DISCUSSION PAPER NO. 287

CONSUMER PREFERENCE AXIOMS:

Behavioral Postulates for Describing and Predicting Stochastic Choice */

by

John R. Hauser +/

July 1977

*/ (Revised) November 1976

+/ Department of Marketing
Graduate School of Marketing
Northwestern University

1. INTRODUCTION

Two important managerial problems addressed by marketing research are: (1) explaining how consumers form preferences and (2) predicting their purchase behavior. Explanatory models provide diagnostic information to managers so that they can modify demand by altering product characteristics, advertising appeals, or other aspects of the marketing strategy. Predictive models provide information to managers so that they can evaluate alternative strategies or plan production, inventory, and salesforce.

Four distinct streams of research in marketing and economics have addressed aspects of these problems. Consumer behavioralists have postulated and tested models which identify the process by which consumers form preferences. Von Neumann-Morgenstern utility theorists have axiomatically studied models to prescribe rational behavior. Both sets of models study behavior deterministically and at the level of the individual consumer. Stochastic modellers have postulated and tested models which identify the structure of a market and the distribution of preferences across the population. These models explain behavior stochastically and at the aggregate level. Finally, econometricians have postulated and estimated models based on observations of past behavior in an attempt to predict future behavior. These models explain behavior stochastically and at an intermediate level (individual choice predictions, but the same choice process for everyone).

These streams of research have often been viewed as competing, but in actuality they are complementary. Stochastic assumptions can be directly coupled with the axiomatic strengths of von Neumann-Morgenstern utility theory, the measurement strengths of consumer behavior, and the predictive strengths of econometrics to provide both explanation and prediction at the level of the individual consumer. This paper provides a common theory (definitions, axioms, and theorems) to combine these diverse disciplines and to develop a usable managerial tool which can: (1) identify the appropriate forms for preference models, (2) handle new products with uncertain attributes, (3) directly measure preference functions at the individual level, and (4) test fundamental behavioral assumptions at an understandable level.

The structure of the paper is to briefly review existing literature, present the formal preference theory, discuss how it relates to probability of choice models, and provide an empirical example and comparison with existing models.

2. EXISTING LITERATURE

Consumer behavior: The multi-attributed preference theories (reviewed by Wilkie and Pessimier [33]) have been devoted to models which predict preference or attitude toward a product as a weighted sum of a consumer's perceptions of the levels of the attributes describing that product. I.e.:

$$p_j = \sum_k I_k x_{jk} \qquad (1)$$

where p_j is a measure of the consumer's preference for product j, x_{jk} is the perceived level of attribute k for product j, and I_k is the importance of attribute k. This model implies that the consumer should deterministically select the product with the highest preference value. In practice, the correlations between preference and choice are not perfect but rather range from .1 (Sheth and Talarzyk [28]) to .8 (Ryan and Bonfield [27]). Other researchers have relaxed the restrictive linear form (Green and Rao [8], Green and Devita [7], and Johnson [17]) and have used conjoint measurement (Tversky [31]) on additive, multiplicative, and pairwise interactive models to estimate consumer preference from consumers' perceptions of a product's attribute levels. Nonetheless, they too have not predicted behavior with certainty.

Utility theory: Rather than arbitrarily selecting the algebraic structure of preference functions (also called utility functions), economists have proceeded deductively from verifiable postulates about an individual's preference ordering. In particular, the von Neumann-Morgenstern [32] postulates (reformulated by Friedman and Savage [6], Marschak [23], Herstein and Milner [14], Jensen [15], and possibly others) have been particularly useful in axiomatically specifying the conditions under which a unique+ preference scale exists. Later research identified "independence properties" which specified necessary and/or sufficient conditions under which preferences for products could be represented by parametric functions of the attribute levels. (See for example Farquhar [3,4], Fishburn [5], Keeney [20,19], Keeney and Raiffa [21], and Raiffa [26].) Since the parameterized functions are axiomatically derived from basic behavioral assumptions, the parameters provide explicit indications of trade-offs, risk aversion, and interactions among the levels of the product's abbributes.

Stochastic models: In 1974, Bass [1] callenged the field of consumer behavior by stating:

> "Although it is heresey, in some circles, honesty compels one to question the fundamental premise that all behavior is caused. If there is a stochastic element in the brain which influences choice, then it is not possible, even in principle, to predict or to understand completely the choice behavior of individual consumers."

+Unique to an additive and multiplicative constant.

He goes on to postulate and test empirically a theory of stochastic preference and brand switching that tries to predict aggregate stochastic behavior while making no claims about a specific individual's behavior. Bass' model, like the related entropy (Herniter [13]) and Hendry models (Butler and Butler [2]), does not try to measure, model, or predict preference as a function of the perceived levels of product attributes. Thus its use is limited to identifying aggregate market characteristics. Further work is necessary to make it sensitive to decision variables, such as advertising, promotion, or product characteristics, which effect the perceived levels of product attributes.

Econometrics: Recognizing that for practical purposes it is impossible ever to specify fully a utility function, econometricians have postulated that the "true" utility function can only be partially observed. McFadden [24,25] has operationalized this concept by postulating that the true ordinal utility of product j, u_j, consists of an observation portion, v_j, plus an error term, e_j. In other words:

$$u_j = v_j + e_j \tag{2}$$

Using this error structure and the definition of ordinal utility, McFadden shows how to derive choice probabilities. In fact, if the error terms are identically distributed, independent, zero-meaned Weibull random variables, the probability, L_j, of choosing product j is given by the analytically simple logit model. I.e.:

$$L_j = \exp(v_j) \;/\; \sum_{\ell} \exp(v_{\ell}) \tag{3}$$

where the denomimator is summed over all products in the consumer's choice set. In practice, this model has proven powerful because if v_j is linear in a set of parameters, the parameters can be easily and consistantly estimated by maximum likelihood techniques (McFadden [24]). Although McFadden's theory allows arbitrary functions for v_j as long as the functions are linear in their parameters, most empirical applications have dealt with functions represented by weighted sums of attribute levels.

Discussion

Examining the various approaches to understanding or predicting consumer preferences, we see a diverse set of approaches. Consumer behavior theory postulates preference models and experimentally tests them. Prescriptive utility theory ignores the prediction problem but develops powerful deductive theory to identify the appropriate preference models. Econometrics admits imperfection and statistically searches for preference models which explain as much behavior as possible.

The common goal of each of these multi-attributed preference models is not to predict behavior deterministically, but rather to explain and predict as much about behavior as is possible. To do this successfully, these models should use preference functions that are as strong as is feasible but which explicitly incorporate the concept of stochastic preference. Thus, what we would like to do is combine the deductive power of prescriptive utility theory with the stochastic behavior models of econometrics and the measurement strengths of consumer behavior.

For example, we might try using a von Neumann-Morgenstern utility function in equation (3) to estimate probabilistics for Bass' model. Unfortunately, this simple combination may not work because each theory or set of theories is based on behavioral assumptions which may or may not conflict. For example, neither attitude models nor the von Neumann-Morgenstern models specify the measurement scale, whereas equation (3) is extremely sensitive to the scale of v_j; and even if the scale were estimated, it remains to be shown that the theories are consistant.

This paper sets forth an axiomatic structure which draws on the strengths of each theory to build a comprehensive theory for describing choice. The theory extends von Neumann-Morgenstern theory to describe stochastic choice; it provides an axiomatic structure based on fundamental assumptions to identify functional forms for consumer behavior models; it expands the econometric models to handle non-linear utility functions which vary individual by individual; and it provides a behavior-based link from individuals' perceptions of product attribute levels to aggregate stochastic choice.

3. FORMAL THEORY

This section begins by defining the mathematical goals of the theory.

Formal Definitions

Let $A = \{a_1, a_2, \ldots, a_J\}$ be set of choice alternatives; let x_k be a performance measure, such as "quality," describing at least one alternative, $a_j \epsilon A$.

Let $X = \{X_1, X_2, \ldots, X_K\}$ be a complete set of performance measures and let $\underline{x}_j = \{x_{1j}, x_{2j}, \ldots, x_{Kj}\}$ be the values that the performance measures take on for a deterministic alternative a_j. Let $\underline{x}_{\bar{j}} = \{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{j-1}, \underline{x}_{j+1}, \ldots \underline{x}_J\}$ In other words, $\underline{x}_{\bar{j}}$ is the set of performance measures for all alternatives except a_j. Furthermore, let $p_i(a_j | \underline{x}_1, \underline{x}_2, \ldots, \underline{x}_J; \underline{\lambda}_i)$ be the probability that individual i chooses alternative a_j given specific levels of the performance measures $\{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_J\}$, and given a vector of real-valued "preference" parameters, $\underline{\lambda}_i$, for individual i.

Compaction: What we need is a real-valued function, call it a compaction function,[+] which tells us how individual i evaluates the performance measures to form his (stochastic) preferences. In particular, if we hold all other products fixed the compaction function for a given product should produce numbers which are monotonic in the choice probabilities. Formally:

Definition 1: A real-valued function, $c_j(\underline{x}_j, \underline{\lambda}_i)$, is said to be a compaction function if for any fixed $\underline{x}_{\bar{j}}$

$$c_j(\underline{x}_j,\underline{\lambda}_i) > c_j(\underline{x}_j',\underline{\lambda}_i)$$

implies

$$p_i(a_j|\underline{x}_j,\underline{x}_{\bar{j}};\underline{\lambda}_i) \geq p_i(a_j|\underline{x}_j',\underline{x}_{\bar{j}};\underline{\lambda}_i)$$

and

$$c_j(\underline{x}_j,\underline{\lambda}_i) = c_j(\underline{x}_j',\underline{\lambda}_i)$$

implies

$$p_i(a_j|\underline{x}_j,\underline{x}_{\bar{j}};\underline{\lambda}_i) = p_i(a_j|\underline{x}_j',\underline{x}_{\bar{j}};\underline{\lambda}_i).$$

Uniformity: An analyst tries to identify a set of performance measures which are complete. He then would hope that tradeoffs and interdependencies among those performance measures would not be alternative specific. In other

---

[+]The word compaction has been chosen rather than utility because we are not attributing utility properties to the function. Semantically, the idea is to "compact" the K attribute measures into 1 scalar measure of goodness for each choice alternative.

words, knowing the performance measures, $\underline{x}_j$, for alternative a_j and the preference parameters, $\underline{\lambda}_i$, for individual i would be sufficient to compute individual i's compaction value, c_j, for alternative a_j. Thus a uniform compaction function has the same functional form for all alternatives (drop the j subscript on $c_j(\underline{x}_j;\underline{\lambda}_i)$). Formally:

Definition 2: A compaction function is uniform for an alternative set, A, if

$$c_j(\underline{x}_j,\underline{\lambda}_i) = c(\underline{x}_j,\underline{\lambda}_i) \qquad \text{for all } a_j \epsilon A.$$

Notice that alternative specific <u>terms</u> can be included as performance measures as long as the <u>functional form</u> is the same for all alternatives in A.

<u>Symmetry</u>: Uniformity deals with the functional form of the compaction function, symmetry deals with the functional form of the conditional probability law. Symmetry implies that a specific value of the scalar measure of goodness has the same implications for each alternative. To better understand this consider the new notation:

$p_i(a_j|c_1,c_2,\ldots, c_J)$ = the probability of choosing alternative a_j given that $c_1(\underline{x}_1,\underline{\lambda}_i) = c_1, c_2(\underline{x}_2,\underline{\lambda}_i) = c_2$, etc. This notation is consistent by the definition of compaction. Furthermore, define $c_{\overline{jk}} = \{c_1,c_2,\ldots, c_{j-1}, c_{j+1},\ldots, c_{k-1},c_{k+1},\ldots, c_J\}$, that is, $c_{\overline{jk}}$ is the set of all scalar measures of goodness (for individual i) except c_j and c_k. Thus formally:

Definition 3: A compaction function (and the probability law it evokes) is said to be <u>symmetric</u> for an alternative set A if for all pairs of j and k, $a_j,a_k \epsilon A$:

$$p_i(a_j|c_j = x, c_k = y, c_{\overline{jk}}) = p_i(a_k|c_j = y, c_k = x, c_{\overline{jk}})$$

and

$$p_i(a_\ell|c_j = x, c_k = y, c_{\overline{jk}}) = p_i(a_\ell|c_j = y, c_k = x, c_{\overline{jk}})$$

for all $a_\ell \neq a_j, a_k$.

Less formally, switching the compaction values for any j-k pair switches the choice probabilities for j and k but leaves all other choice probabilities unchanged.

Stochastic Preference Defined

As you can see by the definitions of compaction, we are looking for a function which captures the essence of an individual's evaluation process such that the compaction values are sufficient to predict probabilities. This compaction function is parallel to a utility function except that it predicts choice probabilities rather than deterministic choice. Let us formally define stochastic preference and stochastic indifference. Note that the definition is conditioned on the alternative set.

Definition 4: Let A be a set of alternatives, let a_j, a_k be elements of A, then $>_A$ is a stochastic preference operator on AxA if $(a_j >_A a_k)$ **implies** the probability of choosing a_j from A is greater than the probability of choosing a_k from A.

Define stochastic indifference, written $a_j \sim a_k$, for equal probabilities and make the obvious definitions for $\gtrsim_A$, $<_A$, and $\lesssim_A$.

Definition 4 deals with stochastic preferences but deterministic alternatives. In practice, consumers rarely have perfect information about the attributes of products. Thus, we would like to consider products which are not perfectly perceived. To this end we generalize the alternative set to include alternatives with uncertain characteristics. Those familiar with utility theory will notice that the following definition is simply a von Neumann-Morgenstern standard gamble.

Definition 5: A <u>lottery</u>, $L(a_j,a_k;p)$: $AxAx[0,1] \rightarrow A^*$, is an alternative which has the characteristics of a_j with probability p, and the characteristics of a_k with probability (1-p). (A^* is the range of L). (See Figure 1.)

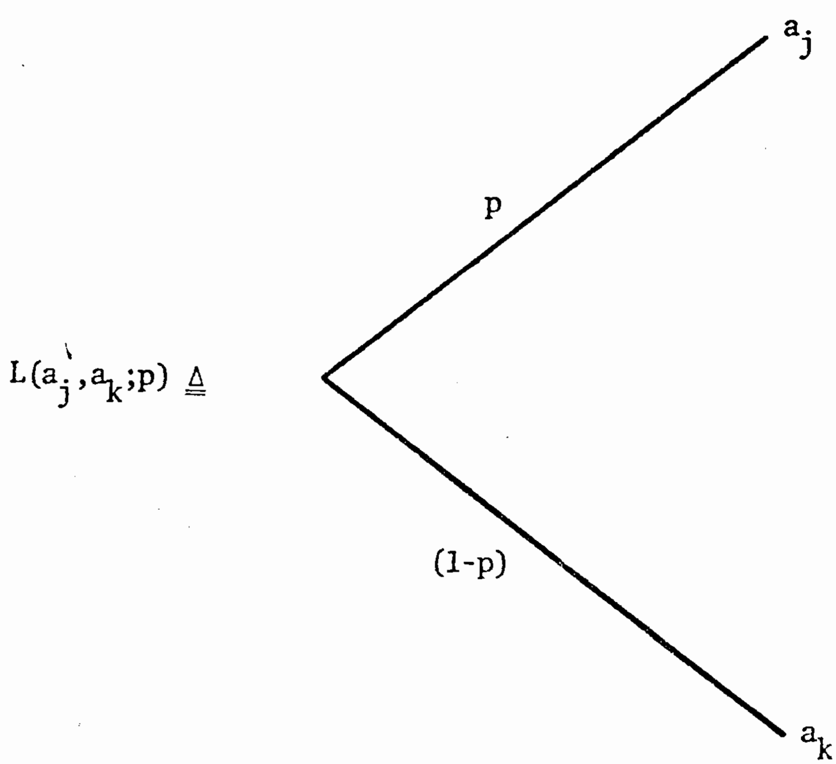

Figure 1: Lottery Definition

We now have the notation to present the axioms. The first three axioms are the von Neumann-Morgenstern [32] axioms restated for stochastic preference, the fourth axiom is the choice axiom which deals with the structure of the choice set. Intuitive explanations follow the formal statements.

The Axioms

Suppose A^*, $>$, $\sim$, and L satisfy the following axioms[+]:

Axiom 1: $>$ is a complete ordering on A^*.

(a) For any two a_j, a_k exactly one of the following holds.

$$a_j > a_k,\ a_j \sim a_k,\ a_j < a_k$$

(b) $a_j > a_k$ and $a_k > a_\ell$ implies $a_j > a_\ell$

(c) $a_j \sim a_k$ and $a_k \sim a_\ell$ implies $a_j \sim a_\ell$

Axiom 2: Ordering and combining:

(a) $a_j \left\{\begin{matrix} > \\ < \end{matrix}\right\} a_k$ implies $a_j \left\{\begin{matrix} > \\ < \end{matrix}\right\} L(a_j, a_k; p)$ for all $p \varepsilon (0,1)$.

(b) $a_j > a_k > a_\ell$ implies the existence of $p_1, p_2, p_3\ \varepsilon (0,1)$ such that

$$L(a_j, a_\ell; p_1) < a_k$$

$$L(a_j, a_\ell; p_2) \sim a_k$$

$$L(a_j, a_\ell; p_3) > a_k$$

[+]The A^* subscript on $>_{A^*}$ is temporarily dropped for notational simplicity.

Axiom 3: Algebra of combining

(a) $L(a_j,a_k;p) \sim L(a_k,a_j;1-p)$

(b) $L[L(a_j,a_k;p), a_k;q] \sim L(a_j,a_k;pq)$

Axiom 4: Choice axiom. Let α be any finite subset of A^*, let $a_j,a_k,a_j',a_k' \in \alpha \subseteq A^*$, then $a_j \sim_{A^*} a_j'$ and $a_k \sim_{A^*} a_k'$ implies $\text{Prob}\{a_j \text{ from } \alpha-a_k\} = \text{Prob}\{a_j' \text{ from } \alpha-a_k'\}$ where $\alpha-a_k$ is the set α with the element a_k deleted.

Interpretation of the Axioms

Axiom 1 (Complete ordering): (a) In utility theory this is a reasonably strong assumption, i.e., that an individual can state his preferences and that they are temporally stable. The new preference definition allows stochastic behavior, thus the new interpretation is that an individual's "average" behavior has no unmeasurable long-term trend. (b+c) This property is actually induced by the preference definition because > and = are transitive for the real numbers. It is stated explicitly to maintain a parallel with the utility axioms.

Axiom 2 (Ordering and combining): (a) This states simply that if a_k is stochastically preferred to a_j, then a lottery with even a slight chance of a_k is preferred to a_j. ("Losing" the lottery gives a_j.) (b) If $a_j > a_k > a_\ell$, then given a lottery, $L(a_j,a_\ell;p_1)$, the influence of a_j can be made sufficiently small (p_1 close to 0) such that a_k is still preferred

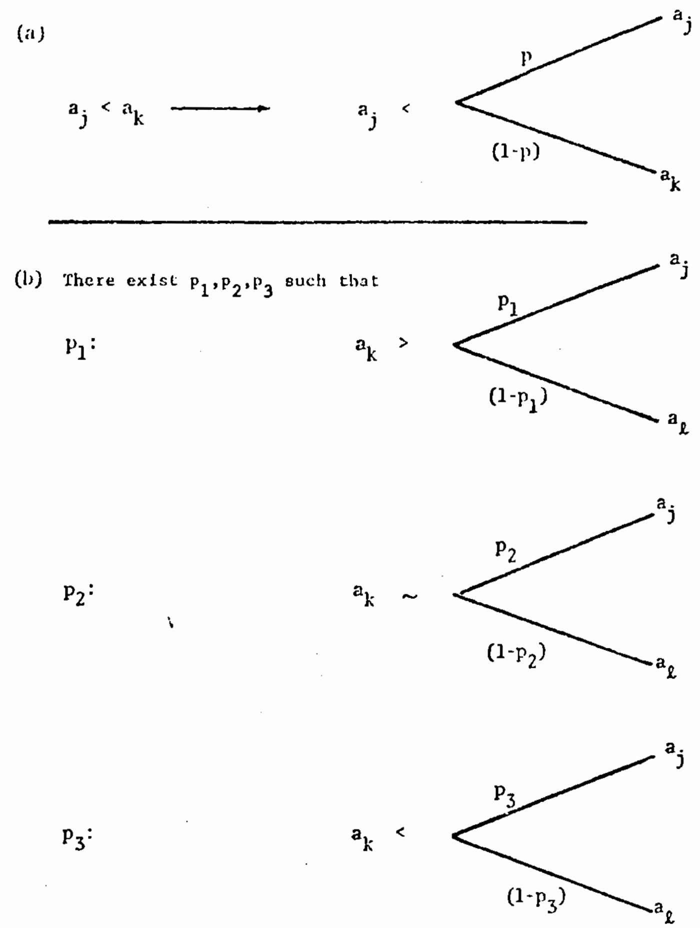

Figure 2: Schematic of Ordering and Combining Axiom

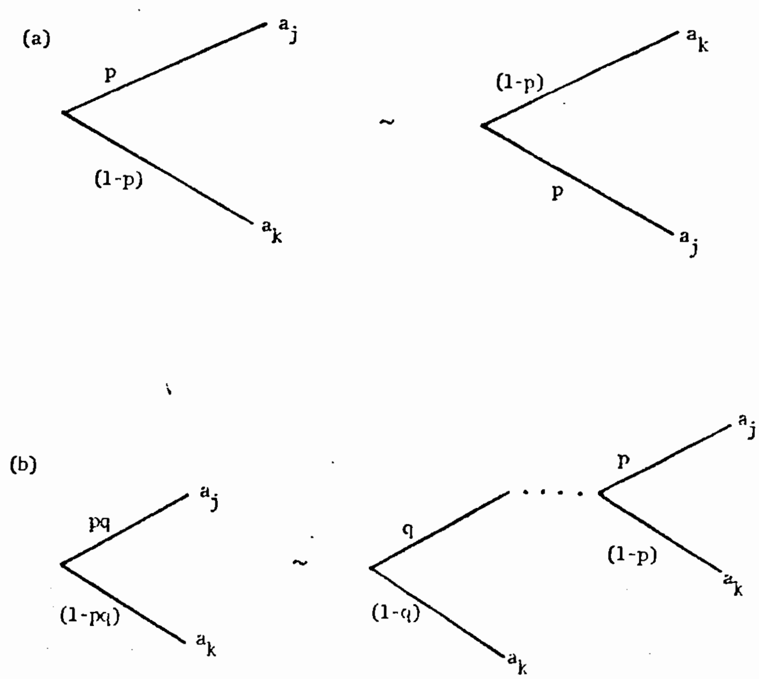

Figure 3: Schematic of Algebra of Combining Axiom

to the lottery. (Review Figure 2.) Furthermore, each individual can conceive of some probability, p_2, which makes him stochastically indifferent between the middle alternative, a_k, and a lottery involving the extreme alternatives, a_j and a_ℓ. Taken together, parts (a) and (b) of this axiom imply a reasonable continuity assumption. Since they are stated for stochastic preference, these axioms imply properties of the probability model.

Axiom 3 (Algebra of combining): (a) This states simply that the lottery operation is commutative, i.e., it does not matter in which order the elements of the lottery are named. (b) This statement of associativity is perhaps the strongest assumption in the utility axioms and hence in our axioms. It states that a series of successive lotteries can be treated as an equivalent one-step lottery. In other words, it states that every individual can conceive of a complex lottery and that he will rationally react to it as if it were a simple lottery with equivalent probabilities.

Axiom 4 (Choice axiom): This axiom states that if the probability of choosing an alternative, a_j, is equal to the probability of choosing another alternative, a_j', when all alternatives are available, then for any subset of the alternatives this equality of probabilities remains the same if some alternatives (other than a_j,a_j') are deleted from consideration. Furthermore, if a_k and a_k' are indifferent on A, then deletion of one or the other is equivalent in terms of stochastic indifference on the respective subsets. In other words, if two alternatives are equivalent on the entire choice set, then they are equivalent in their presence or their absense from any subset. This is certainly a reasonable assumption for distinct choices, but for certain types of choices, particularly hierarchical choices, it can break down.

For example, suppose a student has the following choice probabilities for health care delivery: Boston Group Practice (BGP), .4; private care with Dr. Jones, .3; private care with Dr. Smith, .3; and suppose these choices represent an exhaustive list. Now suppose Dr. Smith is no longer available. Will BGP still be stochastically preferred to Dr. Jones? Maybe, but perhaps the student's decision rule is to first choose between group practice and private care and then randomly select a doctor if he decides on private care. This might imply that Dr. Jones>BGP (.6>.4) after Dr. Smith departs.

This example cautions us not to blindly apply models derived from the axioms. Instead, the axioms must be verified before models are built, and if the choice process is hierarchical (sequential) it must be modeled as such. There are a number of ways to identify hierarchies in the choice set. See the stochastic models referenced earlier as well as Kalwani and Morrison [18]. Axiom 4 is needed because alternatives will be represented by sets of performance measures and compaction functions will be inferred from questions about stochastic indifference among abstract alternatives (represented by values for the performance measures). Thus, compaction functions will be determined on uncountable choice sets, $\{X_1, X_2, \ldots, X_N\}$, and applied to finite subsets, $\{a_1, a_2, \ldots, a_j\}$.

The next subsection discusses some mathematical implications of the axioms.

Existence and Uniqueness Theorems

The first and most significant implication of the axioms is the existence of a real-valued function on the expanded alternative set, A^*, which preserves (stochastic) preference and for which mathematical expectation applies. The proof of this result is quite tedious and exactly parallels the proof for utility functions contained in the appendix of von Neumann and Morgenstern. Thus, it is stated here without formal proof. (Let R represent the real numbers.)

Theorem 1 (Existence): There exists a real-valued function, c^*, on A^*, $c^*:A^* \rightarrow R$, with the following properties:

$$\text{(a)} \quad a_j \begin{Bmatrix} > \\ \sim \\ < \end{Bmatrix}_{A^*} a_k \iff c^*(a_j) \begin{Bmatrix} > \\ = \\ < \end{Bmatrix}_R c^*(a_k)$$

$$\text{(b)} \quad c^*[L(a_j,a_k;p)] = p\, c^*(a_j) + (1-p)\, c^*(a_k)$$

where $a_j, a_k \in A^*$, $p \in [0,1]$.

The function, c^*, looks similar to a preferential compaction function, but with the additional property of mathematical expectation being appropriate. Later, a related function will be shown to be a compaction function, but first let us investigate how c^*'s properties behave under transformation.

Theorem 2 (Uniqueness): The function $c^*:A^* \rightarrow R$ is unique up to a positive linear transformation.

Proof: Suppose there exists a function $d^*:A^* \rightarrow R$. Then there must exist a function $f:R \rightarrow R$ such that for any $a_j \in A^*$, $f(c^*(a_j)) = d^*(a_j)$. (See Figure 4.) Since c^* and d^* both satisfy the properties of theorem 1

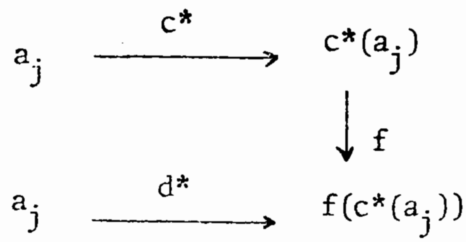

Figure 4: Schematic of Uniqueness Proof

then for $x, y \in R$

(a) $x \begin{Bmatrix} > \\ = \\ < \end{Bmatrix}_R y$ implies $f(x) \begin{Bmatrix} > \\ = \\ < \end{Bmatrix}_R f(y)$

(b) $f(px + (1-p)y) = pf(x) + (1-p)f(y)$

But (b) is just the definition of a linear function, thus $f(x)$ is linear in x. Furthermore, (a) implies that $f(x)$ is monotonically increasing in x, thus $f(x)$ must be a positive linear transformation. Thus if any function, d^*, satisfies theorem 1, it must be a linear transformation of another function, c^*, which satisfies theorem 1.

(This proof is similar to the uniqueness proof in utility theory.)

Empirical Use Requires Representation by Performance Measures

So far axioms 1, 2, and 3 have implied the existence and uniqueness of a scale function, c^*, which indicates stochastic preference over A^*. This is an interesting result but the goal of a compaction function is not simply to predict stochastic preference for products but also to indicate how consumers make judgments relative to attributes that describe the products. To this end, definition 1 defined compaction in terms of an attribute set, X. For empirical use, we would like to have consumers indicate stochastic preference for abstract alternatives represented by elements of X. We would then hope that if particular vectors of attribute levels, say $\underline{x}_\ell$ and $\underline{x}_m$, are realized as products, say a_ℓ and a_m, then judgments relative to X would be valid for the expanded set of products which now includes a_ℓ and a_m, i.e., for $\alpha = A \cup \{a_\ell, a_m\}$. For example, if $\underline{x}_\ell >_X x_m$ then hopefully $a_\ell >_\alpha a_m$. If this is true, then the preference information captured by a compaction function could be used to understand and predict consumer response to potential products. This assumption is formalized by the following axiom:

Axiom 5: Abstract Alternatives

(a) Consumers can indicate stochastic preference and indifference relative to X.

(b) There exists a vector-valued function, $\underline{g}: A \rightarrow X \subseteq R^N$ such that

$$\underline{g}(a_j) \neq \underline{g}(a_\ell) \qquad \text{for all } a_j \neq a_\ell$$

(c) If $\underline{x}_m$ is realized as a physical product, a_m, and if

$$\underline{x}_m \begin{Bmatrix} > \\ \approx \\ < \end{Bmatrix}_X \underline{x}_j \quad \text{then} \quad a_m \begin{Bmatrix} > \\ \approx \\ < \end{Bmatrix}_\alpha a_j \qquad \text{for all}$$

$$a_j \in \alpha = A \cup \{a_m\}. \quad \text{Where } \underline{x}_j = \underline{g}(a_j) \text{ and } a_m \neq a_j.$$

We can now construct a compaction function which has the desired properties.

Theorem 3: Let $A = \{a_1, a_2, \ldots, a_J\}$ be set of two or more products. Suppose there exists a complete set of performance measures, X, such that axiom 5 holds. Suppose axioms 1, 2, 3, and 4 hold on $A \cup B$ for all finite $B \subseteq \underline{g}^{-1}(X)$. ($\underline{g}^{-1}(X)$ is the inverse image of X.) Let $c: X \to R$ be a real-valued function on X such that $c(\underline{g}(a_m)) = c^*(a_j)$ for all $a_j \in A \cup B$, then c is a uniform, symmetric compaction function on X.

Proof: (Compaction) By axioms 1, 2, and 3 applied to $A \cup B$ for any $B \subseteq \underline{g}^{-1}(X)$, there exists a $c^*_{A \cup B}$ relative to $A \cup B$. Since by axiom 4, Prob$\{a_j$ from $(A \cup B)^*\}$ $\begin{Bmatrix} > \\ = \\ < \end{Bmatrix}$ Prob$\{a_\ell$ from $(A \cup B)^*\}$ implies Prob$\{a_j$ from $A^*\}$ $\begin{Bmatrix} > \\ = \\ < \end{Bmatrix}$ Prob$\{a_\ell$ from $A^*\}$ (note $A^* \subseteq (A \cup B)^*$) then $c^*_{A \cup B}$ satisfies theorem 1 for $a_j \in A^*$ and thus is a positive linear transform of c^*_A. But by hypothesis $c(\underline{g}(a_\ell)) = c^*_{A \cup B}$ and $c(\underline{g}(a_\ell)) = c^*_A$ for a least two a_ℓ such that $a_\ell \in A \subseteq A \cup B$. Thus, the appropriate $c^*_{A \cup B}$ is the particular transformation such that $c^*_{A \cup B}(a_\ell) = c^*_A(a_\ell) = c(\underline{g}(a_\ell))$ for all $a_\ell \in A$. Suppose $a_j \neq a_m$ are contained in A, then since $c(\underline{g}(a_j)) = c^*(a_j)$ and $c(\underline{g}(a_m)) = c^*(a_m)$, c satisfies the definition of compaction (see theorem 1). Suppose $a_j \in A$, $a_m \notin A$ but $a_m \in \underline{g}^{-1}(X)$. Let $B = \{a_m\}$. Suppose $c(\underline{g}(a_m)) = c(\underline{g}(a_j))$. Then by axiom 5b $c^*_{A \cup B}(a_m) = c^*_{A \cup B}(a_j)$. Thus by above and by theorem 1 Prob$\{a_m$ from $A \cup \{a_m\}\}$ = Prob$\{a_j$ from $A \cup \{a_m\}\}$. By axiom 4 Prob$\{a_m$ from $A \cup \{a_m\} - a_j\}$ = Prob$\{a_j$ from $A \cup \{a_m\} - a_m\}$. Thus, by switching notations we have shown that $p_i(a_m | \underline{x}_m, \underline{x}_{\bar{j}}) = p_i(a_j | \underline{x}_j, \underline{x}_{\bar{j}})$ is implied by $c(\underline{x}_m) = c(\underline{x}_j)$. The proof is similar but more tedious for $c(\underline{g}(a_m)) \begin{Bmatrix} > \\ < \end{Bmatrix} c(\underline{g}(a_j))$. You must first show the obvious result that axioms 1 to 3 allow you to extend axiom 4 to the case of $a_j \begin{Bmatrix} > \\ < \end{Bmatrix}_A a_j'$. Thus, c is a compaction function for $a_j, a_m \in A$ and for $a_j \in A, a_m \notin A$. The proof is similar for $a_j, a_m \notin A$.

(Uniformity): Since $c(\underline{x}_j) = c(\underline{g}(a_j)) = c^*(a_j)$ for any $a_j \in A$, c is uniform.

(Symmetry): Suppose we have a choice set $\alpha = \{a_j', a_\ell'\} \cup A$ with $a_j', a_k' \in g^{-1}(X)$. Suppose $c^*_\alpha(a_j) = c^*_\alpha(a_k') = y$ and $c^*_\alpha(a_k) = c^*_\alpha(a_j') = z$. By theorem 1, $a_j \sim_\alpha a_k'$ and $a_k \sim_\alpha a_j'$. Thus, by axiom 4 Prob$\{\underline{a}_j$ from $\alpha - a_j'\}$ =

Prob$\{a_k{}'$ from $\alpha-a_k\}$ and again Prob$\{a_j$ from $\alpha-a_j{}'-a_k{}'\}$ = Prob$\{a_k{}'$ from $\alpha-a_j-a_k\}$. But by switching notation again, the last statement just says $p_i(a_j | c_j = y, c_k = z, c_{\overline{jk}}) = p_i(a_k{}' | c_j{}' = z, c_k{}' = y, c_{\overline{jk}})$ since c is a compaction function. This is the definition of symmetry.

Summary of Formal Theory

Axioms 1 through 5 identify a set of fundamental behavioral postulates which are sufficient for the existence and uniqueness of a uniform, symmetric compaction function on a set of product attributes. The theory sets up a rigorous framework for the measurement of "utility" or compaction functions. But "utility" measurement is not new to marketing. For example, Green and Rao [8] and Green and Wind [9] in some of their applications ask consumers to rank order alternatives represented by levels of product attributes, Johnson [16] asks consumers to make pairwise judgments relative to alternatives represented by levels of performance measures, and Hauser and Urban [11] ask consumers to rank order alternatives represented by concept descriptions.+ Each of these applications have successfully produced managerial insight. But these applications can be strengthened as follows.

Functional form: In marketing (conjoint measurement, preference regression, and attitude theory) and in econometrics an important analytic judgment is the functional form of the estimated scale function. Von Neumann-Morgenstern theory (see Farquhar [4]) identifies which functional forms are appropriate under which conditions, but perhaps more importantly

+Green and Rao, Green and Wind, and Johnson base their measurements on conjoint theory which allows a "compaction" function to be measured if it exists and the form is known.

identifies parametric forms whose parameters explicitly identify important effects such as risk aversion and attribute interaction.

Risky alternatives: Both marketing and econometrics have dealt primarily with deterministic products, but for new products consumers do not perceive the levels of the attributes with certainty. As market research applications move to major products such as health care, automobile ownership, housing decisions, and durables, managers will want to measure risk preferences and analysts will want to have compaction functions which are appropriate for risky alternatives. Von Neumann-Morgenstern theory makes it possible to measure compaction functions which deal explicitly with the problem of products with uncertain attribute levels.

Direct measurement: Identification of an appropriate parametric functional form for the compaction function makes direct measurement possible. For example, suppose it is known that $c(\underline{x}_j,\underline{\lambda}_i) = \lambda_{i0} + \lambda_{i1}(x_{j1} + \lambda_{i2}\, x_{j2} + \Lambda_i\, x_{j1}x_{j2})$, $(\lambda_{i1} > 0)$. Suppose further that individual i is indifferent between pairs: $\{x_{j1} = 10,\ x_{j2} = 0\} \sim \{x_{j1} = 0,\ x_{j2} = 5\}$ and $\{x_{j1} = 6,\ x_{j2} = 4\} \sim \{x_{j1} = 3,\ x_{j2} = 6\}$. The first indifference pair gives $\lambda_{i2} = 2$, the second indifference pair gives $\Lambda_i = 1/6$. These two tradeoff questions have completely specified $c(\underline{x}_j,\underline{\lambda}_i)$. (Since a von Neumann-Morgenstern compaction function is unique to a positive linear transformation, λ_0 and λ_{i1} are arbitrary.) This example uses an extremely naive functional form, realistic applications require careful measurement and more tradeoff questions. For empirical measurement of a 10 parameter, non-linear form see Hauser and Urban [12].

Fundatmental assumptions: It is extremely easy to postulate the existence of a compaction function and almost as easy to postulate a functional form. Axioms 1 through 5 give sufficient conditions for existence but they are given at a level that is realistic and testable. Note that these axioms are somewhat less restrictive than the original von Neumann-Morgenstern axioms because they require stochastic preference rather than deterministic preference, but in some cases they are more restrictive because they require analysts to identify hierarchies in the choice set. Since theorem 3 proves only sufficiency, it remains for future theory to relax these axioms and search for necessary conditions.

4. PROBABILITY OF CHOICE MODEL

The proceeding section presented the formal theory which enables market researchers to use the strength of von Neumann-Morgenstern theory to improve the measurement of consumer preferences. But this is only half of the story. The compaction function explains preference, we need a probability function to predict choice. This is a difficult problem with no simple solution. We will now present two pragmatic solutions, each of which is theoretically incomplete Finally, a general model is presented that is consistant with axioms 1 through 5 but not yet practical. It is up to the practitioner to weigh the relative benefits to choose the appropriate practical technique. It remains for future research to develop a model that is both theoretically complete and practical.

Econometrics

Perhaps the most widespread and empirically most powerful probability of choice model is McFadden's multinomial logit model. As was indicated earlier, this model assumes an ordinal utility function and a particular error structure which gives the following model:

$$\text{Prob}\{a_j \text{ from } A\} = \exp(v_{ij}) \;/\; \sum_{a_\ell \in A} \exp\,(v_{i\ell}) \tag{4}$$

Where the subscript i is now added to indicate a different set of compaction values for each individual in the population. Note that in this model v_{ij} is a symmetric compaction function, but the von Neumann-Morgenstern properties, axioms 2 and 3, are not guaranteed for v_{ij}. It is nonetheless possible that v_j is a monotonic transformation of a von Neumann-Morgenstern compaction function, $c(\underline{x}_j,\underline{\lambda}_i)$. I.E., $v_{ij} = f[c(\underline{x}_j,\underline{\lambda}_i)]$. It is empirically feasible to parameterize f and estimate the appropriate parameters. [A particularly simple f is the range adjusting model, $v_{ij} = \beta\ c(\underline{x}_j,\underline{\lambda}_i)$.] This two-step process has the advantages of measuring individual specific parameters, $\underline{\lambda}_i$, for the compaction function and of using a functional form for $c(\underline{x}_j,\underline{\lambda}_i)$ which explicitly measures important effects. The disadvantage is that f is now arbitrary and the postulates of the two theories are not entirely compatible.

Ranked Probability Model

Axioms 1 through 5 guarantee functions which rank order risky alternatives in terms of stochastic preference. Suppose we work with this property and no other, i.e., suppose we use the property that von Neumann-Morgenstern compaction functions rank order products, with or without uncertain attributes, in terms of their probability of being selected. The most naive model of this form is to assume there exists a probability, p_1, that an individual will choose his first ranked product, a probability, p_2, that an individual will choose his second ranked product, and so on. These probabilities could be observed for a sample situation and used to predict for situations in which new products are introduced.

This model was tested on two empirical data sets. In the first test, the dependent variable was self-reported last brand purchased (299 consumers, 18 brands of deodorant, see Silk and Urban [29]) and the compaction values were ratio scaled preference measures obtained from constant sum paired comparisons (Torgensen [30]). The ranked probabilities were

$p_1 = .83,\ p_2 = .15,\ p_3 = .02,\ p_4 = 0,\ p_5 = 0, \ldots.$

In this case, the ranked probability model explains 82.1% of the uncertainty+ while a range adjusting logit model explained 79.6% of the uncertainty.

In the second test, the dependent variable was first preference (76 consumers, 4 Health Maintenance Organizations, see Hauser and Urban [12]) and the compaction values were von Neumann-Morgenstern utility functions assessed over measures of "quality," "convenience," "personalness," and "value." The ranked probabilities were $p_1 = .52,\ p_2 = .21,\ p_3 = .19,\ p_4 = .08.$ In this case, the ranked probability model explained 14.5% of the uncertainty while a range adjusting logit model explained 11.6% of the uncertainty.

Clearly, these are weak tests of the ranked probability model since the range adjusting logit model is one of many possible monotonic transformations, but these tests do indicate that the rank order effect is worth further investigation.

General Probability Model

So far, two empirical approximations have been suggested for using von Neumann-Morgenstern compation functions to estimate choice probabilities. Each approach is feasible for empirical applications, but the range adjusting

---

+Uncertainty is measured by the entropy of the system and explanation is measured by the information provided by the model. See Hauser [10].

logit model sacrifices some of the theoretical rigor of axioms 1 through 5 while the ranked probability model sacrifices some of the power of axioms 1 through 5. This section examines the general requirements of the axioms and develops a generalized form for a probability model which is consistent with those axioms. Since this general model is not yet feasible, it is presented purely to spark research.

We begin the discussion by re-examining equation 4. To empirically use this model, we must estimate f for a segment of the population, call it $\mathcal{S}$. A consequence, among others, of this model is that any two individuals with the same vector of compaction values will have the same set of choice probabilities. I.e., $\{v_{i1}, v_{i2}, \ldots, v_{iJ}\} = \{v_{h1}, v_{h2}, \ldots, v_{hJ}\}$ implies $\{L_{i1}, L_{i2}, \ldots, L_{iJ}\} = \{L_{h1}, L_{h2}, \ldots, L_{hJ}\}$. This is a strong assumption about the accuracy of interpersonal comparisons, but it, or something like it, must be assumed to empirically estimate a model. If we measure compaction functions for each individual and assume the above hypothesis, we can then hypothesize a conditional probability model of the following form:

$$L_{ij} = p(a_j | v_{i1}, v_{i2}, \ldots, v_{iJ}) \tag{5}$$

and if we observe enough individuals we can estimate some parametric model -- an arbitrary parametric model.

Unfortunately, such a model is of little practical use because the model could not account for new products which are introduced to the choice set. Clearly, we must restrict the form of equation 5.

One obvious restriction is to require equation 5 to be symmetric in the sense of definition 3. I.e., if we switch two compaction values, $v_{ij} \leftrightarrow v_{ik}$, then we must switch the choice probabilities for those products, $L_{ij} \leftrightarrow L_{ik}$. The other choice probabilities must remain unaffected. If we restrict equation 5 in this way, the identity of the product no longer matters, only its compaction value. [Note equation 4 is of this form.] We can now introduce the general ranked probability model.

Let $rv_{i1} = \max_j \{v_{i1}, v_{i2}, \ldots, v_{iJ}\}$

$rv_{i2} = \text{second-largest}_j \{v_{i1}, v_{i2}, \ldots, v_{iJ}\}$

.
.
.

$rv_{iJ} = \min_j \{v_{i1}, v_{i2}, \ldots, v_{iJ}\}$

e_1 = the event that an individual chooses the product with the largest compaction value

e_2 = the event that an individual chooses the product with the second largest compaction value

.
.
.

e_J = the event that an individual chooses the product with the J^{th} largest compaction value.

If we can observe enough individuals, we can estimate the model:

$$pr_{ij} \triangleq p(e_j | rv_{i1}, rv_{i2}, \ldots, rv_{iJ}) \tag{6}$$

and it is simply a matter of bookkeeping to determine $\{L_{i1}, L_{i2}, \ldots, L_{iJ}\}$ from $\{pr_{i1}, pr_{i2}, \ldots, pr_{iJ}\}$. [Note that any symmetric probability model can be put in this form.] Let $\underline{rv}_i = \{rv_{i1}, rv_{i2}, \ldots, rv_{iJ}\}$ and apply Bayes Theorem to equation 6. This gives[+]

$$pr_{ij} = p(e_j) * \left[\frac{p(\underline{rv}_i | e_j)}{\sum_\ell p(e_\ell)\, p(\underline{rv}_i | e_\ell)}\right] \tag{7}$$

[+]Note that when a new product is introduced and the choice set expanded, we are concerned with $p(e_{J+1})$. But in empirical cases, this approaches zero.

But $p(e_j)$ is simply the unconditioned probability, p_j, that an individual will choose the product with the j^{th} largest compaction value. Thus, equation 7 is a generalization of the naive ranked probability model. Furthermore, because the range adjusting logit model is symmetric, equation 7 is also a generalization of it.

At present, the general model (equation 7) is not practical. What the general model does show is that it is feasible to construct a probability model that is both consistent with axioms 1 through 5 and uses their strengths.

5. EMPIRICAL EVIDENCE

The theoretical strength of the von Neumann-Morgenstern compaction functions would be of purely academic interest were it not feasible to directly measure compaction functions for individual consumers. For each of 76 consumers, Hauser and Urban [12] measured the following 10 parameter compaction functions with 8 indifference questions, 5 of which were lotteries.

$$c(x_{j1}, x_{j2}, x_{j3}, x_{j4}; \underline{\lambda}_i) = \sum_{k=1}^{4} \lambda_k u_k(x_{jk}) + \sum_{\ell>k} \sum_{k} \Lambda \lambda_k \lambda_\ell u_k(x_{jk}) u_\ell(x_{j\ell})$$

$$+ \sum_{m>\ell} \sum_{\ell>k} \sum_{k} \Lambda^2 \lambda_k \lambda_\ell \lambda_m u_k(x_{jk}) u_\ell(x_{j\ell}) u_m(x_{jm})$$

$$+ \Lambda^3 \lambda_1 \lambda_2 \lambda_3 \lambda_4 u_1(x_{j1}) u_2(x_{j2}) u_3(x_{j3}) u_4(x_{j4}) + \lambda_0$$

where

$$u_k(x_{jk}) = a_k - b_k e^{-r_k x_{jk}}$$

The parameters $\{a_k, b_k : k=1 \text{ to } 4\}$, λ_0', and Λ were set by scale conventions and the managerially significant "preference" parameters, $\underline{\lambda}_i = \{\lambda_k, r_k : k=1,4\}$, were determined by 8 indifference questions. (The resulting λ_k measure relative importance of attributes, the r_k measure risk aversion, and Λ, which is determined by $\sum_k \lambda_k$, measures attribute interaction.) Those readers interested in the properties of this particular function are referred to Keeney [19]. Those readers interested in the measurement, the results, and the managerial implications of this application are referred to Hauser and Urban [11].

Empirically, this assessment gave reasonable predictions as is evidenced by Table 1, which compares actual vs. predicted market shares for the four health care plans. Table 1 also gives the share which was predicted by a logit model estimated in the same study. It is possible that the non-linear risk averse utility functions performed slightly better than the linear logit model, because they were sensitive to the perceived risk involved in switching from existing care to a new health plan.

	Existing Care	Harvard Community Health Plan	M.I.T. Health Plan	Massachusetts Health Foundation
actual share	.34	.11	.42	.13
predicted share (utility)	.30	.08	.42	.20
predicted share (logit)	.22	.23	.35	.20

Table 1: Actual vs. Predicted Market Shares

6. SUMMARY

This paper began with a definition of stochastic preference and five basic axioms about stochastic choice behavior. These axioms imply the existence and uniqueness of a "compaction" function, that is, a function which identifies how consumers evaluate products in terms of attributes and produces real numbers which are monotonic in choice probabilities.

The measurement of such functions is important for describing and predicting choice. This paper indicates the conditions under which such functions exist, how they can be measured if they exist, and how one might use such functions to estimate choice probabilities.

Hopefully this paper will lead to improved synergy between the theoretic rigor of von Neumann-Morgenstern utility theory and empirical experience of marketing research. This area of investigation is fertile in both theoretical and practical problems, and it deserves attention from both utility theorists and marketing researchers.

REFERENCES

1. Bass, F., "The Theory of Stochastic Preference and Brand Switching," Journal of Marketing Research, Vol. XI, February 74, pp. 1-20.

2. Butler, D.H. and B.F. Butler, Hendrodynamics: Fundamental Laws of Consumer Dynamics, The Hendry Corporation, August 1966.

3. Farquhar, P.H.,"A Fractional Hypercube Decomposition Theorem for Multiattribute Utility Functions," Operations Research, Vol. 23, No. 5, Sept.-Oct. 1975, pp. 941-967.

4. Farquhar,P.H., "A Survey of Multiattribute Utility Theory and Applications," Management Science (forthcoming, 1976).

5. Fishburn, P.C., Mathematics of Decision Theory, UNESCO, 1972.

6. Friedman, M. and L.J. Savage, "The Expected-Utility Hypothesis and the Measurability of Utility," Journal of Political Economy, Vol. 60, 1952, pp. 463-474.

7. Green, P. and M. Devita, "An Interaction Model of Consumer Utility," Journal of Consumer Research, Vol. 2, September 1975.

8. Green,P. and V.R. Rao, "Conjoint Measurement for Quantifying Judgmental Data," Journal of Marketing Research, Vol. 8, August 71.

9. Green,P. and Y. Wind, "New Way to Measure Consumer's Judgments," Harvard Business Review, July-August 1975.

10. Hauser, J.R., "Testing the Accuracy, Usefulness, and Significance of Probabilistic Choice Models: An Information Theoretic Approach," Working Paper, Northwestern University, Transportation Center, April 1976. (Submitted to Operations Research.)

11. Hauser,J.R. and G.L. Urban, "A Normative Methodology for Modeling Consumer Response to Innovation," (Operations Research, forthcoming 1977)

12. Hauser, J.R. and G.L. Urban, "Direct Assessment of Consumer Utility Functions: von Neumann-Morgenstern Theory Applied to Marketing," Working Paper, M.I.T. Sloan School, April 1976. (Submitted to Management Science, 1976)

13. Herniter, J., "An Entropy Model of Brand Purchase Behavior," Journal of Marketing Research, Vol. X, November 1973, p. 361-375.

14. Herstein, I.N. and J. Milner, "An Axiomatic Approach to Measurable Utility," Econometrica, Vol. 21, 1953, p. 291-297.

15. Jensen, N.E., "An Introduction to Bernoullian Utility Theory. I. Utility Functions," Swedish Journal of Economics, Vol. 69, 1967, p. 163-183.

16. Johnson, R.M., "Beyond Conjoint Measurement: A Method of Pairwise Tradeoff Analysis," Proceedings of the Association of Consumer Research, October 1975.

17. Johnson, R.M., "Tradeoff Analysis of Consumer Values," Journal of Marketing Research, Vol. II, May 1974, pp. 121-127.

18. Kalwani, M.U. and D.G. Morrison, "A Parsimonious Description of the Hendry System," Working Paper 830-76, M.I.T. Sloan School, January 1976.

19. Keeney, R.L., "Multiplicative Utility Functions," Operations Research, Vol. 22, No. 1, January 1974, pp. 22-33.

20. Keeney, R.L., "Utility Functions for Multiattributed Consequences," Management Science 18, 1972, pp. 276-287.

21. Keeney, R.L. and H. Raiffa, Decision Analysis with Multiple Conflicting Objectives, (New York: John Wiley & Sons, 1976).

22. Luce, R.D., Individual Choice Behavior, (New York: John Wiley & Sons, 1959).

23. Marschak, J., "Rational Behavior, Uncertain Prospects, and Measurable Utility," Econometrica, Vol. 18, 1950, pp.111-141.

24. McFadden, D., "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zaremblea, ed., Frontiers in Econometrics, (New York: Academic Press, 1970), pp. 105-142.

25. McFadden, D., "The Revealed Preferences of a Government Bureaucracy: Theory," the Bell Journal of Economics, Vol. 6, No. 2, Autumn 1975.

26. Raiffa, H., Decision Analysis: Introducting Lectures on Choices Under Uncertainty, (Reading, MA: Addison-Wesley, 1970).

27. Ryan, M.J. and E.H. Bonfield, "The Fishbein Extended Model and Consumer Behavior," Consumer Research, Vol. 2, No. 2, September 1975.

28. Sheth, J.N. and M.W. Talarzyk, "Perceived Instrumentality and Value Importance as Determinants of Attitudes," Journal of Marketing Research, Vol. 9, February 1972.

29. Silk, A. and G.L. Urban, "Pretest Market Evaluation of New Packaged Goods: A Model and Measurement Methodology," Working Paper, M.I.T. Sloan School, November 1975.

30. Torgenson, W.S., Methods of Scaling, (New York: John Wiley & Sons, 1958), pp. 105-112.

31. Tversky, A., "A General Theory of Polynomial Conjoint Measurement," Journal of Mathematical Psychology 4, 1967, pp. 1-20.

32. von Neumann, J. and O. Morgenstern, The Theory of Games and Economic Behavior, 2nd. ed., (Princeton, N.J.: Princeton University Press, 1947).

33. Wilkie, W.L. and E.A. Pessemier, "Issues in Marketing's Use of Multi-Attribute Attitude Models," Journal of Marketing Research, Vol. 10, (November 1973), pp. 428-441.