and value. Regressions were run across consumers and stimuli to determine population importance weights. Logit analysis uses a linear preference function, but uses an alternative estimation procedure based on maximum likelihood conditions. The dependent measure is first preference and the explanatory variables are the factor scores.[7]

All the models indicate quality as the most important attribute. But vN-M utility theory estimates personalness at a higher level of importance than the other methods.

As we are concerned with preference orderings, an appropriate measure of goodness of fit is the proportion of first preferences correctly predicted. The standard error of this measure is approximately 5.6 percent. Von Neumann-Morgenstern utility assessment is significantly superior to an equally weighted model (0.05 level), which correctly predicted 40 percent of the observations, but is only marginally superior to logit analysis (0.15 level) and preference regression (0.25 level). But preference recovery is only one way to compare the alternative models.

The methods can also be compared by examining their ability to recover the share of choices of each of the four health plans (Table 2). The $\chi^2$ statistic is based on the frequency of actual and predicted choices for each alternative (three degrees of freedom). Von Neumann-Morgenstern utility theory again fits best. In this case, vN-M utility theory was not significantly different from actual at the 0.200 level, while all the other models were significantly different from actual at the 0.005 level.

In comparing the models, it should be noted that the preference regression and logit analyses used the stated preferences in fitting the importance weights, while the vN-M utility theory analysis used separate measures from lotteries and tradeoffs. The vN-M utility theory result is actually a predictive test, while the regression and logit results are a measure of goodness of fit. On the other hand, vN-M utility theory uses more parameters and individual specific preference functions.

One reason for the superior performance of vN-M utility theory is that the linear models overpredict switching from existing care to the new HMOs (Table 2). One possible explanation for this hesitancy to switch might be risk aversion. It is better incorporated in the utility model than in the linear model. Another explanation might be the higher marginal importance weight for personalness that resulted from including nonlinearities and interactions in the utility function. In either case, prediction is improved by including the nonlinearities, which appear to have an intuitive explanation in terms of risk aversion.

## Measurement Feasibility

In this sample, respondents related well to the tasks. They found them interesting and felt that their answers reflected how they felt about health care. Students had no trouble understanding the task when the proper warm-up and educational questions were asked. The empirical results indicate the measures were of good quality and allowed adequate utility functions to be parameterized.

## Tradeoffs

In a 45-minute interview, it was possible to parameterize a fairly complex utility function to provide information on relative importances of attributes, risk aversion with respect to those attributes, and a measure of attribute interaction. The assumptions were supported for most respondents and the model fit the data adequately. Although more assumption testing would be desirable, these tradeoffs were made to conclude the interviews in 45 minutes and were acceptable for a four-variable, multiplicative utility function. Based on our experience, either more complex functions can be measured with less assumption testing or more assumptions can be tested for fewer attributes.

## Psychologically Scaled Attributes

Careful attention to detail and iterative development of psychological scales resulted in utility functions that adequately represent consumer preferences. Our experience indicates that the use of psychological at-

[7] For details on logit analysis, see McFadden (1970; 1975; 1978).

tributes does not represent a barrier to the use of vN-M utility theory in modeling consumer choice.

## CONCLUSION

This initial application of vN-M utility theory indicates that consumer measurement is feasible, psychological attributes can be included, and empirical results were equal to or better than two competing approaches. These are encouraging results because this implies vN-M utility theory's attractive features of modeling risk, indifference measurement, and identification of practical form have potential for application in consumer research.

Von Neumann-Morgenstern utility theory can be a valuable tool for understanding and predicting consumer behavior. It can be most effective if: (1) risk aversion and interaction phenomena are deemed to be important in the consumer's behavior, (2) a sufficient budget is available for the personal interviews, (3) individual utility parameters are important to the research or managerial question, and (4) consumers are well educated. It is particularly effective if the number of decision-makers is small and the choice decision large. For example, purchase of large computers, aircraft, automated machine tools, or other industrial products might be good applications. Other useful examples might be consumer durables, such as washer/dryers or automobiles. In services it might be applicable to health care, college selection, and career selection.

Several future areas of research are appropriate. Our application acknowledges measurement error, but does not explicitly include it in parameter estimation. Research is needed to allow degrees of freedom and to develop distributional assumptions for parameter estimation. The utility and conjoint axioms appear compatible, but research is needed to develop a common set of consistent axioms. Needed also are statistical tests of assumptions so that confidence limits can be set for the repeated lottery and tradeoff questions used in testing utility and preferential independence.

Another topic is the development of more efficient measurement techniques, thus allowing some combination of more complexity, more attributes, or more assumption testing. A final need is to develop simpler measurement methods. Our sample was MIT students. Von Neumann-Morgenstern utility theory requires further testing to determine whether an average respondent could accurately answer lottery questions, even with careful training.

[*Received June 1977. Revised September 1978.*]

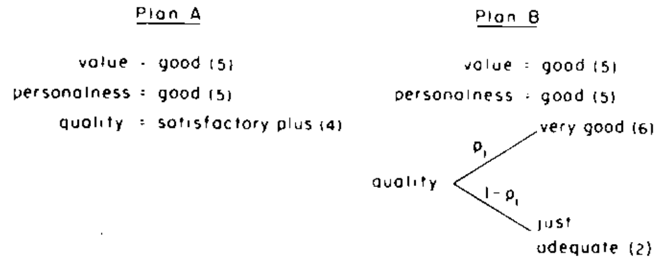
## APPENDIX

### An Example of Utility Assessment

A simplified version of vN-M utility assessment relative to the quality, personalness, and value of health care is presented, using simplified measures of the attributes, extreme values in the lottery and tradeoff questions, and three rather than four attributes. The extensions to empirical measurement and to other applications are straightforward. All calculations are done separately for each consumer in the sample.

**FIGURE E**

SCHEMATIC OF LOTTERY FOR QUALITY




The utility theory axioms imply that if the utility function is appropriate and the consumer is indifferent between two stimuli, then the utilities of each stimuli are equal. Since the functions are unique to positive linear transformations, the choice of anchors for the scale are arbitrary and thus, for simplicity, we define the range of the multiattributed utility function to vary between 0 and 1. Let $x_1$ = the level of quality, $x_2$ = level of personalness, and $x_3$ = the level of value.

Assume that we have verified the assumptions of utility independence and constant risk aversion for each attribute and preferential independence for each pair of attributes. In practice, these would be checked by pretests and/or repeated measures. Then the multiplicative function is given by:

$$\begin{aligned} U(x_1x_2x_3) &= w_1u_1(x_1) + w_2u_2(x_2) \\ &\quad + w_3u_3(x_3) + Ww_1w_2u_1(x_1)u_2(x_2) \\ &\quad + Ww_1w_3u_1(x_1)u_3(x_3) + Ww_2w_3u_2(x_2)u_3(x_3) \\ &\quad + W^2w_1w_2w_3u_1(x_1)u_2(x_2)u_3(x_3), \quad (5) \end{aligned}$$

where $u_k(x_k) = a_k + b_k exp(-r_kx_k)$.

The first task is to measure the uniattributed scale functions, $u_k(x_k)$. We will illustrate the measurement for quality. As quality is presumed to be utility independent of personalness and value, we can fix personalness and value, and allow only quality to vary over the lottery (schematically represented in Figure E). In actual practice, 3- × 5-inch cards representing the health care plans and a lottery wheel are used to indicate probabilities. Theoretically, we can fix $p_1$ (at say $p_1$ = 0.50) and have the consumer specify the quality of Plan A, but in practice consumers relate better to setting the probability wheel than setting the quality of Plan A.

Suppose we assign a numerical scale to the attributes as indicated by the numbers in parentheses in Figure E. Since $U$(Plan A) = $U$(Plan B), we get:

$$U(x_1 = 4, x_2 = 5, x_3 = 5) = p_1 U(x_1 = 6, x_2 = 5, x_3 = 5) + (1 - p_1) U(x_1 = 2, x_2 = 5, x_3 = 5). \quad (6)$$

Now, we arbitrarily scale $u(x)$ as:

$$u_1(x_1 = 6) = 1, \quad u_2(x_2 = 6) = 1, \quad u_3(x_3 = 6) = 1$$

and

$$u_1(x_1 = 2) = 0, \quad u_2(x_2 = 2) = 0, \quad u_3(x_3 = 2) = 0$$

and

$$U(x_1 = 6, x_2 = 6, x_3 = 6) = 1,$$
$$U(x_1 = 2, x_2 = 2, x_3 = 2) = 0.$$

Then, substituting Equation 5 in Equation 6, and cancelling terms yields:

$$u_1(x_1 = 4) = p_1 u_1(x_1 = 6) + (1 - p_1) u_1(x_1 = 2). \quad (7)$$

The scaling also gives $u_1(x_1) = (1 - exp[-r_1(x_1 - 2)])/(1 - exp[-r_1(6 - 2)])$. Substituting in Equation 7, and solving for $r_1$ gives:

$$r_1 = (1/2)\, ln\, [p_1/(1 - p_1)], \quad (8)$$

except in the special case of $p_1 = 0.50$, where $u_1(x_1)$ is indeterminate (zero divided by zero), in which case $u_1(x_1)$ becomes linear and $u_1(x_1) = (x_1 - 2)/4$ (L'Hospital's rule). For example, if $p_1 = 0.80$, then $r_1 = 0.69$. Similar questions and computations would apply for personalness and value, i.e., $u_2(x_2)$ and $u_3(x_3)$.

The next task determines the relative tradeoffs among the attributes, i.e., the ratio of $w_2$ to $w_1$ and the ratio of $w_3$ to $w_1$. We will illustrate the measurement for $w_2/w_1$. As quality and personalness are assumed in the example to be preferentially independent of value, we can fix value and ask a tradeoff question between quality and personalness (Figure F).

Suppose the consumer assigns a value of $q$ to Plan D. As $U$(Plan C) = $U$(Plan D), we can then write:

$$U(x_1 = 2, x_2 = 6, x_3 = 4) = U(x_1 = q, x_2 = 2, x_3 = 4). \quad (9)$$

Substituting Equation 5 in Equation 9, and cancelling terms yields:

$$w_2 = w_1 u_1(q). \quad (10)$$

**FIGURE F**

SCHEMATIC OF TRADEOFF QUESTION FOR PERSONALNESS AND QUALITY

| Plan C | Plan D |
|---|---|
| value : satisfactory (4) | value : satisfactory (4) |
| personalness : very good (6) | personalness : very poor (2) |
| quality : just adequate (2) | quality : ------- |

As $u_1(x_1)$ is known from Equation 8, this gives $w_2(q)$, which is a computable value. For example, if the consumer specifies that the quality of Plan D should be satisfactory ($q = 3$), then for $p_1 = 0.80$ we get $w_2/w_1 = 0.533$. A similar question and computation would yield $w_3/w_1$. Suppose $w_3/w_1 = 0.250$.

We now have the relative importances, but to determine $W$, we must first determine $w_1$. Remember we want to scale $U(x_1,x_2,x_3)$ between 0 and 1. To determine $w_1$, we must ask a lottery question with more than one attribute varying, or a tradeoff question with more than two attributes varying. For simplicity of arithmetic, we will ask a lottery question where all attributes vary between their extreme values ($x_k = 2$ vs. $x_k = 6$), schematically represented in Figure G).

In actual practice, the consumer responds to this lottery or tradeoff after having previously responded to a number of uniattributed lotteries or attribute tradeoffs. Probability wheels and 4- × 6-inch cards are used, and the interviewer must work carefully with the consumer to make sure he can visualize the alternatives represented by the lottery.

Suppose the consumer assigns a value of $P$ to the lottery. As $U$(Plan E) = $U$(Plan F), we can write:

$$U(x_1 = 6, x_2 = 2, x_3 = 2) = PU(x_1 = 6, x_2 = 6, x_3 = 6) + (1 - P) U(x_1 = 2, x_2 = 2, x_3 = 2). \quad (11)$$

Substituting Equation 5 in Equation 11 and, as $U(x_1,x_2,x_3)$ is scaled between 0 and 1, we get:

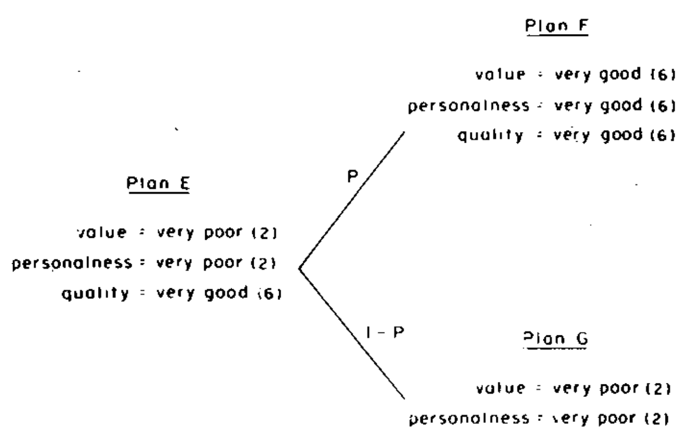
$$w_1 = P. \quad (12)$$

Equation 12 together with the ratios $w_2/w_1$ and $w_3/w_1$ yield the exact values for the $w_k$'s. For example, if $P = 0.65$, then $w_1 = 0.65$, $w_2 = 0.35$, and $w_3 = 0.16$.

The relative importances $w_2/w_1$ and $w_3/w_1$ have behavioral significance because they indicate the relative "value" a consumer would place on changing an attribute from its lowest level to its highest level, if all

**FIGURE G**

SCHEMATIC OF INTERACTION LOTTERY



other attributes were at their lowest levels. The exact values result from the choice of scale, e.g., setting $u(x_1 = 6, x_2 = 6, x_3 = 6) = 1$. Once the scale is selected, then the exact values have intuitive meaning relative to that scale, and together $w_1$, $w_2$, and $w_3$ indicate the degree of substitutability among the attributes. They also mathematically determine $W$.

To see this intuitively, suppose there is no substitutability or complementarity among the attributes. Then, if we independently move $x_1$ then $x_2$ then $x_3$ to their highest level, we get the total utility to its highest level. In this case (see Equation 5), $\sum_k w_k = 1$ and $W = 0$. If we cannot get the total utility to its highest level, i.e., $\sum_k w_k < 1$, then the attributes must be complementary and $W > 0$. For example, high quality would be more important to the consumer when there is more rather than less personalness. Finally, substitutability would imply $\sum_k w_k > 1$ and $W < 0$. These intuitive and mathematical relationships can be derived from Equation 5. (Keeney 1974.)

To algebraically determine $W$, we simply substitute Equation 5 into the normalization equation $u(x_1 = 6, x_2 = 6, x_3 = 6) = 1$. This yields:

$$w_1 + w_2 + w_3 + W(w_1w_2 + w_1w_3 + w_2w_3) + W^2w_1w_2w_3 = 1.0, \quad \textbf{(13)}$$

which is a quadratic equation that can be solved for $W$, since $w_1$, $w_2$, and $w_3$ are known. Keeney (1974) shows there is exactly one root of this polynomial in the relevant range, which is $W \in [-1,0)$ if $\sum_k w_k > 1$, $W \in [0,0]$ if $\sum_k w_k = 1$, and $W \in (0,\infty)$ if $\sum w_k < 1$. These results hold for an arbitrary number of dimensions. In our example, $w_1 + w_2 + w_3 = 1.16$ and, by Equation 13, $W = -0.431$.

This completes the numerical example. Each of the parameters was calculated from a consumer task closely related to the behavioral construct that the parameter measures. The three risk aversion parameters, $r_k$, came from lotteries that were risky with respect to one attribute; the relative importance, $w_k/w_l$, came from pairwise tradeoff questions; and the interaction coefficient, $W$, came from the exact values of the $w_k$'s, which came from a lottery question involving more than one varying attribute. Thus for three attributes, the parameters in Equation 5 were determined by three simple lotteries, two pairwise tradeoff questions, and one complex lottery. $K$ attributes would require $K$ simple lotteries, $K - 1$ tradeoffs and one complex lottery. In general, $2K$ nonredundant questions are required.

This example was for three attributes in a multiplicative, constantly risk averse utility function. The technique is extendible to more attributes and alternative functional forms. To verify assumptions, more questions are asked. To estimate, rather than algebraically solve for the parameters, different methods of analysis are required.

## REFERENCES


Brunswik, E. (1952). *The Conceptual Framework of Psychology*, Chicago: University of Chicago Press.

Carroll, J. Douglas (1972), "Individual Differences and Multidimensional Scaling," in *Multidimensional Scaling: Theory and Application in the Behavioral Sciences*, eds. R. N. Shepard, A. K. Romney, and S. Nerlove, New York: Seminar Press, Inc., 1.

Farquhar, Peter H. (1975), "A Fractional Hypercube Decomposition Theorem for Multiattribute Utility Functions," *Operations Research*, 23:5, 941–67.

——— (1977 a), "A Survey of Multiattribute Utility Theory and Applications," *Studies in Management Service*, 6, 59–89.

——— (1977 b), "Interdependent Criteria in Utility Analysis," working paper, Department of Industrial Engineering and Management Science, Northwestern University.

Fishbein, Martin (1967), "Attitudes and the Prediction of Behavior," in *Readings in Attitude Theory and Measurement*, ed. Martin Fishbein, New York: John Wiley & Sons, Inc.

Fishburn, Peter (1971), "Additive Representations of Real-Valued Functions on Subsets of Product Sets," *Journal of Mathematical Psychology*, 9, 225–36.

Friedman, Milton, and Savage, L. J. (1952), "The Expected-Utility Hypothesis and the Measurability of Utility," *Journal of Political Economy*, 60, 463–74.

Green, Paul E., and Devita, Michael (1975), "An Interaction Model of Consumer Utility," *Journal of Consumer Research*, 2, 146–53.

———, and Srinivasan, V. (1977), "Conjoint Analysis in Consumer Behavior: Status and Outlook," *Journal of Consumer Research*, 5, 103–23.

———, and Wind, Yoram. (1973), *Multiattribute Decisions in Marketing: A Measurement Approach*, Hinsdale, Ill: The Dryden Press.

Haley, Russell I. (1968), "Benefit Segmentation: A Decision-Oriented Research Tool," *Journal of Marketing*, 32, 30–5.

Hauser, John R. (1978), "Consumer Preference Axioms: Behavioral Postulates for Describing and Predicting Stochastic Choices," *Management Science*, 24, 1331–41.

———, and Shugan, Steven M. (1977), "Extended Conjoint Analysis with Intensity Measures and Computer Assisted Interviews," *Advances in Consumer Research*, 5.

———, and Urban, Glen L. (1977), "A Normative Methodology for Modeling Consumer Response to Innovation," *Operations Research*, 25:4, 579–619.

Herstein, I. N., and Milner, John (1953), "An Axiomatic Approach to Measurable Utility," *Econometrica*, 21, 291–7.

Jensen, Niels Erik (1967), "An Introduction to Bernoullian Utility Theory. 1. Utility Functions," *Swedish Journal of Economics*, 69, 163–83.

Kahneman, Daniel, and Tversky, Amos (forthcoming), "Prospect Theory: An Analysis of Decision Under Risk," presented at "Interfaces Between Marketing and Economics," University of Rochester, April 7–8, 1978, *Econometrica*.

Keeney, Ralph L. (1972), "Utility Functions for Multiattributed Consequences," *Management Science*, 18, 276–87.

——— (1974), "Multiplicative Utility Functions," *Operations Research*, 22:1, 22–3.

———, and Raiffa, Howard (1976). *Decision Analysis with Multiple Objectives*, New York: John Wiley & Sons, Inc.

Krantz, David H., Luce, R. Duncan, Suppes, Patrick, and Tversky, Amos (1971). *Foundations of Measurement*, Vol. 1, New York: Academic Press, p. 419.

Luce, R. Duncan (1959). *Individual Choice Behavior*, New York: John Wiley & Sons, Inc.

——— (1977), "The Choice Axiom After Twenty Years," *Journal of Mathematical Psychology*, 15, 215–33.

Marschak, Jacob (1950), "Rational Behavior, Uncertain Prospects, and Measurable Utility," *Econometrica*, 18, 111–41.

McFadden, Daniel (1970), "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. P. Zaremblea, New York: Academic Press, 105–42.

——— (1975), "The Revealed Preferences at a Government Bureaucracy: Theory," *The Bell Journal of Economics and Management Science*, 6, 2.

——— (forthcoming), "Econometric Models for Probabilistic Choice Among Products," presented at "Interfaces Between Marketing and Economics," University of Rochester, April 7–8, *Journal of Business*.

Raiffa, Howard (1970). *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*. Reading, MA: Addison-Wesley.

Richard, Scott F. (1975), "Risk Aversion, Utility, Independence, and Separable Utility Functions," *Management Science*, 22, 12–21.

Tversky, Amos (1967), "A General Theory of Polynomial Conjoint Measurement," *Journal of Mathematical Psychology*, 4, 1–20.

Von Neumann, John, and Morgenstern, Oscar (1947). *The Theory of Games and Economic Behavior*, 2nd ed., Princeton, N. J.: Princeton University Press.

Wilkie, William I., and Pessemier, Edgar A. (1973), "Issues in Marketing's Use of Multiattribute Attitude Models," *Journal of Marketing Research*, 10, 428–41.