# 3

# PROBABILISTIC MODELS OF CONSUMER CHOICE BEHAVIOR

Robert J. Meyer
*University of Pennsylvania*

Barbara E. Kahn*
*University of Pennsylvania*

In this chapter, we review recent developments in the literature on models of individual discrete choice. These models seek to represent the process that individuals use to integrate information about a set of alternatives when making a choice of a single preferred option. We focus on the major theoretical model forms and discuss some of the most important generalizations. These generalizations include the effects on choice of similarity and dominance among items in a choice set and temporal dependencies among choices. We conclude by suggesting an agenda for future work in the area.

## INTRODUCTION

Much of the modern research on consumer decision making is based on a simple conjecture: Human cognitive operations, no matter how complex, can be represented by mathematical analogs. While the validity of this conjecture is certainly not foregone, its implications have attracted the interest of researchers across numerous disciplines. Specifically, if correct, the conjecture implies the possibility of building models that would allow one to forecast choices consumers would make from a set of options, and perhaps, to alter the outcome of those choices.

Over the past twenty years the possibility of such "consumer engineering" has served to spawn a large literature of mathematical models of consumer judgment and choice. Since 1980, for example, there have been over 200 articles published on the subject in the

*This chapter was written while the authors were at U.C.L.A. We thank Richard Batsell, Lee Cooper, Sunil Gupta, Donald Lehmann, Jordan Louviere, William Moore, and David Schmittlein for their comments and suggestions on an earlier draft of the chapter.

literature of marketing,[1] and this reflects only a small fraction of the total literature on the subject that has appeared across disciplines. Indeed, the work has proliferated to such a degree that the field of quantitative models of behavior is probably best thought of as a discipline in its own right.

In this chapter we provide a review of recent developments in one aspect of this discipline, models of individual discrete choice. These are models that seek to represent the process by which individuals integrate information about a set of alternatives when making a choice of a single preferred option. Within the study of consumer behavior, these models are usually used to describe the choice of a brand from a competitive set of brands.

We should stress that our survey is by no means meant to be exhaustive. Because a complete review of even the most recent developments in individual choice modeling is beyond the scope of a single review article, our focus will be limited. We provide a discussion of the major theoretical model forms that are used to represent consumer choices from sets of options, and we look at the major generalizations of these model forms that have appeared in the recent literature. The specific generalizations we focus on are choice models which recognize the effects of the similarity and dominance among items in a choice set on evaluation processes, and models which recognize temporal dependencies in choice. Excluded from our review will be discussion of statistical issues in the estimation of these models and their application in managerial settings. For a review of these topics we refer the reader to recent texts by Ben-Akiva and Lermann (1985), Cooper and Nakanishi (1988), Hensher and Johnson (1981), and Train (1986), as well as to review articles by Amemiya (1981), Corstjens and Gautschi (1983) and Wrigley, Longley, and Dunn (1984; 1988). In addition, we also exclude from discussion work in a number of areas that have strong traditional ties to the literature of probabilistic choice models; in particular, the literatures of multiattribute attitude models (e.g., Green and Wind 1973; Louviere 1988), models of purchase quantity and timing (e.g., Ehrenberg 1959; Massy, Montgomery, and Morrison 1970; Neslin, Henderson, and Quelch 1985), and models of group choice (e.g., Corfman and Lehmann 1987; Davis 1976; Webster and Wind 1972).

We begin by offering a general taxonomy of the literature on choice models, which will be useful in both organizing our subsequent discussion and in identifying major future research directions. We will then review the dominant theoretical paradigm in the study of individual discrete choice, economic random utility theory. We then survey attempts to generalize this model in two directions: models that recognize a dependency of preferences on the external appearance of the set of choice options, and models that recognize a dependency among a sequence of choice (dynamic models). We conclude with a discussion of research challenges facing the area.

## The Literature on Individual Choice Models

The contemporary literature of consumer choice models does not have a simple genealogy. While several taxonomies have been suggested (e.g., Corstjens and Gautschi 1983; McFadden 1986), they belie the complex history of many of the key ideas in the area, and the disagreements which often exist about the relationship among these concepts. Much of this lack of clarity stems from the multidisciplinary nature of research in the area; modeling ideas are often developed in parallel by several authors in several disciplines. In addition, ideas have sometimes been proposed and then left idle in one discipline, only to be rediscovered and developed further in another discipline. As a result, models that are seen as the general form of one class of representations by one author are often seen as the special case

[1] This compilation spans entries in the *Journal of Consumer Research*, the *Journal of Marketing*, the *Journal of Marketing Research*, *Marketing Science*, conference proceedings in Marketing, and original book compilations.

of another class by another author.[2] The field is thus best viewed as a collage of works rather than an organized system; it reflects the independent efforts of researchers in different disciplines, unified by a common interest in how individuals make choices from sets of options.

The central tenet of our view of the field is that there exists two intellectual traditions of models (see Figure 3.1), each reflecting different views of the ability of humans to process information. At one extreme is the view that individuals make choices by considering all relevant information available to the decision maker at the time of choice, and that individuals choose that option which maximizes some utility function defined across this information set. Central to the basic forms of such models is an assumption that preferences for items can be defined independently from the set of options under consideration, and that any errors which exist in the measurement of preferences are also independent of the consideration set. Termed *simply scalable* choice models, they are illustrated by the choice models of Luce (1959), the Multinominal Logit (McFadden 1981), and the Independent Probit (Domenich and McFadden 1975).

At the other extreme is the view that individuals are inherently limited in their ability to process information, and thus they make choices through simplified heuristics, which do not use all the information available at the time of choice. Central to these models is the assumption that preferences are inherently context dependent. Specifically, these models assume that the likelihood an option will be chosen is a function of both the attractiveness of its features and the extent to which these features are shared by other options under consideration. Work in the area is best associated with the attribute elimination models of Restle (1959), Tversky (1972), and Tversky and Sattath (1979). These models, although rich in explanatory power, are often difficult to estimate empirically.

Perhaps paradoxically, although each para-

[2] To illustrate, as an economist, McFadden (1981) has described the multinominal logit model as following in a line of microeconomic theories of choice. In contrast, Yellott (1978) has described the same model as a descendent of psychological theories of comparative judgment developed in the late 1920s.

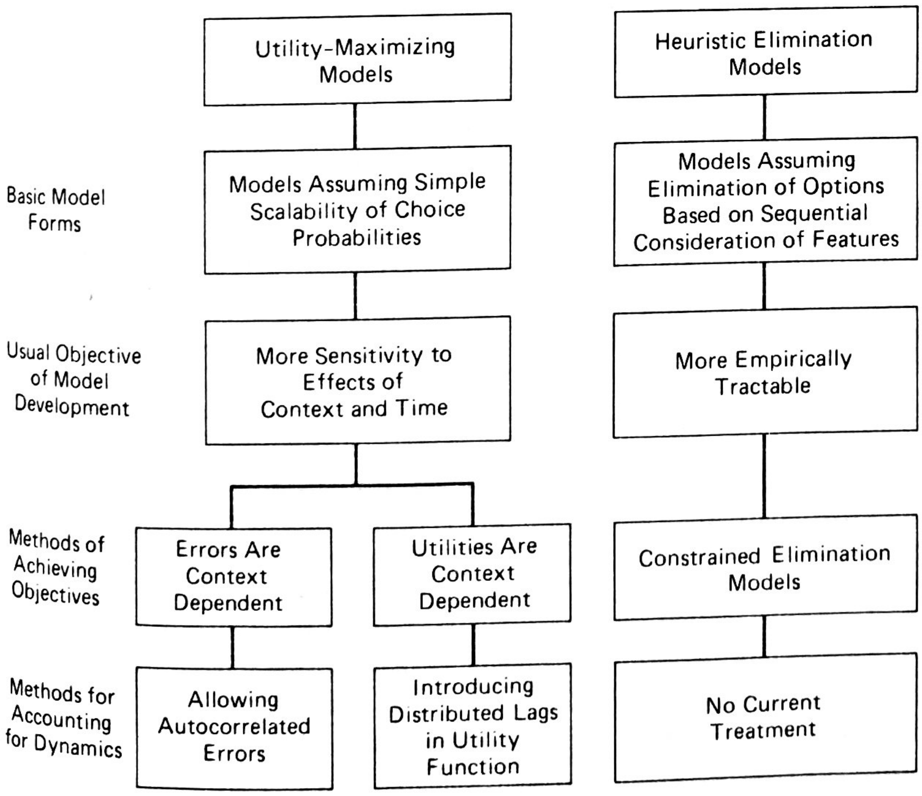


**FIGURE 3.1** A Taxonomy of Theoretical Choice Model Forms

digm starts with quite different views of decision making, research within each area has been characterized by movement to a common ground; specifically, research in utility-maximization models centers on attempts to make them sensitive to the effects of choice context, while research in heuristic elimination models centers on attempts to make them empirically tractable.

Within the paradigm of utility-maximization models, two general approaches to relax assumptions of context independence have emerged. One stream of work retains the assumption that utilities can be defined independently of context, but relaxes the assumption of independence among errors in the measurement of utilities. These representations are best illustrated by McFadden's (1979) Generalized Extreme Value model and Hausman and Wise's (1978) Covariance Probit model. The other major approach relaxes the assumption that utilities can be defined independently of context, but retains the assumption that errors are independent. Such representations are illustrated by the differential cross-effects choice models of Batsell and Polking (1985) and Cooper and Nakanishi (1988). The two streams have also been extended to the treatment of temporal context; dynamics have been handled both through the addition of lagged effects in the utilities of options (while assuming that errors are not time dependent), and by allowing time dependencies in measurement errors.

Research on heuristic elimination models has tended to search for ways to approximate a general sequential choice model, Elimination-by-Aspects (EBA) (proposed by Tversky (1972)), by simpler forms which are more easily estimable. In contrast, only limited work has been directed to adding greater generality to the EBA models, such as by allowing them to account for learning.

In this chapter we will review in greater detail both the structure of the basic model forms within each of these paradigms and their recent refinements. Because the utility-maximization models form the dominant paradigm in modern choice analysis, they will form the central focal point of the chapter. Our discussion of heuristic-based models will appear as it is usually positioned in the literature, as a discussion of an alternative approach to modeling the effects of item similarity on choice probabilities (e.g., Corstjens and Gautschi 1983).

## THE ECONOMIC THEORY OF INDIVIDUAL CHOICE

### The Consumer as a Random Utility Maximizer

Most models of consumer choice are based on a simple behavioral postulate: When faced with a set of options, consumers choose that option that is thought to deliver the highest level of perceived gratification or utility. As noted by McFadden (1981; 1986), this postulate is essentially tautological; if an individual chooses an option, it must be that which was anticipated to be the best among the array of alternatives, at least according to some criterion.

Although there exist a number of formal theories of how consumers seek to maximize utility within a given choice setting, by far the most often cited is that first attributed to Thurstone (1927) in psychology, and later generalized in economics by McFadden (1974). When faced with a choice, a consumer is presumed to view each option as a bundle of attributes. The consumer is then thought to form an overall evaluation of each option by combining perceptions of the option's attributes through a cognitive "integration rule" or utility function. The chosen alternative is then that option for which this overall evaluation is the highest.

Due to temporal variation in tastes and unmeasured influences on choice, however, it is assumed that these utilities are not fully observable by the analyst. Rather, the utility of each option is considered a random variable, defined by a distribution of possible values. Models that yield estimates of the probability that each option will be chosen based on assumptions about the form of these distributions are termed *random utility models*.

This formal modeling process can be described as follows: Let $X_i$ be a vector of the measured attributes of choice option $i$, and let $V_i(X_i)$ be a preference mapping which links the vector $X_i$ to a summary indicator of the overall utility or value of $i$. Although $V$ may be any multiattribute function, in most applications it is assumed to be a linear combination of the set of observed attributes of $i$:

$$V_i\,(X_i) = b_i + \sum_{k=1}^{m} b_k x_{ik}\,, \tag{3.1}$$

where $x_{ik}$ is the observed value of option $i$ on attribute $k$ (such as its price), $k = 1, \ldots, m$, and $b_i$ and $b_k$, $k = 1, \ldots, m$ are scaling parameters. Expression (3.1) is the most well-known representation of measurable utility used in random utility models (cf. Hensher and Johnson 1981). It presumes an intercept $b_i$ which is unique to each alternative, and a set of generic attribute effects $b_k$, $k = 1, \ldots, m$. The alternative-specific intercept is designed to capture the systematic constant component in the attractiveness of option $i$ not captured by the attribute vector $X_i$. As such, it plays a role similar to a "subject-specific" blocking effect in a repeated-measures analysis of variance (Hensher and Johnson 1981).

We might note that there is no a priori reason for constraining the attribute effect coefficients to be equal across alternatives except for reasons of parsimony. Cooper (1988), Cooper and Nakanishi (1988), and Train (1986), for example, note that in many contexts the effect of variations in attribute values will differ across alternatives, implying alternative-specific attribute effect coefficients. Similarly, in many instances overall impressions will not be an additive combination of attribute values (e.g., Anderson 1981, Louviere 1988). In this case expression (3.1) might be expanded to included higher-order cross products among attributes (e.g., Lynch 1985, Louviere and Meyer 1980).

We presume that $V(X_i)$ is an imperfect indicator of the true utility an individual assigns to option $i$ at the time of choice, $U_i$. Specifically, $U_i$ is assumed to be linked to $V(X_i)$ through an independent disturbance $\epsilon_i$, such that

$$U_i = V_i(X_i) + \epsilon_i\,, \tag{3.2}$$

where $\epsilon_i$ reflects the observed tastes of the individual with respect to $i$. And $V_i(X_i)$ is termed the *strict utility* of $i$, while $\epsilon_i$ is the *random utility* (McFadden 1981).

Given a choice set $A$, the likelihood that a decision maker will choose option $i$ from this set [$Pr(i/A)$] is thus the likelihood that the latent variable $U_i$ is the highest at the time of choice. Formally,

$$Pr(i|A) = Pr(U_i > U_j)\,\forall j \in A, j \neq i\,.$$

This expression, in turn may be rewritten in terms of (3.1) as:

$$Pr(i|A) = Pr([V_i(X_i) + \epsilon_i] > [V_j(X_j) + \epsilon_j]),\ \forall j \in A, j \neq i\,.$$

or, more conveniently,

$$Pr(i|A) = Pr([\epsilon_j < V_i(X_i) - V_j(X_j) + \epsilon_i])\ \forall j \in A,\ j \neq i\,. \tag{3.3}$$

The probability that $i$ is chosen is then obtained by making an assumption about the form of the distribution of the random variables $\epsilon_i$ and $\epsilon_j$, and integrating (3.3) over a continuum of possible values of $\epsilon_i$. Specifically,

$$Pr(i|A) = \int_{-\infty}^{+\infty} Pr(\epsilon_i = \epsilon)(Pr(\epsilon_j < V_i(X_i) - V_j(X_j) + \epsilon)d\epsilon)\ \forall j \in A, j \neq i\,, \tag{3.4}$$

where $\epsilon$ is a constant of integration (cf. Hensher and Johnson 1981).

The most common assumption made about $\epsilon_i$ and $\epsilon_j$ is that they have independently and identically distributed Type I Extreme Value distributions; that is,

$$Pr(\epsilon_i \leq \epsilon) = e^{-\epsilon^{-\epsilon}}.$$

Under this assumption, integration of (3.4) yields the closed-form probability model,

$$Pr(i|A) = \frac{e^{V(X_i)}}{\Sigma_{j\epsilon A} e^{V(X_j)}}, \quad (3.5)$$

which is the well-known multinomial logit model of choice (Hensher and Johnson 1981; McFadden 1981; Schmittlein 1986; Yellott 1977).

Of course, other assumptions can (and have) been made about the distribution of $\epsilon_i$ and $\epsilon_j$, however none yield probability expressions which are as computationally tractable as the logit. In his early work on comparative judgment, Thurstone (1927), for example, posited a normal distribution for the random components, an assumption which yields the binary probit model (Ben-Akiva and Lerman 1985). The primary limitation of the assumption of normality, however, is that it does not offer a closed-form solution for the probability of choice, and extensions to cases with more than two alternatives have proved computationally difficult (an issue we will elaborate on later).

Because of its simplicity, the multinomial logit has emerged as the most widely used form of individual choice model. Applications have been diverse, including the study of consumer's choices of packaged goods (e.g., Cooper 1988; Currim, Meyer, and Le 1988; Guadagni and Little 1983; Gupta 1988; Malhotra 1984), travel mode choice and automobile ownership (Ben-Akiva and Lerman 1985; Gensch and Recker 1978; Train 1986), and college choice (Punj and Staelin 1979), to cite just a few examples. In all of these instances the model is shown to provide a good account of the relationship between the attributes of sets of alternatives and the choices made from these sets.

Adding to the appeal of the mutinominal logit is that it is linked to a number of other well-known models for discrete data anlysis, such as the models suggested by Bradley and Terry (1957) in statistics and Luce (1959) in psychology for the scaling of paired-comparison data, and the models used for the analysis of market share data in marketing (e.g., Nakanishi and Cooper 1974; Ghosh, Neslin, and Shoemaker 1984). Because Luce's model is perhaps the best known of these variations, choice models which can be written in the form of expression (3.5) are often generically referred to as "Luce Models" (e.g., Yellott 1977).

### The Random Utility Models as a Representation of Population Choice Behavior

Although random utility models such as the multinomial logit are most conveniently presented as representations of an individual's choice from a set of options, in practice they are more often used as models of aggregate behavior. In principle, this application is straightforward: When viewed as a model of population behavior, the strict utility function $V(X_i)$ of the multinomial logit is an indicator of the tastes of a "representative" member of the population, and the random utility $\epsilon_i$ reflects random individual variation about this mean (Hensher 1984; McFadden 1981).

In most settings, however, such application will be complicated by the presence of systematic variations in tastes across the population; for example, older consumers may have different loyalties than younger consumers, and consumers with higher incomes may be less price elastic. There exist two standard approaches for handling such variation: The analyst can either expand the strict utility function to include a battery of characteristics for each individual (e.g., Fisher and Nagin 1981), or models can be estimated for differing consumer segments, each homogeneous with respect to a set of consumer characteristics (e.g., Currim 1981; Gensch 1985). The assumption in both of these cases is that while tastes may vary systematically across consumers with differing characteristics, within groups heterogeneity is a double exponentially distributed random error.

While these approaches to dealing with taste variation will usually be adequate in most settings, they are less tenable in cases where the analyst has limited access to useful socio-demographic/economic measures, or residual taste variation is not well described by a symmetric probability density function. To address this problem several authors have explored the use of compound probability models of choice: individual choice probabilities are characterized by a simple choice model (such as the multinomial logit), and variation in probabilities is characterized by a distribution that allows for a flexible representation of taste variation, such as a beta distribution (e.g., Dunn and Wrigley 1985; 1987). A good pedagogic introduction to this approach to modeling aggregate choice behavior is provided by Massy, Montgomery, and Morrison (1970), with more recent applications in the context of random utility models being provided by Wrigley, Longley, and Dunn (1988), Jones and Zufryden (1980), and Steckel and Vanhonacker (1988).

## MODELS RECOGNIZING THE EFFECTS OF THE ITEM SIMILARITY AND DOMINANCE ON CHOICE PROBABILITIES

Despite its frequent applications as a model of both individual and aggregate behavior, the multinomial logit model of choice has been criticized for offering an overly restrictive view of the process that underlies choice. Although proponents point out that the model is often an extremely robust predictor of choice even when its underlying assumptions—such as the independence of the random components of utility—are violated (e.g., Louviere and Woodworth 1983), critics have claimed that the model does not provide an accurate description of the *process* which underlies choice, except in highly limited contexts.

At the heart of most criticism is the model's assumption that the process by which options are evaluated and chosen is independent of the external appearance of the option set. On one hand, this assumption of independence is perhaps the simple logit's greatest asset; if the assumption holds, one can, in principle, estimate a model in one context and use it to predict the choices individuals will make in any other. On the other hand, it is an assumption that many have argued will rarely, if ever, be empirically verified. Specifically, there is extensive behavioral evidence that choice processes tend to be constructed only *after* the composition of a choice set is first considered by consumers (e.g., Bettman 1987; Payne 1983).

The limitations of the context independence assumption are perhaps most readily seen by considering the model's predictions about how choice probabilities will be redistributed given the introduction of a new option to the choice set that is highly similar to one of the existing options. One of the key properties of models of the form of expression (3.5) is that the ratio of the choice probabilities of any two options is invariant under changes in the composition of the choice set in which both appear. This result is called the "constant ratio rule," or the "independence of irrelevant alternatives" (IIA) property (McFadden 1981). Its central implication is that if a new option is added to a choice set, the shares of existing options will always decrease in direct proportion to the size of their original shares.

Debreu (1960) is usually credited with offering the best known counter-example to this prediction. Although there have been many variants of the original Debreu discussion, all share the following basic structure: An individual is initially faced with a choice between two dissimilar choice alternatives, such as between ordering lamb or fish entree from a menu. A new option is then made available which is nearly identical to one of the originals except for an unimportant difference, say a new lamb entree with slightly different seasoning. The intuitive effect of this new option would seem clear: Because the new option does little to resolve the basic dilemma of whether to order lamb or fish for dinner, if it has any effect at all, it will be to diminish the probability of ordering the original lamb entree, while leaving the probability of ordering the fish largely unchanged. Unfortunately, this intuition would not be reflected in the predictions of a simple logit model; the constant ratio rule dictates that the new option would "draw" from each option in proportion to their original (binary) probabilities of being chosen. Hence, if the individual was initially indifferent between the fish and lamb entrees, the new lamb entree would draw shares equally from both options.

Since the early 1970s a major thrust of research in mathematical choice modeling has been to develop representations of choice that provide more intuitively plausible accounts of how changes in the relative similarity of options in a choice set affect choice probabilities. These modeling efforts can generally be categorized as following one of three strategies for generalization:

1. Models which assume that choices are not made by a utility-maximization process, but rather by a feature elimination process;
2. Models which retain the assumption that strict utility functions can be defined independently of the particular set of options under consideration, but which allow dependencies among the measurement errors; and
3. Models which retain the assumption that measurement errors are independent among options, but which allow the strict utility function to be dependent upon the consideration set.

We will briefly review each of these strategies.

## Feature Elimination Models

Process-tracing studies of human choice offer a dramatically different portrait of the process by which decisions are made than that presumed by most random utility models. When individuals are asked to "think aloud" when making choices (e.g., Bettman 1971; Hayes-Roth and Hayes-Roth 1979), when eye movements are traced (e.g., Russo and Dosher 1983), and when patterns of information search are monitored (e.g., Bettman 1981; Johnson, Meyer, and Goshe 1988; Payne 1976), there is little evidence that individuals form independent assessments of the overall value of each in a set of options, and then select that with the highest value, even in highly simplistic settings (Russo and Dosher 1983). Rather, the process of choice seems better characterized by the application of a sequence of discrete elimination heuristics, in which only a limited subset of the total amount of information in a choice set is considered at any one time. In such processes the external appearance of the choice set is an integral part of choice; the criteria used for eliminating options and the ultimate likelihood that an option will be chosen appears driven by the perceived structure of the entire option set.

Restle (1959) was the first to propose a model characterizing choice as a process of sequential eliminations. More recent work in the area, however, has tended to be dominated by Tversky's (1972) *elimination-by-aspects* (EBA), a generalization of Restle's original model.

The psychological process which underlies the EBA model is straightforward. When faced with a choice, the individual is hypothesized to view each alternative in the set as a collection of measurable attributes or features. Each of these features are shared by at least one, but not all, alternatives in a choice set. The individual then selects one of these features for consideration with a probability proportional to its relative desirability among all features. Options which share this feature are retained for further consideration, while all others are eliminated. If there is more than one option in the surviving set, a second feature is selected by the same probabilistic process as before, with the tournament continuing until only one option remains.

For example, consider the choice of an entree from a menu. If "spiciness" were the most important single aspect in the set, then whether an entree was spicy or mild would have the highest probability of being the first criterion used for evaluation. Assuming "spiciness" was the first aspect, then all entrees which were not spicy would be eliminated. Given the remaining alternatives, another aspect, say "lamb" would be selected and all alternatives which were not made of lamb would be eliminated. This would continue until one entree was left.

The EBA model can be formally stated as follows. Let $T = \{x, y, z, \ldots\}$ be the total set of options faced by the decision maker, and let $A$, $B$, $C$ denote the nonempty subsets of $T$. Let $P(x|A_\alpha)$ be the probability of choosing alternative $x$ from that subset of options offered in $A$ which share aspect or feature $\alpha$. Next, let $A'$ be a set of aspects that belong to at least one alternative in A (note A' is the set of aspects and $A_\alpha$ is a set of alternatives). Finally let $x'$ be a set of aspects of $x$, $x' = \{\alpha, \beta, \ldots\}$ and $u(\alpha)$ be a scale value defining the desirability of feature $\alpha$. EBA is defined in terms of the recursive formula,

$$P(x|A) = \frac{\Sigma_{\alpha\epsilon x'}\ u(\alpha)P(x|A\alpha)}{\Sigma_{\beta\epsilon A'}u(\beta)}, \tag{3.6}$$

where $u(\beta)$ is a scale value defining the desirability of feature $\beta$ which may or may not be possessed by option $x$. Equation (3.6) may also be rewritten:

$$\frac{\Sigma_{\alpha\epsilon x'}\ u(\alpha)P(x|A\alpha)}{\Sigma_{y\epsilon A}(\Sigma_{\alpha\epsilon y'}\ u(\alpha)P(y|A\alpha))}. \tag{3.7}$$

Expression (3.7) illustrates one of EBA's central properties: It collapses to a model of the simple logit form (expression 3.5) when either each option is defined in terms of unique features or when the conditional probabilities of choice given a feature are equal across options. In such cases (3.7) simplifies to:

$$P(x|A) = \frac{V(x)}{\Sigma_{y\epsilon A}\ V(y)}, \tag{3.8}$$

where

$$V(x) = \sum_{\alpha\epsilon x'} u(\alpha). \tag{3.9}$$

A key property of the EBA model is that the "features" used to eliminate options are not necessarily measured by the analyst prior to choice. Rather, the objective is to conduct an *internal* analysis of preferences, where the analyst's objective is to derive a set of psychological scales for options which best reproduce an observed set of choices or judgments (cf. Carroll 1972; Cooper and Nakanishi 1983). This contrasts with the *external* analysis provided by the logit model, where the observed set of choices is linked to a set of externally derived scales, such as measures of the prices of options and their perceived attractiveness.

The central appeal of the EBA model is that it provides a highly flexible theoretical scaling model for choice data in contexts where options can be assumed to be "screened" through a sequential consideration of features. The major drawback is difficulty of estimation: Given a set of $n$ options, the EBA model requires the estimation of $2^n - 3$ scaling parameters, which, in turn, require estimates of the conditional probabilities of choice within all possible $(2^n - 1)$ nondegenerate subsets of alternatives in a choice set of interest (Tversky 1972). Because this data requirement rapidly becomes excessive with increases in set size, published applications have been limited to small illustrative examples (e.g., Tversky 1972), and we are aware of no attempts to apply the model to "normally available" econometric data.

In an attempt to overcome this estimation problem, Tversky and Sattath (1979) suggested that computational tractability could be improved if the analyst imposed a prior hierarchical structure. The logic behind the approach is that number of scaling parameters required to reproduce a set of choice probabilities is driven by the complexity one assumes for the underlying choice process. For example, if expression (3.8) can be assumed to hold for a set of choice data (i.e., a simple Luce model holds), only $n$ (rather than $2^n - 3$) parameters would need to be estimated; that is, one scale parameter for each option. The intermediate case is where choice is assumed to follow a hierarchical agenda, in which case $2n - 3$ parameters would be sufficient (Tversky and Sattath 1979). Based on this notion, Tversky and Sattath developed a modeling procedure and estimation algorithm called PRETREE, which allows the calibration of a hierarchical-elimination choice model mirroring EBA, but with fewer parameters and correspondingly less stringent data demands.

There have been a number of published applications of PRETREE, all reporting reasonable success in recovering the parameters of a feature-based hierarchical choice process from paired-comparison data (e.g., Kahn, Moore, and Glazer 1987; Moore, Lehmann, and Pessemier 1986). These investigations, however, also suggest some caveats (e.g., Lehmann and Moore 1985). Preeminent among these is the difficulty of identifying the "correct" decision tree in a given setting; the fit and form of the final choice tree is often sensitive to how it is initially "seeded," hence repeated applications of the estimation algorithm are required in most settings. A central consequence of this limitation is that it becomes difficult to apply the procedure at the individual level, which inhibits investigations into heterogeneity in decision rules within a sample.

***Recent Developments in Hierarchical Elimination Models.*** Given the strong intuitive appeal of viewing choice as a process of sequential elimination, a number of authors have sought alternatives to the EBA and PRETREE models in an attempt to overcome their discussed drawbacks. Efforts have focused in three general areas:

1. those which have explored alternative approaches to the estimation of feature-elimination models;
2. those which have attempted to extend feature-elimination models to the case of continuous attributes; and
3. those which have sought means for the estimation of models at a disaggregate level without a need to prespecify a decision sequence.

Illustrative of work in this first area are the efforts of Lehmann and Moore (1985), who explore the possibility of using a hierarchy of logit models as a tool for developing feature models similar to those yielded by PRETREE (we will elaborate on this procedure shortly). Likewise, DeSarbo and De Soete (1984), DeSarbo, De Soete, Carroll, and Ramaswamy (1987), and Rao and Sabavala (1981) have explored the possibility of inferring hierarchical feature models through the use of hierarchical clustering algorithms—a possibility which was also explored by Tversky and Sattath (1979).

Manrai and Sinha (1989) and Rotondo (1986) have considered the problem of extending feature-elimination models to cases where options are described by continuous attribute dimensions rather than by discrete features. In Manrai and Sinha's (1989) approach, termed *Elimination-by-Cutoffs* (EBC), options are initially rank-ordered with respect to their scores on each of a number of prescribed dimensions. Options are chosen by sequentially eliminating alternatives that share a common maximum preference distance from another option in the set; for example, the first elimination stage might be to eliminate all alternatives which are no more than *k* rating points better than the worst option in the set on a dimension. The limitation of the current version of EBC, however, is that attribute scales have to be defined by the user prior to analysis, and that it does not allow recognition of similarity effects accruing to groups of options possessing unique discrete features, such as the example of choices among dinner entrees.

The problem of estimating hierarchical models of choice that recognize individual differences in elimination strategies has proven to be more elusive. Gensch and Svestka (1984), for example, provide an algorithm which provides a lexicographic screening of options at the individual level; however, it requires as input individuals' rankings of the relative importance of product attributes, as well as their ratings of the attractiveness of each option on these attributes. More recently, Currim, Meyer, and Le (1988) and Greene and Smith (1988) have illustrated how tree-structured regression procedures such as AID and CART (Breiman, Olshen, and Stone 1983) can be used to derive hierarchical models of binary choice at an individual level from normally available consumer panel data, without requiring assumptions about the form of the decision hierarchy. Unfortunately, because the decision trees yielded by such procedures are derived from an analysis of independent choice rather than proximities (such as in PRETREE or hierarchical clustering), they may not be directly interpreted as implying competitive groupings within an option set—the normative role of partitions in PRETREE models. Nevertheless, if the trees turn out to be a reasonably close match with those yielded by more complex methods, they would provide an efficient solution to the problem of assessing the pattern of heterogeneity in choice structures that exists in a sample.

## Random Utility Models Relaxing the Assumption of Error Independence

Within the framework of random utility models the task of developing models that do not possess the IIA assumption would seem straightforward: because IIA follows from an assumption that the random comments of utility are independent across options, non-IIA models can, in principle, be derived simply by allowing a more flexible distribution of errors. The behavioral rationale for similarity effects that underlies such a modeling approach would be quite different than that presumed by feature-elimination models. The consumer would be thought to be attempting to maximize utility across a set of options, and similarity would be presumed to be driven by unobserved influences on utility that are correlated across subsets of options. For example, when choosing among entrees on a menu, unmeasured factors which detract from or contribute to the attractiveness of, say, one fish entree would be presumed to affect all fish entrees. The consequences is a dependence of the choice probabilities for any one option on its similarity with other options in the set—the same context effects the EBA model captures by assuming choices are made through a sequential elimination process.[3]

There are two major classes of random utility models which are derived from an assumption of a potential dependence among the unobserved components of utility: Generalized Extreme Value Models and Generalized Probit Models. We will briefly review each in turn.

***The Generalized Extreme Value Model.*** The Generalized Extreme Value (GEV) Model (McFadden 1979) is a family of logit-like models which follows from the assumption that the components of utility in the general random utility model given in expression (3.4) have a multivariate extreme value distribution. The "standard" GEV (Amemiya 1981) model can be stated as follows: Assume that at the time of choice the decision maker initially partitions the total set of available options into $n$ subsets $A_p$, $p = 1, \ldots, n$, of similar alternatives such that the IIA property of the basic logit holds within these sets but not necessarily between these sets. The GEV model asserts that the probability that an individual will choose option $i$ which is an element of subgroup $q$, $P_{iq}$ is given by:

$$P_{iq} = \frac{e^{V(X_{iq})/\lambda q}\left[\sum_{j\epsilon q} e^{V(X_{jq})/\lambda q}\right]^{(\lambda q - 1)}}{\sum_{p=1}^{n}\left[\sum_{j\epsilon p} e^{V(X_{jp})/\lambda p}\right]^{(\lambda_p - 1)}}, \quad (3.10)$$

where $V(X^{iq})$ is the strict utility of option $i$ in subset $q$ as in expression (3.2), and $\lambda_p$ is an inverse measure of the correlation among the unobserved components of utility within subset $p$. An important property of (3.10) is that in the case where unobserved components are all uncorrelated ($\lambda_q = 1$ for all $p$), (3.10) reverts to a simple (independent) multinomial logit.

Although seemingly complex, the probability given by Expression (10.3) turns out to have a rather simple intuition, and it has a straightforward computational form. Specifically, the expression implies a decision tree in which the probability that a given option is chosen is the product of a series of conditional probabilities that the sets to which it belongs are chosen and then that the option is chosen from those sets.

Train (1986) offers the following intuition for this result: Imagine that we decompose the strict utility associated with any option ($V(X_{iq})$)

---

[3] Several authors have pointed out that when a random utility model is used as a model of aggregate choice, violations of IIA can also be viewed as a problem of unobserved heterogeneity in preferences in a population. Specifically, one could presume there exist subsets of consumers who are attracted to similar types of options, such that when one consumer becomes attracted to an option of a given type, the aggregate attractiveness of all options of that type increases.

into two components: those attributes which are shared by all options in subset $q$ and those which are unique to each option. For example, if a subset under study is the set of all fish entrees on a menu, the first component would be those attributes shared by fish entrees (e.g., seafood), and the second component would be made up of attributes which discriminate among the entrees (e.g., fish varieties and seasonings). Formally,

$$V(X_i) = W(X_{.q}) + \theta_q Y(X_{iq}), \qquad (3.11)$$

where $W(X_{.q})$ is a utility mapping defined over the attributes common to all options in set $q$, $Y(X_{iq})$ is a utility mapping defined over the attributes which vary across options within set $q$, and $\theta_q$ is a scaling constant.

This decomposition allows us to separate the probability in (3.6) into the product of the marginal probability that subset $q$ (for example, seafood) is chosen from the total set of $p = 1, \ldots, n$ subsets (for example, all food groups) times the conditional probability that option $i$ (for example, a particular seafood entree) is chosen given $q$. Specifically,

$$P_{iq} = P(A_q) \cdot P(i|A_q),$$

where $P(A_q)$ and $P(i|A_q)$ are simple multinomial logit models of the probability of choosing subset $q$ from all subsets and choosing option $i$ given a choice of subset $q$, respectively. In particular,

$$Pr(A_q) = \frac{e^{W(X_{.q}) + \theta_q I_q}}{\Sigma_p e^{W(X_{.p}) + \theta_p I_p}},$$

$$Pr(i|A_q) = \frac{e^{Y(X_{iq})}}{\Sigma_{i\epsilon q} e^{Y(X_{iq})}}, \qquad (3.12)$$

where

$$I_q = \sum_{i\epsilon q} e^{Y(X_i)}.$$

Expression (3.12) is the probability that subset $q$ will be chosen from all subsets. The term $I_q$, which appears in (3.12) is the sum of all the strict utilities of the options in subset $q$ and is termed the "inclusive value" of the subset. It is the inclusive value which causes the GEV model to be referred to as a "nested logit"; choice is modeled as a sequence of logit models, in which the utilities derived from a model of conditional choice are nested within the model of marginal choice.

The two-stage GEV or nested logit model described here easily generalizes to the case of multiple stages of subset selection, such as an individual first partitioning the set of menu items into "spicy" versus "nonspicy" items, then meat types within each seasoning group, and so forth. In the multiple-stage situation, the modeling process is a straightforward one: The analyst first estimates a model for the choice of an item within a given subset (for example, the entree from the set of spicy meat dishes). The sum of the calibrated strict utilities for all of the items in that subset then serves as an explanatory variable in a higher-level model of choice among subsets, with such a model being augmented by attributes which vary across subsets [the vector $W(X_{.q})$]

The GEV model captures one of the simplest ways in which the appearance of a choice set affects choice probabilities, and provides an intuitively plausible means of accounting for the Debreu counter-example to the simple logit discussed previously. Because when GEV choice options are partitioned into discrete groups with differential degrees of substitutability, new alternatives to a set will not uniformly affect existing options. Specifically, the effect of a new option depends upon the number of existing options which exist in the subgroup to which it is assigned; the less "new" it brings to the overall choice set, the smaller its incremental share.

We should stress that while the GEV and EBA model discussed earlier both characterize choice as a hierarchical process, the behavioral process which underlies them is quite different. Specifically, the GEV model characterizes a full-information, "bottom-up," decision process, in which choices among subsets of options at one stage explicitly consider the utilities of all alternatives which might be available at later stages. The EBA and related models, in contrast, characterize a limited-information, "top down," decision process, in which elimination decisions at one stage do *not* consider the value of options available in later stages. Despite this difference, GEV and the EBA model might nevertheless be viewed as parallel developments: In the same way that the EBA model generalizes the psychology-based Luce model to the case of sequential considerations of product features, GEV generalizes the economic-based multinomial logit model to the sequential consideration of subsets of options.

Because of its strong intuitive appeal and relative ease of estimation, the GEV model has been applied to contexts where analysts have a priori reasons to suspect that the IIA assumption of the basic logit may be empirically untenable (e.g., Dubin 1986; Guadagni 1983). For example, Onaka and Clark (1984) and McFadden (1979) report applications of the model in studies of residential choice. In these models the individual is first postulated to choose a neighborhood, then residence within a neighborhood. Likewise, Brown (as presented in Hensher and Johnson 1981) reports an application to the study of choice of vacations; here, the individual is modeled as first deciding upon the duration of the vacation, then selecting a destination conditional upon duration.

Despite these successes, the GEV model has some limitations that have caused it to be viewed as only a partial solution to the problem of modeling choice where the IIA assumption is likely to be violated. The three most salient limitations are the following:

1. the need to prespecify how choice options are partitioned prior to choice and, in the case of multiple-stage GEV model, the order in which subset decisions are made;
2. the continued assumption that within subsets, the IIA assumption holds; and
3. the assumption that sequential consideration process itself is independent of context.

The first of these is usually the most troublesome. In the absence of any prior knowledge about how an individual partitions the option set, the analyst must proceed by trial and error, estimate a GEV model for one possible decision structure, note its fit, and then compare this fit to that derived for others (e.g., McFadden 1986). While this process is perhaps only an inconvenience in instances where the number of possible structures is small, it becomes burdensome when there are numerous partitioning possibilities, such as in cases where there is the possibility of heterogeneity in decision structures across segments of decision makers.

Although progress is being made towards the development of methods that would allow the partitioning structure of an option to be set prior to estimation of the GEV model (e.g., Currim, Meyer, and Le 1988), these still fail to relieve the second and third limitations, which are far more fundamental. In many instances the IIA assumption will continue to be untenable within option sets in a way which eludes discrete partitioning of these options. For example, when building a model of choices from a menu, one might presume that an individual first partitions the menu into "fish" and "meat" entrees, then, perhaps, into different types of fish and meat (e.g., "steak" or "lamb"). While each final set will thus be homogeneous with respect to these features, it will always be heterogeneous with respect to any remaining features, such as "spiciness" and "price." As such, there will always be the possibility of clusters of items which are more similar than others (unless, of course, the final set contains only one or two options). Finally, while the model's predictions are "context sensitive," the functional form itself is not: It assumes that both the partitioning of options and the order in which subgroups are considered is invariant under changes in the option set. If individuals formulate their eliminations strategies in a strategic fashion depending upon the characteristics of a choice set, the model thus becomes of limited value as an empirical tool.

## The Multinomial Probit Model

The multinomial probit model represents the second major class of models that relaxes the IIA property of the simple multinomial logit by allowing dependencies among the unobserved components of utility. As we mentioned briefly at the outset, the probit is the form of random utility model which arises when the unobserved components of utility in choice are assumed to have a multivariate normal rather than extreme value distribution. Formally, following expression (3.4), the multinomial probit (MNP) model can be expressed as follows:

$$P(i|A) = \int_{-\infty}^{+\infty} d\epsilon_i(\pi j \neq i) \int_{-\infty}^{V(X_i) - V(X_j) + \epsilon_i} d\epsilon_j \phi_N(\epsilon_1, \ldots, \epsilon_N; \Sigma), \tag{3.13}$$

where $\phi_N$ is an N-dimensional multivariate normal density with variance covariance matrix $\Sigma$. The off-diagonal elements of $\Sigma$ capture pairwise dependencies which may exist among the unobserved components of utility.

The central advantage of the MNP model is that it provides the analyst with a much greater degree of flexibility in capturing differing patterns of substitution among options than the GEV model. Specifically, the GEV model's parsimony comes from its assumption that unobserved taste variation can be modeled by a multivariate extreme value (MEV) distribution. A drawback of the MEV, however, is that it is not does yield an arbitrary covariance matrix; as such, the GEV model captures only a prespecified pattern of substitution among options (which are prescribed by a posited hierarchical structure). In contrast, because the covariance matrix of a multivariate normal is flexible, the MNP model can, in principle, be used to represent any pattern of substitution among options.

Because of these potential advantages, the MNP model attracted considerable interest as soon as software for its estimation became available (e.g., Daganzo 1979; Hausman and Wise 1979), and cross-method comparisons with the GEV model indicated that it might be a preferable modeling approach in settings where a hierarchical structure is difficult to prespecify (e.g., Currim 1982). Applications of the procedure continue (e.g., Johnson and Hensher 1982; Kamakura and Srivastava 1985, 1986), although on a more limited scale than the GEV approaches.

The primary limitation of the model, which has inhibited broader application, is that the model poses perhaps the most formidable estimation problems of all the generalized forms that we have discussed. Because the model does not provide a closed-form solution for the probability of choice, estimation is cumbersome, and currently available packages (e.g., Daganzo 1979) provide only approximate solutions for the parameters. Most critically, the accuracy of this approximation has been a point of major controversy in the literature, particularly when the model is applied to choice problems involving more than three alternatives (e.g., Horowitz 1980; Wrigley 1985). As such, serious further consideration of the multinomial probit is probably best put on hold until reliable methods of estimation are devised.

## Accounting for Context Effects within the Simple Logit Form

While generalized distribution models such as the GEV and feature elimination models attract a considerable amount of interest among researchers in the area of formal choice models, they also attract a large number of criticisms. The most important is that even the simple forms pose formidable estimation problems in practice, and this tends to outweigh the advantage of offering a more general description of choice (e.g., Batsell and Polking 1985). In light of these concerns, a number of researchers have developed choice models that were not subject to the IIA property, yet retained the computational simplicity of the basic multinomial logit model. The essence of their approach is to allow the strict utility argument of the multinomial logit to be dependent upon the set of options under consideration.

The psychological rationale for such models is that while the "true" cause of context effects may be the tendency of individuals to make choices in a staged, contingent fashion, within any single choice there may be instability in the sequential structure (e.g., options eliminated and then readmitted for consideration) such that the most parsimonious description may well be a simultaneous consideration model, such as the simple Luce model (Tversky 1972). While context variations would still be presumed to affect choice probabilities, the influence would be on the perceived overall attractiveness of each option, not on altering the basic choice process (the approach taken by deterministic hierarchical models, such as PRETREE or GEV).

Although most such "generalized logit models" have been presented as ad hoc formulations (e.g., Gaudry and Dagenais 1979), they can also be viewed as approximate choice probabilities which arise by assuming a stochastic elimination process, such as in Tversky's elimination-by-aspects (EBA) model. To illustrate, recall that the EBA model, given earlier in expression (3.6), can be written in the form (3.7) reproduced here:

$$P(x|A) = \frac{\Sigma_{\alpha\epsilon x'} u(\alpha)P(x|A_\alpha)}{\Sigma_{y\epsilon A}(\Sigma_{\alpha\epsilon y'} u(\alpha)P(y|A_\alpha))}, \quad (3.14)$$

where, as before $P(x|A)$ is probability that option $x$ is selected from choice set $A$, $u(\alpha)$ is the utility or weight associated with feature $\alpha$ and $A_\alpha$ is that subset of options of $A$ which share this feature. If we multiply the numerator and the denominator by

$$\frac{\Sigma_{\alpha\epsilon y'} u(\alpha)}{\Sigma_{\alpha\epsilon y'} u(\alpha)},$$

(3.14) may be rewritten in the form

$$P(x|A) = \frac{\Sigma_{\alpha\epsilon x'} u(\alpha)\, K_x}{\Sigma_{y\epsilon A}\, (\Sigma_{\alpha\epsilon y'} u(\alpha))\, K_y}, \quad (3.15)$$

where

$$K_y = \frac{\Sigma_{\alpha\epsilon y'} u(\alpha)\, P(y|A_\alpha)}{\Sigma_{\alpha\epsilon y'} u(\alpha)}. \quad (3.16)$$

Expression (3.15) provides an important result: Choice probabilities consistent with EBA can be derived from a simple logit or Luce model in which the overall utility of some option is defined by an independent sum of its attribute worths, multiplied by a scaling factor ($K_y$). Any model which takes on the general form of (3.15) with a consistent scaling of $K_y$ thus yields choice probabilities consistent with EBA.

We should stress, of course, that simply rewriting the EBA model in terms of (3.15) does not make the representation any easier to estimate, since the definition of $K_y$ provided in expression (3.16) remains recursive. Nevertheless, inspection of (3.16) suggests that scaling factor $K_y$ has some rather simple properties which may render approximation straightforward:

1. it is theoretically bounded by 0 (when the features of any one option are shared by an infinite number of other options) and 1;
2. it increases as the number of shared aspects of $x$ decreases; and
3. it achieves the upper bound of 1 in the limiting case where all aspects = $\alpha$ are unique to each option (when $P(y|A_\alpha) = 1$ for all $\alpha$).

Hence, $K_y$ might be seen as a measure of the distinctiveness of option $y$ in the choice set $A$; when all options are unique and hence are equally substitutable ($K_y = K$ for all $y$), (3.15) reduces to a simple logit or Luce model. In general, however, adding new options to a set will affect existing options differentially, depending on their similarity to the new option.

Adjusted logit models can be seen as rescalings of the strict utility argument in the mutinomial logit in a fashion similar to (3.15), but with a more easily estimable function for the scaling factor $K_x$. For example, the first effort in this area was that by Gaudry and Dagenais (1979), who proposed the generalized logit model,

$$Pr(i|A) = \frac{e^{V(X_i)} + \theta_i\, \Sigma_{j\epsilon A}\, e^{V(X_j)}}{(1 + \Sigma_{j\epsilon A}\, \theta_j)\, \Sigma_{j\epsilon A}\, e^{V(X_j)}}, \quad (3.17)$$

where $\theta_i$ was a nonnegative parameter reflecting the degree to which the strict utility of option $i$ would be affected by the introduction of a new option to the choice set. They called the representation the "DOGIT" model, to capture the idea that it is a logit model which "dodges" the assumption of the independence of irrelevant alternatives.

Although the DOGIT model was widely discussed in the literature at the time of its introduction and was used in at least some applications (e.g., Gaudry and Willis 1979), it still offers only a limited solution to the problem of generalizing the basic logit model. Specifically, the substitution parameter $\theta$ is not modeled as a direct function of the similarity of the attributes of options (hence it is not generalizable beyond a specific estimation context), and the model remained computationally complex.

In an effort to overcome these limitations, Batsell (1981), Huber and Sewall (1982), and Meyer and Eagle (1982), and others, suggested similar generalized logit models which directly mirror the form given in expression (3.15). Although the specific proposals vary somewhat, the basic suggestion is that expression (3.15) is equivalent to the adjusted logit model

$$Pr(i|A) = \frac{e^{V(X_i)}\,\theta_i}{\Sigma_{j\epsilon A}\, e^{V(X_j)}\,\theta_j}, \qquad (3.18)$$

where, following (3.16), $\theta_i$ is a 0-1 bounded inverse measure of the similarity of option $i$ to all other options in the choice set. Here $\theta_i$ is either treated as a single scaling constant to be inferred from observed choice data, or parameterized separately, with $\theta_i$ being computed from attribute measures.[4]

A related development by Cooper and Nakanishi (1983a) suggests that distinctiveness effects might be modeled at the level of the individual attribute. Specifically, they describe a family of "zeta-score" models of choice, where the strict utility of an option is modeled as a weighted sum of a set of distinctiveness measures for individual attributes. Specifically, they suggest that a useful measure of the distinctiveness of attribute $k$ of option $i$ in choice set $A$ is the "zeta-score" index,

$$\zeta ikA = \begin{cases} (1 + z^2_{ikA})^{\frac{1}{2}} \; if \; z_{ikA} \geq 0 \\ (1 + z^2_{ikA})^{-\frac{1}{2}} \; if \; z_{ikA} < 0 , \end{cases}$$

where $z^2_{ikA}$ is the standardized value or $z$-score of attribute $k$ of $i$ in set $A$. In general, attribute scores near the mean in a choice set will have zeta scores near 1, those which are low outliers will have zeta scores approaching zero, and those which are high outliers will have zeta scores greater than one. The overall probability of choice is then derived by estimating a standard multinomial logit model, in which the strict utility argument is defined in terms of zeta scores rather than raw attribute values.

Cooper and Nakanishi have reported a number of applications of choice models defined in terms of zeta scores (see Cooper and Nakanishi, 1988, for a review), and report that in almost all instances the transformation provides modeling fits equal to or better than those provided by models estimated by raw attribute values. In other words, in most cases distinctiveness in attributes helps to explain choices. A limitation of zeta-score models, however, is that they model distinctiveness only as it surfaces in ob-

---

[4]To illustrate, Meyer and Eagle (1982) suggest that a simple consistent proxy for $\theta_i$ might be the average Pearson product-moment correlations between options across attributes; in particular,

$$\theta_i = \left(\frac{1}{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \left|\frac{r_{ij} - 1}{2}\right|\right)^{\beta},$$

where $r_{ij}$ is correlation between the observed attributes of options $i$ and $j$. Batsell (1982), in contrast, suggests the weighted similarity measure:

$$\theta_i = exp\left(\frac{1}{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \left( \sum_{k=1}^{m} w_k |x_{ik} - x_{jk}|\right)\right),$$

where $k = 1, \ldots m$ subscripts a series of $m$ attributes, defined by the measures $x_{ik}$, and weighted by the parameters $w_k$.

served attributes of options; hence, violations of IIA which accrue to correlations among measured attributes (such as attributes of appearance) would not be overcome by the model. In such cases the model would have to be augmented by empirical "correction factors."

The most general class of generalized logit models that have been explored are those which modify the strict utility term in a basic logit through a battery of factors representing the pattern of substitution which exists among individual options in a set (e.g., Batsell and Polking 1985; Carpenter, Cooper, Hanssens, and Midgley 1988; Cooper 1988). Termed *competitive-effect models,* they provide a variance decomposition of the single "similarity factor," $\theta_i$, contained in the simple adjusted logit model, expression (3.18). Specifically, the proposed model forms can be shown to be equivalent to letting $\theta_i$ be a linear or multilinear combination of competitor-specific parameters, each measuring the degree of substitutability provided by a particular competing option in the choice set.

To illustrate, imagine that we define $\theta_i$ in expression (3.18) in terms of the additive decomposition,

$$\theta_i = exp\left(\sum_{\substack{j\epsilon A\\ j\neq i}} b_j + \sum_{\substack{j\epsilon A\\ j\neq i}} \sum_{k=1}^{m} b_{jk}\, x_{jk}\right), \tag{3.19}$$

where $x_{ik}$ is the observed score of option $i$ on attribute $k$, and each $b$ is a parameter to be estimated. In addition, assume that the strict utility of option $i$ $V(X_i)$ is also additive in form, as in expression (3.1):

$$Vi(Xi) = b_i + \sum_{k=1}^{m} b_k\, x_{ik} .$$

Substituting these expressions for $V(i)$ and $\theta_i$ in (3.18) the following generalized choice model immediately follows:

$$Pr(i|A) = \frac{e^{V(X_i|A)}}{\Sigma_{j\epsilon A}\, e^{V(X_j|A)}}, \tag{3.20}$$

where

$$V(i|A) = \sum_{j\epsilon A} b_j + \sum_{j\epsilon A} \sum_{k=1}^{m} b_{jk}\, x_{jk} . \tag{3.21}$$

More complex competitive decompositions might, of course, be postulated. Batsell and Polking (1985) and Louviere and Woodworth (1983), for example, note that in some contexts multilinear or higher-order interactive decompositions will be needed to fully characterize the pattern of substitution in a choice setting. The most extreme case is that where the effect of one option on another is fully conditional upon the remaining options in the set (an interaction of order $n$).

The central appeal of decompositions is that they provide a detailed look at the way in which changes in the composition of a choice set affects the likelihood that a given option will be chosen. Competitor-specific parameters, for example, measure the extent to which the mean likelihood of choosing option $i$ will be affected by the presence or absence of another option $j$ in the choice set, controlling for the observed attributes of $i$ and $j$. Likewise, the competitor attribute effect measures of the extent to which $i$'s choice share will be differentially affected by competitors' attribute changes (Cooper and Nakanishi 1988). Hence, in addition to providing a predictive model of choice, the parameters of the model might be the focus of separate analyses in their own right. Cooper (1988), for example, has illustrated how the coefficients of competitive-effects models can be used to infer maps of the structure of competition within a product class.

The major drawback of competitive-effect models is their lack of parsimony; the large number of parameters lend interpretability problems to the model, and complete model forms are not always estimable given normally available data (see, e.g., Currim, Meyer, and Le 1988; Carpenter, Cooper, Hanssens, and Midgley 1988).[5] In addition, the model parameters simply serve to describe competitive effects, not explain them. As such, the models are of limited usefulness in contexts where new options are being added to a set, or when there are changes in the characteristics of options which have the potential of altering the competitive relationship among options. Nevertheless, that the models potentially allow one to model complex substitution patterns among options within an analytic framework of a simple logit model has caused them to be widely seen as the state of the art in mathematical choice modeling.

## Context Effects on Choice: What Current Models Can and Cannot Explain

***The Effects of Item Similarity and Dominance.*** The models discussed all recognize that the process by which an individual chooses an option from a set will be sensitive to the external appearance of that set. These efforts, however, have been generally designed to capture only one type of context effect, the so-called Debreu counter-example to IIA, or the phenomenon that items which are more similar in a choice set compete more closely (given a fixed set of choices) than those which are more distinctive. After twenty years of work toward this goal, it can reasonably be concluded that a range of satisfactory solutions has been obtained. There exist three major families of models which capture such similarity effects, each offering a possible approach. If similarities among a set of items seem best described in terms of a set of hierarchical partitions, such as in cases where the decision maker faces a large number of choice options, the GEV model provides a straightforward, and usually easily estimable, modeling solution. If a hierarchical representation is difficult to specify or appears theoretically inappropriate, an adjusted logit model can be specified. Finally, if the analyst is modeling choice among a small set of options and the source of violations of IIA are thought best attributed to correlations among the unobserved components of utility, the multinomial probit model is a plausible option.

Unfortunately, there exist a broader range of empirically observed context effects which have proven more difficult to model. For example, Huber and Puto (1983) note that the effect of item similarity can actually, at times, be a complex one. While existing non-IIA models tend to predict that as two options become more similar their choice shares should also become similar, the opposite may be the case in reality; because similar options are also more easily comparable, their choice shares may in fact *diverge* if it becomes apparent that one has a decisive edge on an attribute. In particular, given two options which are identical on all attributes except for one, intuition suggests that the option which is inferior on this remaining option will have little rational probability of being chosen. While this "dominance" effect is theoretically captured in feature-based models such as EBA and PRETREE, it is not captured in algebraic multiattribute models, such as GEV, the adjusted logit, or MNP.[6]

Some potential solutions have been offered for how the simple logit model might be gener-

[5] Louviere and Woodworth (1983) point out that the multicollinearity problems which plague econometric applications of competitive-effect models are less of a barrier in laboratory applications of the model, in which data is drawn from experimentally designed choice sets.

[6] The GEV, or nested logit model, can capture the effect of dominance if the choice model in each stage is composed on only one attribute, or if the nested logit process is the same as PRETREE. Our critique centers on the more general case where choice model at each stage is composed of multiple attributes. To illustrate, if two options are identical on two attributes and the choice probabilities are 0.5, ML estimates of the attribute parameters will also be equal (or indeterminant). If one option then became a slight bit better on one attribute, intuition suggests that probabilities would go to 1 and 0 for each option, but equal weights on the attributes would predict choice probabilities remaining near 0.5.

alized to account for dominance effects, however none seem truly satisfactory. For example, Meyer and Eagle (1982) suggested that the dominance effect might be modeled by presuming a "weight-shifting" process in multiattribute evaluations: When making a choice between two options, the amount of weight given to a particular attribute in choice will depend on the amount of variance displayed by that attribute relative to the variability displayed by other attributes. As options become increasingly similar on a set of dimensions, choice thus becomes increasingly driven by the subset of attributes which continue to differ across options. Hence, as intuition would suggest, given a choice between two alternatives which differ on only one attribute, choice probabilities polarize to one and zero. Meyer and Eagle apply this model to a study of binary store choice and find good explanatory ability. Unfortunately, a drawback of the approach is that it is not parsimoniously generalized to multinomial choice contexts; more critically, estimation requires choice probabilities for all possible paired comparisons of options with a given choice setting.

A more perplexing effect of item similarity has been noted in a series of studies by Huber, Payne, and Puto (1982), Huber and Puto (1983), and Ratneshwar, Shocker, and Stewart (1987). They report that when an option which is identical to another on several dimensions, but worse on one, is added to a choice set—an alternative which is normatively irrelevant—it actually serves to *increase* the choice share of the option which dominates it. They characterize this as part of a general "attraction effect," in which the joint likelihood of purchasing a group of similar items in a choice set tends to be greater than would be expected based on their independent preferences alone.

Kahn, Moore and Glazer (1987) and Kahn and Lehmann (1989) suggest that the phenomenon is perhaps most plausibly explained as a "portfolio effect": when asked to choose among groups of items from which one will eventually be selected, individuals exhibit a tendency toward selecting that group with the largest number of elements, presumably because it is thought to give them most flexibility when a final choice is made. When there are differential patterns of similarity in a choice set, options naturally fall into similar subgroups. Individuals may be analogously lured toward initially focussing on the set with the largest number of elements, and then choosing the most preferred option from the subset.

***Choice Set Formation.*** The "attraction effect" may be a consequence of a much broader problem facing applications of current choice models: How does the process by which individuals form evoked sets, or the subset of options which are seriously evaluated at the time of choice, affect the final choice outcome? The pragmatic consequence of ignoring the consideration process (as is almost universally done in field applications of logit-type choice models) is that model parameters have limited psychological meaning: While attribute coefficients are theoretically measures of the effect of variations in an attribute on the strict utility of a choice option, this is not the case if only a subset of the individuals under study were considering the option at the time of choice. Hence, in general, attribute effect parameters will be biased downward from the "true" value which would be obtained if the model was specified only for actively considered subsets of options. As an illustration, in applications of aggregate market-share models analysts often report findings of low attribute elasticities for small market share brands (e.g., Cooper 1988). While one might interpret this result as meaning that such brands are purchased by a small group of "loyal" buyers, another, perhaps more plausible, explanation is that it is an artifact of estimating an aggregate model over a heterogeneous buyer group, most of which never consider the smaller brands, with those who do consider the smaller brands actually being highly attribute sensitive. While this distinction would seemingly be an important one from a managerial perspective, it is confounded in current models.

How to specify choice sets at an individual level has long been a major stumbling block in applications of formal choice models (cf. Fotheringham 1989). Heuristic solutions that have been proposed include defining the relevant choice set as that set of options which have been chosen at least once during a time interval (e.g., Hensher and Johnson 1981), or, if survey information is available, directly asking consumers for their "usual" set of considered options (e.g., Silk and Urban 1979). Models which allow an analyst to predict *variation* in considerations sets from occasion to occasion, however, have proven more elusive. The only such model we are aware of that has been implemented is that proposed by Gensch (1987), who uses his MLH algorithm (discussed earlier) to conduct a hierarchical screening which determines the "best" $n$ options on a given choice occasion (in his application $n$ is arbitarily 4). He then calibrates a multinomial logit model under the assumption that these $n$ options form the relevant consideration set.

The paradox of Gensch's approach, however, is that it assumes that all options are considered (at least initially) in the hierarchical screening. Hauser and Wernerfelt (1988), Meyer (1980), and Richards (1982) have proposed that this difficulty be circumvented by modeling the choice set formation process as a Bayesian search problem, in which options enter the choice set when their hypothesized utility exceeds some reservation level (an idea we will elaborate on later when we discuss learning processes).

***Other Effects of Context.*** Finally, there exist a number of other, more subtle, ways in which choice context has been observed to affect the process of choice. Johnson and Meyer (1984), for example, note that the parameters of choice models tend to become increasingly skewed as the size of the consideration set grows; in particular, with larger choice sets, choice tends to be driven by observed variation on a smaller number of attributes. Similarly, Eagle (1985), Lynch and Hutchinson (1984), and Meyer and Louviere (1982) have found evidence for a general effect of attribute variance: Individuals tend to attend more closely to attributes which exhibit greater variability in a choice set. Although such effects are usually not represented within applied choice models, they are more easily overcome than dominance, attraction, and choice set effects noted earlier. Specifically, as illustrated by Borgers and Timmermans (1987) and by Meyer and Louviere (1982), given choice experiments in which set size and variance are manipulated, these effects can be directly recognized as interactive effects in the strict utility function of a traditional logit.

## CHOICE DYNAMICS

The individual choice models described previously are inherently static models. While they can—and often are—estimated by observing a sequence of choices made over time (e.g., Elrod 1988), such applications are based on an assumption which is questionable in practice: the assumption that choices on one occasion are independent of those made on subsequent occasions. Violations of this assumption arise frequently: Tastes may systematically change as a consequence of learning or habit formation (e.g., Kuehn 1962, Jeuland 1978) or as an explicit desire among individuals to seek variety in choice (e.g., McAlister and Pessemier 1982, McAlister 1982), or both (Bawa forthcoming, Kahn, Kalwani and Morrison 1986).

Over the years there has been extensive modeling research which has explored the nature of temporal effects on choice. The work has proceeded in two ways. The first direction has been to develop model forms for choices made over time among options which are familiar to the decision makers. These models, which generally have focussed on the effects of loyal/habitual or variety-seeking behavior, have traditionally not incorporated explanatory variables, but rather have tried to provide parsimonious constructs for measuring the degree to which past choice behavior influences current choice behavior. Researchers have begun to extend these simple models to include explanatory market variables which further explain the patterns of switching over time (e.g., Carpenter and Lehmann 1985; Zufryden 1986).

The second direction has been to develop model forms for choices made in situations when full information is initially not available for all items in the choice set. In these cases, temporal dependencies result from individuals learning the attributes of options over time through choice.

## Models Which Measure the Effects of Temporal Dependencies

There are three families of models that have been used to measure temporal effects on brand choice: Markov models (primarily first-order), linear learning models, and the incorporation of a term which is the weighted average of past purchases into the simple multinomial logit model. All three of these types of modeling efforts have been used to measure both brand loyal and variety-seeking or switching behavior.

***Markov Models.*** For first-order Markov models the probability of purchasing a brand is dependent upon the purchase which was made at the previous purchase occasion. These conditional probabilities are generally represented in a transition matrix. To illustrate, consider a market where there are two choice options, Brand 1 and Brand 0. In a first-order Markov model the conditional probabilities of a purchase of $i$ at time $t + 1$ given a purchase of $j$ at time $t$, $P(i|j)$, is given by the matrix:

$$\begin{array}{ccc} & & \text{Brand Purchased at } t+1 \\ & & \begin{array}{cc} 1 & 0 \end{array} \\ \text{Brand Purchased at } t & \begin{array}{c} 1 \\ \\ 0 \end{array} & \begin{bmatrix} P(1|1) & P(0|1) \\ & \\ P(1|0) & P(0|0) \end{bmatrix} \end{array}$$

This model could be extended to the study of markets with multiple brands, either by assuming that Brand 0 represents all other brands and Brand 1 represents the brand of interest (e.g., Massy, Montgomery, and Morrison 1970), or by extending the matrix to n brands (e.g., Givon 1984).

If the two rows of this transition matrix have equal entries, then there are no temporal dependencies. In that case, $P(1|1) = P(1|0) = P(1)$ and choice behavior is said to be zero-order, which implies that the probability of purchase for a brand on each trial is independent of any past purchase behavior (Kahn, Kalwani, and Wright 1986).[7] In contrast, if the two rows are not equal, there is evidence for temporal dependency. For example, if the probability of choosing a brand increases if the brand has been chosen before, when $(P1|1)$ is greater than $P(1))$, then there is evidence of brand loyal behavior. On the other hand, if the probability of choosing a brand decreases if the brand has been chosen before, when $(P(1|1)$ is less than $P(1))$, this would reflect variety-seeking behavior (e.g., Givon 1984, McAlister, and Pessemier 1982; McAlister 1982; Jeuland 1978). The exact modeling specification for these two types of conditional probabilities, of course, is open to speculation. We will review proposed approaches to each.

***Markov Models for Brand Loyal Behavior.*** One of the first researchers to use a first-order Markov model to measure brand loyalty was Lipstein (1959). Although he did not offer any specific model, he suggested that the repeat purchase probabilities, for example, $P(1|1)$ and $P(0|0)$, be used as indices of loyalty. Telser (1962) was one of the first researchers to investigate the relationship between a marketing variable, price, and patterns of brand choice switching over time. He specified a linear relationship between the transition probabilities and the difference between the price of the brand considered and the average price of all other brands.

Morrison (1966) explicitly modeled brand loyal behavior in a first-order Markov model. He developed two first-order Markov models: the Brand Loyal model and the Last Purchase Loyal model. In the Brand Loyal model an individual with a high probability of remaining

[7]The simple multinomial logit models described earlier were zero-order models of behavior as they did not recognize the effect of past purchase history on current history.

with Brand 1 would also have a higher probability of choosing brand 1 after a purchase of Brand 0. Morrison modeled this probability as follows:[8]

$$P(1|1) = p$$

$$P(1|0) = kp$$

where:

$p$ = distributed beta across the population
$k$ = a positive constant which is the same for all individuals

In the Last Purchase Loyal model an individual is more loyal to the brand last purchased than any brand in particular. Morrison modeled this dependency as follows:

$$P(1|1) = p$$

$$P(1|0) = 1 - kp$$

where:

$p$ = distributed beta across the population
$k$ = a positive constant which is the same for all individuals

Although Morrison's models allow for behavior on one choice occasion to influence behavior on another, his models do not provide a convenient way of measuring the strength of this influence. Jeuland (1979) has extended Morrison's representations to include a specific parameter to measure inertia or loyal behavior. In his extension, the probability of repeat purchasing brand $i$ is increased by a fraction of the maximum possible increase if brand $i$ was purchased at the previous purchase occasion. His model is characterized as follows:

$$P(1|1) = \theta + I(1 - \theta)$$

$$P(0|0) = (1 - \theta) + I\theta$$

[8]Note that in all first-order matrices $P(1|1)$ and $P(1|0)$ = 1 and $P(1|0)$ and $P(0|0)$ = 1, so the entire matrix is characterized by just defining one entry on each row of the matrix.

where:

$\theta$ = zero-order probability of purchasing brand 1
I = an inertia parameter which varies from 0 to 1

In this model, the parameter $I$ captures the degree of dependency from one choice occasion to the next. If $I$ equals one then the choice on one occasion completely determines the choice on the next occasion; if $I$ equals zero then the choice process reduces the zero-order.

***Markov Models for Variety-Seeking Behavior.*** Jeuland's model provides a way to parsimoniously measure the degree of loyalty or inertia represented from one choice occasion to the next. The next logical step is to use this type of methodology to investigate the "opposite" of brand loyal behavior, or variety-seeking behavior, in which choosing the brand on one occasion decreases the probability of choosing the same brand on the next occasion. Givon (1984) did exactly that. He used the Markov model to measure variety-seeking behavior as well as loyal behavior. In this model, the choice made on the last purchase occasion can either increase the probability of choosing the same brand on this occasion (i.e., loyal behavior), decrease the probability of choosing the same brand on this occasion (i.e., variety-seeking behavior) or have no effect (i.e., zero-order behavior).

He defines a parameter $VS$ which ranges from $-1$ to 1. When $VS$ is positive then brand loyal behavior is occurring, when $VS$ is negative then variety-seeking behavior is occurring, and, finally, when $VS$ is zero, then zero-order behavior is occurring. Although Givon defines his model for n brands, the simple 2-brand conditional probabilities are as follows:

$$P(1|1) = (|VS| - VS)/2 + (1 - |VS|)\theta$$

$$P(1|0) = (|VS| + VS)/2 + (1 - |VS|)\theta$$

where:

$\theta$ = preference for brand 1
$VS$ = measure of the effect of last period's choice on this period choice

Kahn, Kalwani, and Morrison (1986) used Givon's and Jeuland's models to form the basis of a taxonomy that integrates brand loyalty and variety-seeking research into one framework. In addition to the two first-order models, they also defined four second-order Markov models. In these models, the probability of choosing a brand on the current occasion depended on the choices made on the past two occasions. By adding the second order models, Kahn et al. could describe a large spectrum of choice behavior which depended on previous choices made. Bawa (forthcoming) also considers higher-order behavior. He shows how the preference function for a brand behaves nonmonotonically as repeat purchasing or switching take place.

## Markov Models of Brand Switching That Incorporate Explanatory Variables

Although the above stochastic models provide a framework of parameters for measuring the degree of dependency of the current choice on the last period or last two period choices, they do not provide any help in determining why that dependency exists. Some attempt has been made to explain this dependency within the Markov model. Lattin and McAlister (1985) and Feinberg, Kahn, and McAlister (1989) use measures of the similarity between brands to help explain the patterns of dependencies associated with variety-seeking behavior. Specifically, for markets in which there is reason to believe that consumers are influenced by a desire to seek variety in choice, Lattin and McAlister and Feinberg, Kahn, and McAlister offer the following model of the conditional probability that some option $i$ will be chosen given a previous choice of a different option $j$:

$$p_{i|j} = \frac{\pi_i - V(S_{ij})}{1 - V\,\Sigma_k S_{kj}}, \qquad (3.22)$$

where $\pi_i$ is a measure of the overall attractiveness of brand $i$, $S_{ij}$ is a measure of the value of the product features shared by both $i$ and $j$ (a measure of similarity), and $V$ is a (0,1) bounded scaling factor measuring the consumer's need for variety in choice. This model has a straightforward interpretation: It hypothesizes if a consumer has a strong need for variety, the likelihood that he or she will buy brand $i$ after previously buying brand $j$ increases with the independent attractiveness of $i(\pi_i)$, its distinctiveness vis-à-vis brand $j$, and its distinctiveness vis-à-vis all other competing options. Aspects which are unique to an alternative thus have a positive impact on the conditional probability of choosing that alternative. Although their model is restricted to the study of variety-seeking tendencies, clearly similarity could also be used to help explain the dependencies for a brand loyal consumer as well. In an analogous fashion, Kahn and Raju (1990) propose a model which extends the Markov framework to include the effects of frequency of promotion on variety-seeking and loyal behavior.

Other researchers have looked at a broader range of explanatory variables to investigate the effect on choice of brands over time. These researchers have used the simple multinomial logit model discussed earlier to model the transition probabilities of the Markov model. For example, Carpenter and Lehmann (1985) used a multinomial logit model to express the conditional probabilities as a function of marketing mix variables, product features, and their interactions. In their model, individuals are grouped to form segments according to their last purchase. For example, consider a segment $j$, defined as the set of consumers who purchased brand $j$ in time $t$. Assume that for this segment switching to brand $i$ of a set of $M$ brands yields measurable utility $v_{i|j}$. The probability of switching from brand $j$ at time $t$ to brand $i$ at time $t + 1$, $p_{i|j}$ (which is equivalent to the conditional probability entry in a Markov matrix) is then modeled as:

$$P_{i|j} = e^{v_{i|j}} \Big/ \sum_{k \in M} e^{v_{i|k}}$$

Carpenter and Lehmann express $v_{i|j}$, the utility of switching to product $i$ from $j$, as a function of the attributes and marketing activities of both products. They have tested their model on consumer nondurable data and have shown how price promotion, advertising, and brand name strength affect both loyal and switching behavior.

In a similar manner, Jones and Zufryden (1980) and Zufryden (1980, 1981) model the transition probabilities within a two-brand Markov model as a binomial logit model. Within this formulation, Zufryden could relate purchase-explanatory variables (e.g., product features, pricing, consumer characteristics, etc.) to switching probabilities. Zufryden (1986) then extended his earlier model to consider multibrand market situations.

*Linear Learning Models.* Another type of modeling form that has been used to represent the dependencies of past choices on the current choice is the linear learning model. This model, originally proposed in psychology by Bush and Mosteller (1955), represents the probability of purchasing a brand at time $t$ as a linear function of whether or not the brand is purchased on occasion $t - 1$. The parameters of the linear relationship are a function of the past purchase history. In this way, the model is both first-order, because it directly depends on the purchase made on the last occasion, and infinite order, because the coefficients of the linear relationship are derived from all previous choices.

Kuehn (1962) was the first to apply the linear learning model to study the study of temporal buying patterns, with other applications reported by Aaker (1970), Carman (1966), Massy (1970), and McConnell (1968). In these applications it is assumed that if a brand was purchased on purchase occasion $t$ then the probability of purchasing the brand on occasion $t + 1$ would increase. If the brand was rejected on occasion $t$ then the probability of purchasing it on $t + 1$ would decrease. The two influences are captured through two linear relationships: a purchase operator and a rejection operator. The acceptance operator is given by:

$$p_{t+1} = \alpha + \beta + \lambda p_t,$$

and the rejection operator is given by:

$$p_{t+1} = \alpha + \lambda p_t,$$

where $\alpha$, ß, and $\lambda$ are nonnegative parameters determined from the purchase history, and are subject to the constraint $(\alpha + \beta + \lambda) \leq 1$. Srinivasan and Kesavan (1976) note that in the special case where $\alpha = 0$ and $(\beta + \lambda) = 1$, the linear learning model becomes equivalent to an exponential smoothing model, in which the probability of a brand's choice at one point in time is modeled as the weighted sum of previous choice outcomes, with recent outcomes being weighted most heavily.

Although the linear model has often proven to be a useful descriptive device (e.g., Srinivasan and Kesavan 1976), it has been subject to two sets of criticisms. First, as originally formulated, the model offers a representation of an *individual* consumer's learning process, hence it can be applied to the study of aggregates only under the assumption that individuals in a sample are homogeneous in their learning processes—an assumption which will often be empirically untenable. This problem has been addressed first by Jones (1970) with a model which was difficult to estimate and, subsequently, by Givon and Horsky (1978, 1979). Givon and Horsky (1978) developed a heterogeneous linear learning model which allowed consideration of an entire population and nested a first-order Markov model and a zero-order model as well. A second criticism, suggested by Aaker (1970), is that even in cases where heterogeneity in parameters is not a concern the model may still fail if yet another assumption is violated: that of stationarity in parameters over time. Specifically, in some settings the likelihood that an individual will switch to an option may vary over time in response to both changes in the set of choice alternatives (for example, one option being made more attractive due to a lowering of its price, or due to new options being introduced), as well as changes in the decision maker's tastes (e.g., a desire to seek variety). These sorts of influences, however, are not explicitly recognized by the model.

Unlike the Markov models, the linear learning model has not been readopted by researchers in the 1980s aside from a few attempts to use it to measure variety-seeking behavior by reversing the rejection and acceptance operators (e.g., Sharma and Durand 1980). Instead models of learning have been advanced based on theories of risk aversion, as will be discussed in a later section.

***Exploring the Effect of Past Choices in the Multinomial Logit Model.*** Although as mentioned earlier, the simple multinomial logit model is a zero-order model, it can be extended to represent dynamic choice behavior by introducing distributed lags of previous purchases in the indirect utility function of the logit (e.g., Guadagni 1983, Guadagni and Little 1983, Gupta 1988, Meyer and Cooper 1986). For example, Guadagni and Little (1983) incorporate "brand loyalty" and "size loyalty" in their multinomial logit analysis of coffee purchases by nesting an exponential smoothing model of previous brand and previous size purchases. Specifically, they include an independent variable in their strict utility function which they called "loyalty" and which is defined as follows:

$$x_k^i(n) = \alpha_b x_k^i(n-1) + (1-\alpha_b)z$$

where:

$$z = \left\{ \begin{array}{l} 1 \text{ if customer i bought brand } k \\ \quad \text{on purchase occasion (n-1)} \\ 0 \text{ otherwise} \end{array} \right\},$$

$x_k^i(n)$ = loyalty for brand $k$ for $n$th purchase of customer $i$

$\alpha_b$ = carryover constant

This modeling approach also closely parallels one proposed by Keon (1979). He suggested that, over a series of successive choice occasions, choice patterns will be driven by three elements: (1) the relative zero-order desirability of a product's attribute mix relative to others, (2) the tendency for the attractiveness of a brand to be "forgotten" given the absence of a purchase, and (3) the tendency for consumers to become bored or satiated with a brand given a purchase. Brand loyalty would be modeled by a dominance of the forgetting effect: while the purchase of a brand at one point in time would not enhance its attractiveness per se, it would enhance its relative attractiveness by the gradual reversion of the perceived attractiveness of the options to a neutral level. On the other hand, variety-seeking behavior would be modeled by a dominance of the satiation effect.

Lattin (1987) developed a "balance model" of choice behavior, which uses the multinomial logit form and includes a variable accounting for the effects of past choices on the current choice. Lattin assumes that the utility for a brand is the sum of the utilities provided by the characteristics or attributes constituting the brand. He then further assumes that the utility derived from these product characteristics depends on its salient quality in memory. The "salience" refers to the lingering impact of the characteristics after consumption. The salience of the characteristics can have either a negative (i.e., variety-seeking type behavior) or a positive effect (i.e., brand loyal behavior). The form Lattin used for item salience is identical to the exponential smoothing model used by Guadagni and Little (1983) as their construct of loyalty; however, in Lattin's model the specification permits variety-seeking as well as loyal behavior in the choice of brands and attributes.

***Summary: Models Recognizing Temporal Dependencies.*** Although different in form, Markov models, linear learning models, and multinomial logit models, which include "loyalty" variables, share a common objective: They each are attempts to parametrize or measure the degree to which brand choices made in one period affect the brand choices in subsequent periods. For example, in Kahn, Kalwani, and Morrison (1986), the models provided an estimate of the degree to which the previous choice reinforced (or negatively reinforced) the choice made on the next period. In the linear learning models these measures were provided by the coefficients in the linear model. Finally, in the models similar to Guadagni and Little's (1983) model, the coefficient of the "loyalty" or "salience" terms measures the importance of previous brand choices on the current choice.

While these models of loyalty and variety-seeking behavior recognize the systematic dependencies among choices over time, the effects are limited to those of habituation and/or satiation and not to *changes* in information about the option set.

An increasing number of authors have argued that full-information assumptions may be violated in a number of consumer buying situations. (e.g., Ford and Smith 1987; Levin and Johnson 1985). In many cases, such as in new product decisions, because consumers have limited experience in consuming differing options, choices will be driven by consumers' hypotheses about the utilities of each option, which may bear little resemblance to the "true" utilities these options might hold given a greater range of experiences by the consumers. As a consequence, consumer choices will invariably involve at least some level of uncertainty, with this uncertainty varying over time as a consequence of both individuals becoming more familiar with the option set, and changes in the option set itself.

## Models of Dynamic Choice Under Uncertainty

Several authors have recently offered suggestions for how impression-formation models might be generalized to account for the effects of attribute uncertainty and learning, with Meyer and Sathi (1985) and Roberts and Urban (1987) offering extensions to the study of choice from competitive sets. Although their formulations differ somewhat in structure, both draw their roots from economic expected utility theory.

The basic approach is as follows: As in static choice theory under certainty, we begin with the assumption that individuals choose that option which is thought to deliver the highest level of utility at the time of choice. Because the decision maker is uncertain as to what this value may be given a choice, however, he or she is treated as an expected utility maximizer; that is, the decision maker selects the option which holds the highest expectation given a distribution of possible values.

As before, it is assumed that overall impressions of choice options can be modeled in terms of a multiattribute function of their attributes. In this case, however, it is assumed that these attribute values are not known with certainty prior to choice; rather, they are random variables, each characterized by a distribution of possible values. The expected utility of given option $i$, $EU_i$, is thus represented as a multiattribute function of the expected utility of each attribute, as in the linear-additive form,

$$EV_i = \sum_{k=1}^{m} b_{ik}\, EU(x_{ik}) \ ,$$

where $EU(x_{ik})$ is the expected utility of the random variable $x_{ik}$, which is the actual value of the $k$th attribute of option $i$.

Roberts and Urban (1987) derive a normative expression for $EU(x_{ik})$ by assuming that individuals are risk averse in decisions under uncertainty, and treat $x_{ik}$ as a random draw from a normal distribution with mean $\bar{x}_{ik}$ and variance $\sigma^2_{ik}$. Specifically, following Keeney and Raiffa (1976) they propose:

$$EU(x_{ik}) = \bar{x}_{\mathrm{ik}} - \frac{r}{2}\, \sigma^2_{ik} \ , \tag{3.23}$$

where $r$ is a measure of the decision maker's attitude toward risk. Expression (3.23) has a simple intuition: When faced with the task of evaluating an uncertain choice option, impressions are formed by integrating a set of risk-adjusted expectations for the value of each attribute. If the risk parameter $r$ is positive, implying risk aversion, the model captures a well-known finding in studies of choice under limited information: Increasing uncertainty in attribute values (increases in the variance $\sigma^2_{ik}$) leads to diminished expected utilities (e.g., Ford and Smith 1987; Meyer 1981). Hence, given a choice between a certain option with a known attribute value and an uncertain option with the same mean, most will prefer the certain option (e.g., Huber and McCann 1983; Levin and Johnson 1985; Yakanishi and Hill 1979).

A limitation of expression (3.23) as an empirical device, however, is that its predictions may not perfectly correspond to actual subjective expected utilities provided by individuals. Specifically, while individuals may form expectations through some notion of central tendency and dispersion, it may not necessarily correspond to the simple difference given in expression (3.22). To account for a broader range of processing rules, Meyer and Sathi (1985) suggest the following alternative definition of $EU_i$:

$$EU\,(x_{ik}) \;=\; k_1 u_{\bar{x}}(\bar{x}_{ik}) \;+\; k_2 u_D\,(D_{ik}) \;+\; k_3 u_{\bar{x}}(\bar{x}_{ik}) u_D(D_{ik}) \;, \tag{3.24}$$

where $D_{ik}$ is a measure of the perceived dispersion around the mean expected attribute level $\bar{x}_{ik}$, $u_{\bar{x}}(\bullet)$ and $u_D(\bullet)$ are utility mappings, and $K_1$, $K_2$, and $K_3$ are normalized scaling constants. Like expression (3.23), (3.24) also captures a negative effect of uncertainty in the case where $u_D(\bullet)$ decreases with increases in perceived dispersion $D_{ik}$. The argued advantage of (3.24) is that it allows greater flexibility in the scaling of means and dispersions, and allows for nonadditive integration rules.

Ford and Smith (1987), Levin and Johnson (1985), and Meyer (1982, 1985) have also explored generalizations of mean/variance models, similar to expressions (3.23) and (3.24), that recognize the possibility of cross-attribute inferences (for more extensive discussion see Chapter 11 by van Raaij in this volume). Specifically, the expected utility of a given attribute is modeled as a function of not only the mean and variance of that attribute, but also as a weighted sum of the observed value of other, certain, attributes. In this case the weights correspond to the perceived covariance between the certain and uncertain attributes.[9]

Central to these developments is the notion that expected values are subject to change through learning. Specifically, a choice on one occasion will likely serve to alter the decision maker's inferences about the utility of the chosen option for the next occasion by yielding a more certain estimate of its expected utility. Both Meyer and Sathi (1985) and Roberts and Urban (1987) posit Bayesian-like updating models for this process, in which the means and variances underlying a given expected utility are modeled as a weighted average of a "no information" baseline or prior and a full-information value.

By successively substituting the updating models into a model of expected value, generalized evaluation models are provided which enable simple choice models (such as the multinomial logit) to be applied to contexts in which consumers initially have limited information about the attributes of options but in which the information is updated over time. Applications of models of this sort have been made in laboratory studies of choice and learning (e.g., Meyer and Sathi 1985), and in the real-world application of changing consumer preferences for new automobiles (Roberts and Urban 1987).

***Search and the Formation of Choice Sets.*** Although the above family of models appears to be a reasonable first step in modeling choice dynamics given limited information about options, they hold an important limitation: As noted by Horsky and Raban (1988), the models are limited to characterizing "rational" choice under uncertainty in the limited case where the decision maker seeks to maximize utility for the immediate choice occasion, and does not consider its impact on the *stream* of utilities gained from future choices. Where this assumption is most problematic is in instances where a particular choice is viewed as part of an information-gathering or search process.

A convenient example is selections among restaurants: A primary rationale for trying out a new restaurant may not be that it is thought to deliver the highest utility among all restaurants for a given evening, but rather that one's total future utility gained from dining out will be enhanced if it can be added to the future "choice set."

[9] Meyer (1985) derives a normative cross-attribute inference model by assuming that individuals view attribute values as drawn from an $m$-dimensional multivariate normal distribution with covariance matrix $\sigma$. The forms presented by Ford and Smith (1987), Levin and Johnson (1985), and Meyer (1982) are structurally similar, although are not derived from a normative perspective.

While search effects have been discussed to some degree in the choice modeling literature (e.g., Richards 1982), there is no emerging consensus as to how they might best be represented. The most direct approach would simply be to assume that search effects are manifested in the risk-attitude parameter of a mean/variance model, as in expression (3.22). If the individual is actively seeking to expand his or her choice set, he or she should exhibit a preference for less certain over more certain options. The drawback of this modeling approach, however, is that search strategies and risk attitudes are conceptually quite different constructs; an individual may be generally risk averse, yet still find it desirable to seek out unfamiliar options as an "investment" in the future utility stream. By failing to separate them, one may observe considerable temporal variance in apparent risk attitudes.

Horsky and Raban (1988) model product choice over time as a "multiarmed bandit" problem, in which the individual makes sequential choices from a set of options, each characterized by a distribution of utility values.[10] This presumes that the decision maker's objective on each choice occasion is to maximize a total stream of utilities across a series of choices. The consequence of the model is a prediction about patterns of switching among options similar to that often reported in studies of temporal choice under uncertainty (e.g. Meyer and Sathi 1985); over time, individuals should first act as risk seekers, actively trying out new options, but then progressively act as risk avoiders, repeatedly selecting that limited subset which delivers the "safest" level of utility.

## Recent Issues in Modeling Choice Dynamics

As an increasing number of authors have turned to the problem of estimating models of consumer choice on time-series data, the array of problems posed in such applications has become increasingly apparent. In the previous discussion we focussed on current solutions to the three most well-understood issues in dynamics: brand loyalty, variety-seeking, and attribute learning. These are, however, only a subset of the possible sources of dependency in choice.

For example, several authors (Bass and Pilon 1981, Kahn and Louie 1990, Winer 1983) have noted that there will often be temporal dependencies in attribute effects that are distinct from temporal dependencies in choices. Perhaps the most well known is the effect of advertising: The effect of brand advertising on the choice observed at time $t$ may be a weighted cumulative function of the number of exposures to that advertisement, as well as the time since exposure. In such cases, the problem of correctly specifying a distributed lag model in exposures becomes paramount, and often not easily solvable (e.g., Carpenter, Cooper, Hanssens, and Midgley 1989).

A somewhat different potential temporal effect is that of price variation. Winer (1983), for example, reports that consumers often judge the expensiveness of a brand not simply relative to the prices of competitors (as would be presumed by a traditional multinominal logit model), but also with reference to its "expected" price, or at the price it would "normally" be sold. Exactly how consumers form such expectations is currently an issue of some debate. Winer (1983), for example, offers evidence that expectations can be modeled as a simple adaptive expectations model, in which expectations are represented as a weighted average of previously observed prices. While the adaptive expectations model is intuitively appealing, its weakness is that it becomes inappropriate in cases where consumers base expectations on longer-term trends and periodicities in a price series, such as anticipating monthly discounts.

Finally, perhaps the most difficult technical problem which has surfaced in attempts to apply discrete choice models to the study of behavior over time is that of temporal dependencies among the unobserved components of utility. While time-series methods for the treat-

[10]"Armed bandit problems" refer to a family of sequential sampling problems in which a gambler faces a set of slot machines, and his or her goal is to determine that with the highest average or total payoff.

ment of autocorrelated errors given continuous dependent variables are well developed, this is not the case for discrete multinomial models, such as the multinomial logit (e.g., Heckman 1981). There have been two recent efforts to remedy this shortcoming; Heckman (1981) and Sugita (1986) have proposed probit models for the analysis of binary (two alternative) panel data, in which the unobserved components of utility over time are presumed to be characterized by a multivariate normal distribution. Unfortunately, the models are currently not easily estimable, and we are aware of few applications beyond those reported in the original works. The current "solution" to such problems is either thus one of incorporating a rich enough array of temporal variables in the strict utility function so as to minimize the potential for autocorrelation, or to estimate models on aggregated choice shares, where standard least-squares procedures can be exploited (e.g., Carpenter, et al. 1989).

## FUTURE RESEARCH

### Toward a Theory of Consumer Choice Models

The evolution of the literature in mathematical models of consumer choice reflects a dialogue between two quite different research motivations: that of developing simple and accurate tools for predicting the outcome of individual choices, and that of providing an accurate psychological explanation for these choices. On one hand, these motivations are often conflicting; efforts toward simpler models have often come at the cost of psychological richness, while efforts toward models which are rich in psychological detail have come at the cost of the ease of implementation. On the other hand, this conflict would seem to have been a useful one: by pointing out the psychological limitations of simple choice models our ability to predict choice has been systematically enhanced, and by insistence on developing theories with well-defined mathematical structures has yielded more rigorous theories of behavior.

In light of this diversity in models and modeling approaches, will these streams of thought ever converge to a "unified" mathematical theory of consumer choice? Our answer to this is both yes and no. It seems unlikely that there will ever be a mathematical theory of consumer choice which pervades work in the field, in the same way that, say, the theory of rational choice dominates modern economics. Because there will always be diversity in the objectives of modeling, there will always be diversity in the theoretical model forms which best satisfy those objectives. However, it *does* seem plausible to suggest that we will converge to agreement about a taxonomy of models which offer successive gradations of explanations of consumer choice behavior, and that we will agree to an understanding of the types of contexts in which each is most appropriate.

Given such a direction, perhaps the single greatest future research need is something resembling a meta theory of models. Although the field is replete with tools for capturing a wide range of choice policies, we currently have little understanding of why different models are appropriate in different contexts, and we have limited ability to predict which model will be appropriate a priori. As such, choice modelers are often tied into the role of being passive descriptors; we can describe past choices, but we can forecast only when there is reason to believe that decision policies will remain invariant over time and/or across contexts.

If a meta theory of choice models is to emerge, it would seem to require consideration of three separate influences:

1. statistical factors which influence the robustness of choice data to alternative specifications of choice models;
2. an understanding of how consumers alter their choice processes in response to observed changes in the appearance of the set of options; and
3. an understanding of how consumers alter their choice processes as a result of learning, satiation, or inertia over time.

The first consideration reflects findings that the robustness of a choice model to specification error often depends on the pattern of correlation which exists among attributes in a choice set (Curry and Faulds 1985; Johnson, Meyer, and Goshe 1989; Newman 1977). Specifically, when predicting choices made from sets where attributes are orthogonal or positively correlated across options, simple random utility models seem to provide a good account of the data, even when the underlying process is known to be one based on hierarchical elimination (cf. Johnson, Meyer, and Goshe 1989). In contrast, when attributes are negatively correlated (such as might arise if the consumer makes a choice from efficient sets of options, or choice sets which are screened for dominated alternatives), similar process misspecifications can dramatically hurt model performance. The implication is, therefore, that concern over fitting the "right" choice model may begin with a simple consideration of the correlational structure of the environment; the more forgiving the environment, the less one needs to worry about instability in processes.

Given that correct specification *is* a critical issue in a given modeling setting, a natural concern is whether changes in the set of options might induce changes in the way choices are made, implying a constant need to reconsider the propriety of a given estimated model form. Ideally, we would like to be in a position of predicting such effects a priori; in particular, we need models which characterize how consumers "decide how to decide."

The dominant current conjecture about how consumers formulate strategies for choice is that such strategies are the result of intuitive cost/benefit calculations. Specifically, consumers are presumed to learn the ability of differing choice strategies to yield "optimal" outcomes, and to learn the robustness of differing rules to changes in the decision environment. (See Chapter 2 by Bettman, Johnson, and Payne in this volume for a more extensive discussion.) In any given context the consumer is thus thought to select that rule which yields the highest expected outcome at the lowest cognitive effort. General normative treatments of this problem have been offered by Johnson and Payne (1985) and Shugan (1979), with normative solutions for optimal elimination strategies offered by Grether and Wilde (1984) and Huber and Klein (1989).

Empirical tests of cost/benefit models of strategy selection have yielded mixed results. On one hand, strong apparent support for cost/benefit models of strategy selection has surfaced in a number of studies of the effect of choice set size on choice rules (e.g., Johnson and Meyer 1984; Johnson, Meyer, and Goshe 1989; Payne 1976, 1983; Payne, Bettman, and Johnson 1988). If consumers choose strategies through intuitive cost/benefit calculations, increases in choice set size should be associated with increases in the use of heuristic elimination rules, owing to increases in information load. Experimental results have consistently supported this prediction: As choice set sizes increase, consumers alter information search strategies in a way consistent with a switch from compensatory to noncompensatory rules, and this is manifested in changes in the parameters of discrete choice models (cf. Johnson and Meyer 1984). In addition, there is also evidence that consumers utilize more stringent cutoff criteria when using elimination rules (resulting in a shorter decision time) as the cost of deliberation increases, and when the risk of ending up with a suboptimal final choice decreases—something again consistent with cost/benefit models (e.g., Grether and Wilde 1984; Huber and Klein 1988).

On the other hand, support for more subtle features of context, such as the pattern of correlation among attributes, affecting rule selection has been more limited. For example, Johnson, Meyer, and Goshe (1989) conjectured that if consumers choose decision rules through intuitive cost/benefit calculations, the use of noncompensatory elimination heuristics should increase as interattribute correlations become more positive (due to the increased ability of simple heuristics to mimic compensatory combination rules). In a separate study, Klein and Yadav (1989) made the same prediction about the effect of interattribute correlations, although offering a somewhat different rationale for why positive correlations should be associated with increased use of eliminations strategies: because positive correlations imply the presence of dominated options, we should observe increased use of elimination strategies as a result of consumers screening these dominant options from consideration prior to a final choice. Neither study, however, found a sensitivity of choice strategies to interattribute correlation. A similar lack of support for the effects of correlation have also been reported in a study of how consumers set cutoffs in heuristic elimination strategies reported by Huber and Klein (1988).

Another possible direction work in "deciding how to decide" might take is to link the study of rule choice to the growing literature on judgment under ambiguity or vagueness (e.g., Einhorn and Hogarth 1985; Kahn and Sarin 1990). The rationale for this link is that rule selection clearly involves uncertainty; in general, the consumer will *not* know beforehand exactly how much effort will be involved executing a particular strategy, or what the payoffs of a particular strategy will be. Unfortunately, such uncertainty is not explicitly recognized in current cost/benefit models of strategy selection. An intriguing potential future area for research would thus be to investigate how consumers make rule selections, given varying levels of uncertainty about the likelihood that a given rule is, indeed, optimal and given the cost of using different rules.

Although we are unaware of investigators who have explored this issue, Kahn and Meyer (1990) have reported on a related investigation into how consumers choose decision weights in multiattribute judgment problems under uncertainty. Their central finding is that consumers act as if they are risk averse in weight selection—uncertainty about the weight of attributes which serve to enhance an already acceptable level of utility for a product (e.g., add-ons to a television) tend to be underweighted, whereas attributes which serve to preserve acceptable levels of utility for a product (e.g., insurance) tend to be overweighted. An interesting conjecture for future research is whether risk or vagueness aversion is also manifested in selections of choice rules. One possibility is that, *ceteris paribus,* the less certain consumers are about the attributes of options in a choice set, the less likely they are to use simplified (and risky) heuristic elimination rules.

Finally, if a theory of choice models is to emerge, perhaps the most difficult problem to address is that of understanding learning, or of understanding how choice strategies change over time as consumers gain experience in a decision environment. Although we discussed work which addresses how consumers develop beliefs about the attributes of products, research on how consumers learn how to make choices is more limited. The absence of such work would seem particularly critical in new technology contexts; in such contexts, the key uncertainty facing consumers is how to integrate new technological improvements or attributes during the course of making a choice. As such, one might expect considerable variability in choice processes over time.

A recent body of research which might suggest a template for future work on the learning of choice rules is the literatures on multiple-cue learning (e.g., Castellan 1977; Klayman 1986). This literature examines the ability of individuals to learn algebraic combination rules through induction. Although we are aware of no work which has examined the learning of rules for choosing from discrete sets of options, findings in the learning of judgment rules may hold parallels. For example, Meyer (1987) found that given limited experience in a product category consumers tended to make evaluative judgments by noting how similar a given option was to his or her mental prototype of a "good" option in that product category. Early in learning judgments seem to be made in a noncompensatory (nonadditive) fashion, but over time these judgments became increasingly compensatory. An interesting area for future work would be to extend this domain of study to the acquisition of choice rules.

## Other Future Directions

Perhaps more than any other literature in consumer research, the field of choice modeling is faced with the problem of reconciling the three great disciplinary traditions which underlie the field: psychology, economics, and applied marketing. Because of this position, researchers in the area face the seemingly impossible task of developing models that possess the rich explanatory power of psychological theories of decision making and possess the axiomatic rigor of economic theories, yet are amenable to widespread application in field settings. While the development of a meta theory of the field may be the most important long-term goal, there are also a large number of critical short-term problems which focus on achieving this reconciliation among the various disciplinary traditions.

One such direction is to develop stronger links between the literature of choice models and the broader range of consumer choice application problems. One example is that of consumer planning, or sequential decisions (e.g., Urban and Hauser 1987). This is a context where heuristic models might demonstrate their greatest value; a good illustration of such an application was offered by Hayes-Roth and Hayes-Roth (1978), who described a prototype heuristic model of how individuals plan sequences of activities under time constraints. While, in theory, one could represent this same planning process as a nested string of random utility models (each characterizing an activity choice), the approach would be excessively cumbersome, particularly given multiple decisions.

The largest existing barrier to a wider use of heuristic choice models as an approach to modeling sequential decisions, however, is the difficulty of utilizing traditional inferential techniques to infer models. If the analyst has no prior hypothesis about how a consumer conducts plans, the recovery of a unique hierarchical plan requires the observation of a factorial array of contingent plans (e.g., "given that you did $A$ in time $t - 1$ and $B$ in time $t - 2$, what would you do in time $t$?"). Because this data requirement becomes excessive with increasing numbers of planned activities, such models have traditionally relied on consumers' own accounts of their planning processes (as revealed though verbal protocols) as a basis for estimation. Modeling techniques that minimize the reliance on introspective accounts would thus seem to be an important area for future development. Currim, Meyer, and Le's (1988) and Smith, Clark, and Cotton's (1984) exploration of the use of algorithms to infer heuristic choice rules illustrates one possible direction.

Other related domains of consumer decisions are those of purchase timing and quantity selection. Although choice models of this type have begun to appear (e.g., Gupta 1988), they are largely at an embryonic stage. The largest difficulty facing the formulation of such models is that consideration sets are often difficult to specify, and normative models of how such decisions *should* be made seem implausible as descriptive devices. For example, consider the deceptively simple problem of modeling the number of units of a desired good that a consumer will choose to buy when that good is on sale at a supermarket. While one could model the choice as one among several possible quantities, each having a utility level (e.g., Gupta 1988), such a model would fail to capture the rather complex thought processes which *should* underlie such a decision. Specifically, purchase quantities should be driven by an explicit consideration of how a given purchase quantity will impact later decisions, such as one's ability to stock up on an even lower price which may occur later. An important focus of future research would be to explore the development of parsimonious models which capture such considerations.

Future research would also seem usefully directed toward the problem of group choice, or choices which are driven by a negotiation process. Although there exists a large and growing literature of models of group decision making (e.g., Davis 1976; Corfman and Lehmann 1987; Webster and Wind 1972), it has tended to remain separate from that of individual choice theory. An interesting area for future work would be to assess the extent to which many of the results on individual choice, such as the effects of context, are also manifested in group decisions, and whether parsimonious means for estimating group choice models can be developed.

Finally, efforts should also be directed to ward expanding the domain of outside disciplinary traditions from which insights about choice behavior are drawn. As we have noted, the existing literature in formal choice models is dominated by work drawn from two areas: mathematical psychology and microeconomics. While one presumes that these ties will continue to be pervasive, the recent literature offers examples of how other areas can be useful for offering insights into modeling structures and methods. Illustrations include the work by Roberts and Urban (1987) and Horsky and Rabin (1988), who use theoretical models of sequential decision making developed in economics and statistics to suggest models for new product forecasting. Similarly, Greene and Smith's (1988) work on hierarchical model estimation draws its roots from inference algorithms first used to build expert systems in artificial intelligence. We presume that the future literature will provide an even richer array of examples.

## REFERENCES


AAKER, DAVID A. (1970). "A New Method for Evaluating Stochastic Models of Brand Choice," *Journal of Marketing Research,* VII, 300–306.

AMEMIYA, TAKESHI (1981). "Qualitative Response Models: A Survey," *Journal of Economic Literature,* 19 (December), 1483–1536.

ANDERSON, NORMAN H. (1981). *Foundations of Information Integration Therory.* New York: Academic Press.

BATSELL, RICHARD R., and POLKING, JOHN C. (1985). "A New Class of Market Share Models," *Marketing Science,* 4 (Summer), 177–198.

BATSELL, RICHARD R. (1982). "A Multiattribute Extension of the Luce Model Which Simultaneously Scales Utility and Substitutability," in J. Huber (Ed.), *The Effect of Item Similarity on Choice Probabilities.* Durham, NC: Fuqua School of Business, Duke University, 4–23.

BAWA, KAPIL (forthcoming). "Modeling Inertia and Variety-Seeking Tendencies in Brand Choice Behavior," *Marketing Science.*

BEN-AKIVA, MOSHE, and LERMAN, STEVEN (1985). *Discrete Choice Analysis: Theory and Application to Predicting Travel Demand.* Cambridge, MA: MIT Press.

BETTMAN, JAMES R. (1971). "Information Processing Models of Consumer Behavior," *Journal of Marketing Research,* 7 (August), 370–376.

BETTMAN, JAMES R. (1981). *An Information Processing Theory of Consumer Choice.* Reading, MA: Addison-Wesley.

BETTMAN, JAMES R. (1987). "Process of Adaptivity in Decision Making," in M.J. Houston (Ed.), *Advances in Consumer Research,* 15, 1–4.

BORGERS, ALOYS, and TIMMERMANS, HARRY. "A Context-Sensitive Model of Spatial Choice Behavior," in R. G. Golledge and H. Timmermans (Eds.), *Behavioral Modelling in Geography and Planning.* London: Croom-Helm, 159–178.

BRADLEY, RALPH A., and TERRY, M. E. (1952). "Rand Analysis of Incomplete Block Designs, I: The Method of Paired Comparisons," *Biometrika,* 39, 324–345.

BREIMAN, LEO, FRIEDMAN, JEROME, OLSHEN, RICHARD, and STONE, CHARLES (1984). *Classification and Regression Trees.* Monterey, CA: Wadsworth and Brooks.

BUSH, ROBERT R., and MOSTELLER, FREDERICK (1955). *Stochastic Models for Learning.* New York: Wiley.

CARMAN, JAMES M. (1966). "Brand Switching and Linear Learning Models,"*Journal of Advertising Research,* 6 (June), 23–31.

CARPENTER, GREGORY S., COOPER, LEE G., HANSSENS, DOMINIQUE M., and MIDGLEY, DAVID F. (1989). "Modeling Asymmetric Competition," *Marketing Science,* 7 (Fall), 393–412.

CARPENTER, GREGORY S., and LEHMANN, DONALD R. (1985). "A Model of Marketing Mix, Brand Switching, and Competition," *Journal of Marketing Research,* 22 (August), 318–329.

CARROLL, J. DOUGLAS (1972). "Individual Differences and Multidimensional Scaling," in R. N. Shepard, A. K. Rommey, and S. Nerlove (Eds.), *Multidimensional Scaling: Theory and Applications in the Social Sciences, Vol. 1: Theory.* New York: Seminar Press, 105–155.

CASTELLAN, J. JOHN (1977). "Decision Making with Multiple Probabilitistic Cues," in N. John Castellan, N. B. Pisoni, and G. R. Potts (Eds.), *Cognitive Theory,* Vol. 2. Hillsdale, NJ: Erlbaum.

CHAKRAVARTI, DIPANKAR, and LYNCH, JOHN G. (1983). "A Framework for Exploring Context Effects on Consumer Judgment and Choice," in

R. Bagozzi and A. Tybout (Eds.), *Advances in Consumer Research*, 10, 289–297.

COOPER, LEE G., and NAKANISHI, MASAO (1983a). "Standardizing Variables in Multiplicative Choice Models," *Journal of Consumer Research*, 10, (June), 96–108.

COOPER, LEE G., and NAKANISHI, MASAO (1988). *Market Share Analysis: Evaluating Competitive Marketing Effectiveness*. New York: Kluwer.

COOPER, LEE G. (1988). "Competitive Maps: The Structure Underlying Asymmetric Cross Elasticities," *Management Science*, 34 (June), 707–723.

COOPER, LEE G., and NAKANISKI, MASAO (1983b). "Two Logit Models for External Analysis of Preferences," *Psychometrika*, 48 (December), 607–620.

CORFMAN, KIM P., and LEHMANN, DONALD R. (1987). "Models of Cooperative Group Decision-Making and Relative Influence," *Journal of Consumer Research*, 14 (June), 1–13.

CORSTJENS, MARCEL L., and GAUTSCHI, DAVID A. (1983). "Formal Choice Models in Marketing," *Marketing Science*, 2 (Winter), 19–56.

CURRIM, IMRAN S. (1981). "Using Segmentation Approaches for Better Prediction and Understanding from Consumer Mode Choice Models," *Journal of Marketing Research*, 18 (August), 301–309.

CURRIM, IMRAN S. (1982). "Predictive Testing of Consumer Choice Models Not Subject to Independence of Irrelevant Alternatives," *Journal of Marketing Research*, 19 (May), 308–322.

CURRIM, IMRAN S., MEYER, ROBERT J., and LE, NHAN (1988). "Disaggregate Tree-Structured Modeling of Consumer Choice Data," *Journal of Marketing Research*, 25 (August), 253–265.

CURRY, DAVID, and FAULDS, DAVID (1985). "Indexing Product Quality: Issues, Theory, and Results," *Journal of Consumer Research*, 31(1), 134–145.

DAGANZO, CARLOS (1979). *Multinomial Probit: The Theory and Applications to Demand Forecasting*. New York: Academic Press.

DALAL, S. R., and KLEIN, R. W. (1988). "A Flexible Class of Discrete Choice Models," *Marketing Science*, 7 (Summer), 232–251.

DAVIS, HARRY L. (1976). "Decision Making within the Household," *Journal of Consumer Research*, 2 (March), 241–260.

DAWES, ROBYN M., and CORRIGAN, BERNARD (1974). "Linear Models in Decision Making," *Psychological Bulletin*, 81(2), 95–106.

DEBREU, G. (1960). "Review of R.D. Luce, Individual Choice Behavior: A Theoretical Analysis," *American Economic Review*, 50, 186–188.

DESARBO, WAYNE S., DE SOETE, GEERT, CARROLL, J. DOUGLAS, and RAMASWAMY, V. (1987). "A New Stochastic Ultrametric Unfolding Methodology for Assessing Competitive Market Structure and Deriving Market Segments," Working Paper, Wharton School of Business, University of Pennsylvania, Philadelphia, PA.

DESARBO, WAYNE S., and DE SOETE, GEERT (1984). "On the Use of Hierarchical Clustering for the Analysis of Nonsymmetric Proximities," *Journal of Consumer Research*, 11 (June), 601-610.

DOMENCICH, T., and MCFADDEN, D. (1975). *Urban Travel Demand: A Behavioral Analysis*. Amsterdam: North Holland.

DUBIN, JEFFREY. "A Nested Logit Model of Space and Water Heat System Choice," *Marketing Science*, 5 (Spring), 112–124.

DUNN, RICHARD, READER, S., and WRIGLEY, NEIL (1987). "A Non-Parametric Approach to the Incorporation of Heterogeneity into Repeated Polytomous Choice Models of Urban Shopping Behavior," *Transportation Research A*, 21, 327–345.

DUNN, RICHARD, and WRIGLEY, NEIL (1985). "Beta-Logistic Models of Urban Shopping Behavior," *Geographical Analysis*, 17, 95–113.

EAGLE, THOMAS C. (1984). "Parameter Instability in Disaggregate Retail Choice Models: Experimental Evidence," *Journal of Retailing*, 60, 101–123.

EHRENBERG, A. S. C. (1959). "The Pattern of Consumer Purchases," *Applied Statistics*, 8, 26–41.

EINHORN, HILLEL, and HOGARTH, ROBIN (1985). "Ambiguity and Uncertainty in Probabilistic Inference," *Psychological Review*, 92, 433–461.

ELROD, TERRY (1988). "Choice Map: Inferring a Product Map from Panel Data," *Marketing Science*, 7 (Winter), 21–40.

FOTHERINGHMAN, A. STEWART (1988). "Consumer Store Choice and Choice Set Definition," *Marketing Science*, 7 (Summer), 299–310.

FEINBERG, FRED, KAHN, BARBARA, and MCALISTER, LEIGH (1990). "Market Share Response when Consumers Seek Variety," Working Paper, Fuqua School of Business, Duke University, Durham, NC.

FISCHER, GREGORY, and NAGIN, DANIEL (1981). "Fixed Versus Random Coefficient Quantal Choice Models," in C. F. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications.* Cambridge, MA: MIT Press, 273–304.

FORD, GARY T., and SMITH RUTH ANN (1987). "Inferential Beliefs in Consumer Evaluations: An Assessment of Alternative Processing Strategies," *Journal of Consumer Research,* 14 (December), 363–371.

GAUDRY, M. J. I., and WILLS, MICHAEL J. (1979). "Testing the Dogit Model With Aggregate Time-Series and Cross-Sectional Travel Data," *Transportation Research B,* 13, 155–166.

GAUDRY, MARC J. I., and DAGENAIS, M. G. (1979). "The Dogit Model," *Transportation Research B,* 12, 105–111.

GENSCH, DENNIS, and SVESTKA, J. (1984). "A Maximum-Likelihood Hierarchical Disagregate Model for Predicting Choices of Individuals," *Journal of Mathematical Psychology,* 28 (June), 160–178.

GENSCH, DENNIS, and RECKER, WILFRIED W. (1979). "The Multinomial, Multiattribute Logit Choice Model," *Journal of Marketing Research,* 16 (February), 124–132.

GENSCH, DENNIS H. (1987a). "A Two-Stage Disaggregate Attribute Choice Model," *Marketing Science,* 6 (Summer), 223–240.

GENSCH, DENNIS H. (1985). "Empirically Testing a Disaggregate Choice Model for Segments," *Journal of Marketing Research,* 22 (November), 462–467.

GHOSH, AVIJIT, NESLIN, SCOTT, and SHOEMAKER, ROBERT (1984). "A Comparison of Market Share Models and Estimation Procedures," *Journal of Marketing Research,* 21 (May), 202–210.

GIVON, MOSHE, U. (1984). "Variety-Seeking Through Brand Switching," *Marketing Science,* 3 (Winter), 1–22.

GIVON, MOSHE U., and HORSKY, DAN (1978). "Market Share Models as Approximators of Aggregated Heterogeneous Brand Choice Behavior," *Management Science,* 24 (September), 1404–1416.

GIVON, MOSHE U., and HORSKY, DAN (1979). "Application of a Composite Stochastic Model of Brand Choice," *Journal of Marketing Research,* XVI (May), 158–167.

GREEN, PAUL, and WIND, YORAM (1973). *Multiattribute Decisions in Marketing: A Measurement Approach,"* Hinsdale, IL: Dryden Press.

GREENE, DAVID P., and SMITH, STEPHEN F. (1988). "A Genetic System for Learning Models of Consumer Choice," *IEEE Transactions,* 217–223.

GRETHER, and WILDE, LOUIS (1984). "An Analysis of Conjunctive Choice: Theory and Experiments," *Journal of Consumer Research,* 10 (March), 373–385.

GUADAGNI, PETER M., and LITTLE, JOHN D. C. (1983). "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science,* 2 (Summer), 203–238.

GUADAGNI, PETER M. (1983). "A Nested Logit Model of Product Choice and Purchase Incidence," in F. S. Zufryden (Ed.), *Advances and Practices of Marketing Science 1983.* Providence, RI: Institute of Management Sciences.

GUPTA, SUNIL S. (1988). "Impact of Sales Promotions on What, When, and How Much to Buy," *Journal of Marketing Research,* XXV (November).

HAUSER, JOHN, and URBAN, GLEN L. (1986). "The Value Priority Hypotheses for Consumer Budget Plans," *Journal of Consumer Research,* 12 (March), 446–462.

HAUSMANN, J., and WISE, D. (1978). "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica,* 46(2), 403–426.

HAYES-ROTH, BARBARA, and HAYES-ROTH, FREDERICK (1979). "A Cognitive Model of Planning," *Cognitive Science,* 3, 275–310.

HECKMAN, JAMES J. (1981). "Statistical Models for Discrete Panel Data," in C. F. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications.* Cambridge, MA: MIT Press, 114–178.

HENSHER, DAVID A. (1984). "Achieving Representativeness of the Observable Component of the Indirect Utility Function in Logit Choice Models: An Empirical Revelation," *Journal of Business,* 265–280.

HENSHER, DAVID A., and JOHNSON, LESTER W. (1981). *Applied Discrete-Choice Modelling.* New York: Wiley.

HOROWITZ, JOEL (1980). "The Accuracy of the Multinomial Logit as an Approximation to the Multinomial Probit Model of Travel Demand," *Transportation Research B,* 14, 331–341.

HORSKY, DAN, and RABAN, YOEL (1988). "A Bayesian Updating Model of Dynamic Brand Choice Behavior." Paper presented at the meetings of the Marketing Science Conference, Seattle, WA.

HUBER, JOEL, and KLEIN, NOREEN (1988). "Choosing Which to Choose: A Model of Cutoff Selection in a Two-Stage Decision Process," Working Paper, Fuqua School of Business, Duke University, Durham, NC.

HUBER, JOEL, and MCCANN, JOHN (1982). "The Impact of Inferential Beliefs on Product Evaluations," *Journal of Marketing Research,* 19 (August), 324–333.

HUBER, JOEL, PAYNE, JOHN W., and PUTO, CHRISTOPHER (1982). "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis," *Journal of Consumer Research,* 9 (June), 90–98.

HUBER, JOEL, and PUTO, CHRISTOPHER (1983). "Marketing Boundaries and Product Choice: Illustrating Attraction and Substitution Effects," *Journal of Consumer Research,* 10 (June), 31–44.

HUBER, JOEL, and SEWALL, MURPHY (1982). "Market Structure From Indices of Competitive Substitutability," in A. Shocker and R. Srivastava (Eds.), *Analytic Approaches to Product and Market Planning: The Second Conference.* Boston, MA: Marketing Science Institute, 7–16.

HUTCHINSON, J. WESLEY (1983). "On the Locus of Range Effects in Judgment and Choice," in R. P. Bagozzi and A. Tybout (Eds.), *Advances in Consumer Research,* 10, 305–308.

HUTCHINSON, J. WESLEY (1986). "Discrete Attribute Models of Brand Switching," *Marketing Science,* 5 (Fall), 350–371.

JEULAND, ABEL P. (1979). "Brand Choice Inertia as One Aspect of the Notion of Brand Loyalty," *Management Science,* 25 (7), 671–682.

JEULAND, ABEL P. (1978). "Brand Preferences Over Time: A Partially Deterministic Operationization of the Notion of Variety-Seeking," in Subhash Jain (Ed.), *AMA Research Frontiers in Marketing: Dialogues and Directions,* Educators' Proceedings, Series No. 43, Chicago, IL: American Marketing Association, 33–37.

JOHNSON, RICHARD D., and LEVIN, IRWIN P. (1985). "More Than Meets the Eye: The Effect of Missing Information on Purchase Evaluations," *Journal of Consumer Research,* 12 (September), 169–177.

JOHNSON, ERIC J., and MEYER, ROBERT J. (1984). "Compensatory Choice Models of Noncompensatory Choice Processes: The Effect of Varying Context," *Journal of Consumer Research,* 11 (June), 528–541.

JOHNSON, ERIC J., MEYER, ROBERT J., and GHOSE, SANJOY (1989). "When Choice Models Fail: Compensatory Models in Negatively-Correlated Environments," *Journal of Marketing Research,* 26 (August).

JOHNSON, ERIC J., and PAYNE, JOHN W. (1985). "Effort and Accuracy in Choice," *Management Science,* 30, 1213–1231.

JOHNSON, LESTER, and HENSHER, DAVID (1982), "Application of Multinomial Probit to a Two-Period Panel Data Set," *Transportation Research A,* 16 (5–6), 457–464.

JONES, J. MORGAN (1970). "A Comparison of Three Models of Brand Choice," *Journal of Marketing Research,* VII (November), 466–73.

JONES, J. MORGAN, and ZUFRYDEN, FRED (1980). "Adding Explanatory Variables to a Consumer Purchase Behavior Model: An Exploratory Study," *Journal of Marketing Research,* 17 (August), 323–334.

KAHN, BARBARA E., KALWANI, MANOHAR U., and MORRISON, DONALD G. (1986). "Measuring Variety-Seeking and Reinforcement Behaviors Using Panel Data," *Journal of Marketing Research,* XXIII (May), 89–100.

KAHN, BARBARA E., and LEHMANN, DONALD R. (1989). "Modeling Choice Among Menus," Working Paper, Anderson Graduate School of Management, University of California, Los Angeles, CA.

KAHN, BARBARA E., and LOUIE, THERESE (1990). "The Effects of Retractions of Price Promotions on Brand Choice Behavior for Variety-Seeking and Lost-Purchase-Loyal Consumers," *Journal of Marketing Research XXVII* (August).

KAHN, BARBARA E., and MEYER, ROBERT J. (1990). "A Model of Multiattribute Attribute Judgments Under Attribute Weight Uncertainty," *Journal of Consumer Research* (December 1990.)

KAHN, BARBARA E., MOORE, WILLIAM L., and GLAZER, RASHI (1987). "Experiments in Constrained Choice," *Journal of Consumer Research,* 14 (June), 96–113.

KAHN, BARBARA E., MORRISON, DONALD, and WRIGHT, GORDON P. (1986). "Aggregating Individual Purchases to the Household Level," *Marketing Science* 5 (Summer), 260–268.

KAHN, BARBARA E., and RAJU, JAGMOHAN (1989). "A Stochastic Model of the Effects of Frequency of Promotions on Brand Choice," Working Paper, Anderson Graduate School of Management, University of California, Los Angeles, CA.

KAHN, BARBARA E. and SARIN, RAKESH K. (1990). "Modeling Ambiguity in Decisions Under Uncertainty," *Journal of Consumer Research*, 15 (September), 265-273.

KAMAKURA, WAGNER A., and SRIVASTAVA, RAJENDRA K. (1985). "Predicting Choice Shares Under Conditions of Brand Interdependence," *Journal of Marketing Research*, 21 (November), 420-433.

KAMAKURA, WAGNER A., and SRIVASTAVA, RAJENDRA K. (1986). "An Ideal-Point Probabilistic Choice Model for Heterogeneous Preferences," *Marketing Science*, 5 (Summer), 199-218.

KEENEY, RALPH, and RAIFFA, HOWARD (1976). *Decisions With Multiple Objectives: Preferences and Value Tradeoffs*. New York: Wiley.

KLAYMAN, JEOSHUA (1986). "Experimentation, Observation, and Learning in Probabilistic Environments," Working Paper No. 117, Center for Decision Research, Graduate School of Business, University of Chicago, Chicago, IL.

KLEIN, NOREEN and YADAV, MANJIT S. (1989). "Context Effects on Effort and Accuracy in Choice: An Enquiry into Adaptive Decision Making," *Journal of Consumer Research*, 15 (March), 411-421.

KRISHNAN, D S., and BECKMAN, M. J. (1979). "Dynamic Discrete Choice Models, With an Application in Transportation," *Decision Sciences*, 10, 218-231.

KUEHN, A. A. (1962). "Consumer Brand Choice—A Learning Process?" *Journal of Advertising Research*, Vol. II, 10-17.

LATTIN, JAMES M. (1987). "A Model of Balanced Choice Behavior," *Marketing Science*, 6 (Winter), 48-65.

LATTIN, JAMES M., and MCALISTER, LEIGH (1985). "Using a Variety-Seeking Model to Identify Substitute and Complimentary Relationships Among Competing Products," *Journal of Marketing Research*, 22 (August), 330-339.

LEHMANN, DONALD, and MOORE, WILLIAM L. (1989). "A Combined Simply Scalable and Tree-Based Preference Model," Working Paper, Graduate School of Business, Columbia University, New York, NY.

LIPSTEIN, BENJAMIN (1959). "The Dynamics of Brand Loyalty and Brand Switching," Proceedings, Fifth Annual Conference of the Advertising Research Foundation, New York, NY.

LOUVIERE, JORDAN J., and WOODWORTH, GEORGE (1983). "Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data," *Journal of Marketing Research*, 20 (November), 350-367.

LOUVIERE, JORDAN J., and MEYER, ROBERT J. (1981). "A Composite Attitude-Behavior Model of Traveler Decision Making," *Transportation Research*, 5, 411-420.

LOUVIERE, JORDAN J. (1988). *Metric Conjoint Analysis*, New York: Sage.

LOUVIERE, JORDAN J. (1981). "On the Identification of the Functional Form of the Utility Expression and Its Relationship to Discrete Choice," in D. A. Hensher and L. Johnson (Eds.), *Applied Discrete-Choice Modeling*. London: Croom-Helm, 385-416.

LUCE, R. DUNCAN (1977). "The Choice Axiom After Twenty Years," *Journal of Mathematical Psychology*, 15, 215-233.

LUCE, R. DUNCAN (1959). *Individual Choice Behavior*. New York: Wiley.

LYNCH, JOHN G. (1985). "Uniqueness Issues in the Decompositional Modeling of Multiattribute Overall Evaluations: An Information Integration Perspective," *Journal of Marketing Research*, 22 (February), 1-19.

MALHOTRA, NARESH K. (1984). "The Use of Linear Logit Models in Marketing Research," *Journal of Marketing Research*, 21 (February), 20-31.

MANRAI, AJAY K., and SINHA, PRABHAKANT (1989). "Elimination by Cutoffs," *Marketing Science*, 8 (Spring), 133-152.

MASSY, WILLIAM F., MONTGOMERY, DAVID B., and MORRISON, DONALD G. (1970). *Stochastic Models of Buying Behavior*. Cambridge, MA: MIT Press.

MCALISTER, LEIGH (1982). "A Dynamic Attribute Satiation Model of Variety-Seeking Behavior," *Journal of Consumer Research*, 9 (September), 141-151.

MCALISTER, LEIGH, and PESSEMIER, EDGAR (1982). "Variety Seeking Behavior: An Interdisciplinary Review," *Journal of Consumer Research*, 9 (December), 311-322.

McCONNELL (1968). "Repeat-Purchasing Estimation and the Linear Learning Model," *Journal of Marketing Research,* 3 (August), 304–306.

McFADDEN, DANIEL (1986). "The Choice Theory Approach to Marketing Research," *Marketing Science,* 5 (Fall), 275–297.

McFADDEN, DANIEL (1981). "Econometric Models of Probabilistic Choice," in C. F. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications.* Cambridge, MA: MIT Press, 198–272.

McFADDEN, DANIEL (1979). "Modeling the Choice of Residential Location," *Transportation Research Record,* 673, 72–78.

MEYER, ROBERT J. (1981). "A Model of Multiattribute Judgment Under Attribute Uncertainty and Informational Constraint," *Journal of Marketing Research,* 18 (November), 428–441.

MEYER, ROBERT J. (1982). "A Descriptive Model of Consumer Information Search Behavior," *Marketing Science,* 1 (Winter), 93–121.

MEYER, ROBERT J. (1985). "On The Representation and Measurement of Consumer Choice Under Limited Information," in R. Lutz (Ed.), *Marketing Communications: Proceedings of the Winter Educators' Conference.* Chicago: American Marketing Association.

MEYER, ROBERT J., and COOPER, LEE G. (1988). "A Longitudinal Choice Analysis of Consumer Response to a Product Innovation," in R. G. Golledge and H. Timmermans (Eds.), *Behavioral Modelling in Geography and Planning.* London: Croom-Helm, 424–450.

MEYER, ROBERT J., and EAGLE, THOMAS C. (1982). "Context-Induced Parameter Instability in a Disaggregate-Stochastic Model of Store Choice," *Journal of Marketing Research,* 19 (February), 62–71.

MEYER, ROBERT J., and LOUVIERE, JORDAN J. (1981). "Multiattribute Parameter Shifting: An Approach to Modeling Context and Dominance Effects in Individual Choice Behavior," in J. Huber (Ed.), *The Effect of Item Similarity on Choice Probabilities.* Durham, NC: Fuqua School of Business, Duke University.

MEYER, ROBERT J., and SATHI, ARVIND (1985). "A Multiattribute Model of Consumer Choice During Product Learning," *Marketing Science,* 4 (Winter), 41–61.

MOORE, WILLIAM L., LEHMANN, DONALD R., and PESSEMIER, EDGAR A. (1986). "Hierarchical Representations of Market Structure and Choice Processes Through Preference Trees," *Journal of Business Research,* 14, 371–386.

MOORE, WILLIAM L., LEHMANN, DONALD R., and PESSEMIER, EDGAR A. (1989). "A Paired Comparison Nested Logit Model of Individual Preference Structures," *Journal of Marketing Research,* 26 (November), 420–428.

MORRISON, DONALD G. (1966). "Testing Brand Switching Models," *Journal of Marketing Research,* III (November), 401–409.

NAKANISHI, MASAO, and COOPER, LEE G. (1974). "Parameter Estimation for a Multiplicative Interaction Model-Least Squares Approach," *Journal of Marketing Research,* 11 (August), 303–311.

NESLIN, SCOTT A., HENDERSON, CAROLINA, and QUELCH, JOHN (1985). "Consumer Promotions and the Acceleration of Product Purchases," *Marketing Science,* 4 (Summer), 147–165.

NEWMAN, J. ROBERT (1977). "Differential Weighting in Multiattribute Utility Theory: Where it Should and Where is Does Make a Difference," *Organizational Behavior and Human Performance,* 20 (December), 312–325.

ONAKA, J., and CLARK, WILLIAM A. V. (1983). "A Disaggregate Model of Residential Mobility and Housing Choice," *Geographical Analysis,* 15, 287–304.

PAYNE, JOHN W. (1983). "Contingent Decision Behavior," *Psychological Bulletin,* 92 (September), 382–402.

PAYNE, JOHN W. (1976). "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis," *Organizational Behavior and Human Performance,* 16 (August), 366–387.

PAYNE, JOHN W., BETTMAN, JAMES R., and JOHNSON, ERIC J. (1988). "Adaptive Strategy Selection in Decision Making," *Journal of Experimental Psychology: Human Learning, Memory, and Cognition.*

PUNJ, GIRISH, and STAELIN, RICHARD (1978). "The Choice Process for Graduate Business Schools," *Journal of Marketing Research,* 25 (November), 588–598.

RAO, VITHALA R., and SABAVALA, DARIUS JAL (1981). "Inference of Hierarchical Choice Processes from Panel Data," *Journal of Consumer Research,* 8 (June), 85–96.

RATNESHWAR, SRINIVASAN, SHOCKER, ALLAN D., and STEWART, DAVID W. (1987). "Toward an Understanding of the Attraction Effect: The Implications of Product Stimulus Meaningfulness and Familiarity," *Journal of Consumer Research,* 13 (March), 520–533.

RECKER, WILLIAM W., and GOLOB, THOMAS F. (1979). "A Noncompensatory Model of Transportation Behavior Based on a Sequential Consideration of Attributes," *Transportation Research B*, 13, 269-280.

RICHARDS, ANTHONY (1982). "Search Models and Choice Set Generation." *Transportation Research A*, 16, 403-419.

ROBERTS, JOHN H., and URBAN, GLEN (1988). "Modeling Multiattribute Utility, Risk, and Belief Dynamics for New Consumer Durable Brand Choice," *Management Science*, 34 (February), 167-185.

ROTONDO, JOHN (1986). "Price as an Aspect of Choice in EBA," *Marketing Science*, 5 (Fall), 391-402.

RUSSO, JAY E., and DOSHER, BARBARA E. (1983). "Strategies for Multiattribute Binary Choice," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9 (October), 676-696.

SCHMITTLEIN, DAVID C. (1986). "Which Preference Distributions Lead to Luce's Choice Model?" Working Paper, Department of Marketing, The Wharton School of Business, University of Pennsylvania, Philadelphia, PA.

SHARMA, SUBHASH, and DURAND, RICHARD (1980). "Using the Linear Learning Model to Represent Variety-Seeking Behavior," American Marketing Association Educators' Conference Proceedings, Series NO. 46, 148-151.

SILK, ALVIN J., and URBAN, GLEN (1978). "Pre-Test Market Evaluation of New Packaged Goods: A Model and Measurement Methodology," *Journal of Marketing Research*, 15 (May), 171-191.

SRINIVASAN, V., and KESAVAN, R. (1976). "An Alternate Interpretation of the Linear Learning Model of Brand Choice," *Journal of Consumer Research*, 3 (September), 76-83.

STECKEL, JOEL H., and VANHONACKER, WILFRIED R. (1988). "A Heterogeneous Conditional Logit Model of Choice," *Journal of Business and Economic Statistics*, 6 (July), 391-398.

SUGITO, YOSHIHIRO (1986). *A Dynamic Probit Model of Brand Choice: Model Development and Application to Consumer Dealing*, Unpublished Ph.D dissertation, Anderson Graduate School of Management, University of California, Los Angeles, CA.

TELSER, LESTER G. (1962). "The Demand for Consumer Goods as Estimated from Consumer Panel Data," *Review of Economics and Statistics*, 44, 471-499.

THURSTONE, L. L. (1927). "A Law of Comparative Judgment," *Psychological Review*, 34, 273-286.

THURSTONE, L. L. (1959). *The Measurement of Values*. Chicago: University of Chicago Press.

TRAIN, KENNETH (1986). *Qualitative Choice Analysis*. Cambridge, MA: MIT Press.

TVERSKY, AMOS, and SATTATH, SHUMEL (1979). "Preference Trees," *Psychological Review*, 86 (November) 542-573.

TVERSKY, AMOS (1972). "Elimination By Aspects: A Theory of Choice," *Psychological Review*, 79 (4), 281-299.

WEBSTER, FREDERICK E., and WIND, YORHAM (1972). "A General Model of Organizational Buying Behavior," *Journal of Marketing*, 36 (April), 12-19.

WINER, RUSSELL S. (1986). "A Reference Price Model of Brand Choice for Frequently Purchased Products," *Journal of Consumer Research*, 13 (September), 250-256.

WRIGLEY, NEIL, LONGLEY, PAUL, and DUNN, RICHARD (1984). "Some Recent Developments in the Specification, Estimation, and Testing of Discrete Choice Models," in *Categorical Data Analysis for Geographers and Environmental Scientists*. London: Longman.

WRIGLEY, NEIL, LONGLEY, PAUL, and DUNN, RICHARD (1988). "Some Recent Developments in the Specification, Estimation, and Testing of Discrete Choice Models," in R. G. Golledge and H. Timmermans (Eds.), *Behavioral Modelling in Geography and Planning*. London: Croom-Helm, 96-123.

YAMAGISHI, TOSHIO, and HILL, CHARLES T. (1983). "Initial Impression Versus Missing Information as Explanations for the Set-Size Effect," *Journal of Personality and Social Psychological*, 44 (5), 942-951.

YELLOTT, J. I. (1977). "The Relationship Between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the Double Exponential Distribution," *Journal of Mathematical Psychology*, 15, 109-144.

ZUFRYDEN, FRED S. (1980). "A Multivariate Stochastic Model of Brand Choice and Market Behavior," in J. Sheth, (Ed.), *Research in Marketing*, Vol. 3. Greenwich, CT: JAI Press, 273-303.

ZUFRYDEN, FRED S. (1981). "A Logit-Markovian Model of Consumer Purchase Behavior Based on Explanatory Variables," *Decision Sciences*, 12, (October), 645-660.

ZUFRYDEN, FRED S. (1986). "Multibrand Transition Probabilities as a Function of Explanatory Variables: Estimation by a Least-Squares Based Approach," *Journal of Marketing Research*, XXIII (May), 177-183.