Discussion Paper No. 1146

History Dependent Brand Switching: Theory and Evidence*

by

Itzhak Gilboa+

and

Amit Pazgal++

November 1995

*We are grateful to Dipak Jain and David Schmeidler for many discussions, comments, and references.

+ KGSM-MEDS, Northwestern University. Currently visiting the Department of Economics, University of Pennsylvania.

++ KGSM-MEDS, Northwestern University.

Abstract

We present a model of brand-switching in which a consumer's impression of each brand is based on her memory of past consumption of this brand, and is stochastically updated whenever the brand is consumed. In the ordinal version of the model, consumer's memory is an ordering of the available brands. The top brand is chosen and consumed, and may therefore move to a different ranking. In the cardinal version, the consumer remembers a "cumulative utility index" per brand, and, when a brand is consumed, the index is updated by the addition of a random variable, interpreted as "instantaneous utility." In both versions of the model it may be assumed that the consumer may sometimes be "dormant," choosing the same brand out of inertia, or that she is always "active," re-evaluating her decision based on her cumulative memory.

We prove that, in all versions, the frequencies of choice converge, with probability 1, to limit frequencies which can be computed from the model's parameters. We also show that, under mild assumptions, every sequence of choices would have a positive probability.

We test the ordinal model empirically, using scanner data on purchases of crackers, yogurts, and catsups. We show that both the "order effect" and the "inertia effect" exist. Specifically, the ordinal model performs significantly better than its restricted version, in which only the last brand is recalled. Similarly, the model performs significantly better with the inertia assumption than without it.

1. Introduction

In recent years, marketing researchers have presented considerable empirical evidence that the fundamental nature of frequently-purchased, low priced products is one of frequent switching among different brands in the same product class. This phenomenon is known as "brand switching" or "multi brand buying." While different consumer choices at different times may be explained by changing tastes, the high frequency with which consumers change their purchasing decisions calls for a more parsimonious theory, explaining patterns of choice with relatively few parameters.

One approach to the problem of statistically describing brand switching is known as "random utility" models. According to these models, the consumer observes the realization of a random variable, interpreted as "utility," for each available brand, and then chooses a brand whose utility is maximal. Specifically, for every alternative a there is a random variable $u_a = v_a + \varepsilon_a$, where v_a is a deterministic component which may depend on specific brand and household attributes, and ε_a is a random component varying from one choice occasion to the next. Typically, the ε_a's are assumed to be independent of previous periods' realizations. Some models also assume independence across brands, while others allow correlations to exist. For instance, multinomial Logit and Probit models fall in this category. For a detailed analysis see McFadden (1981) and Ben-Akiva and Lerman (1985).

However, the data typically suggest that consumer choices are history-dependent. Specifically, one may identify such patterns as "brand loyalty" and "variety seeking" behavior. (For an extensive review of the literature on individual consumer behavior see McAlister and Pessemier (1982).) Indeed, these phenomena were introduced into stochastic models by Jeuland (1979) (brand loyalty) and by McAlister (1982) and Trivadi, Bass, and Rao (1994) (variety seeking behavior). Furthermore, Guadagni and Little (1983) propose a model in which choice probabilities depend on the whole history of consumer choices. More recently, Chintagunta, Halder, and Roy

(1994) presented a model incorporating state dependence, habit persistence, and consumer heterogeneity.[1]

The present paper suggests a history-dependent random utility model, attempting to take into account consumer's past choices, as well as the randomness in consumer's satisfaction with the brands. It differs from classical random utility models in a few ways: first, we assume that the consumer's choice is determined by the utility derived in the past from *actual* consumption, rather than by perceived utility from hypothetical consumption in the future. That is, in contrast to the random utility models, in which the consumer first observes realizations of utility variables for the current period, and then makes a choice, our model suggests that the utility values that matter at the time of decision are the past realizations. In this sense, our model is *retrospective*, rather than *prospective*. Correspondingly, while the classical random utility models use, in each period, as many random variables as there are brands, our model presupposes only one random variable per period. Interestingly, this reduction in the number of "hidden" variables does not restrict the class of consumption patterns our model may give rise to with positive probability.

Second, our model is cumulative in nature. It is implicitly assumed that the consumer's choice depends on her "cumulative impression" of each brand. The consumer's memory is not assumed to be very large; in fact, we assume that the consumer recalls only summary statistics of past consumption, and their number does not depend on the number of past consumption periods. However, the consumer is assumed to recall a *ranking* of brands. Thus, the effect of a particular consumption period need not be bounded in time; the consumer may "remember" certain facts forever. In particular, it is possible that the consumer tries a certain brand once, and

---

1 Their model, like those of Carpenter and Lehmann (1985) and Chintagunta, Jain, and Vilcassim (1991, 1994), explicitly incorporates explanatory factors such as brand attributes, prices, advertising campaigns, special displays, and brand mix. Yet the focus of the present paper is the dynamic pattern of choice, and, as in standard random utility models, we will assume that all other marketing factors are encapsulated in the choice distribution.

never wishes to consume it again, due to the "traumatic" memory of the first time. We refer to the effect of brands' ranking on current choice as the "order effect."

The third and final main feature of our model is that the consumer is not assumed to make a conscious decision at each and every consumption period. Rather, we suggest that the consumer may be in one of two modes: a "dormant," or "inert" mode, in which she keeps purchasing and consuming her previous choice without re-assessing it; and an "active," or "conscious" mode, in which she actually contemplates her decision. We assume that the transition between these modes is governed by a stochastic process. For instance, with a certain probability one brand may be featured, or be on sale. In this case, a consumer who chooses a certain brand out of inertia might come to re-assess her choice. It is likely to expect her to become active and choose the featured brand. However, she might also become active and choose a different brand. We do not model the specific marketing factors here; instead, we assume that all relevant information is encapsulated in the (brand-dependent) distributions of the dormancy period's length.

We refer to the possibility of "dormant" states as the "inertia effect." Note that it is independent of the previous two assumptions. That is, one may introduce inertia (in the form of dormant and active states) into other models of brand switching as well. Similarly, our basic model has two versions, namely, with and without the inertia assumption.

The specific formulation of the model outlined above calls for a modeling choice: should the consumer's memory be *ordinal*, retaining only qualitative pairwise comparisons of the different brands, or should it be *cardinal*, quantifying these comparisons? The ordinal model puts relatively little demands on the consumer's memory: only a general, and potentially rather vague impression is assumed to be recorded, such as "brand *a* was overall better than brand *b*," without quantifying this preference. For the same reason, this model does not allow the consumer to distinguish between, say, a weak preference for a "slightly better" brand, and a strong preference for a "good" brand over an "unacceptable" one. By contrast, the cardinal model makes these distinctions, but requires a real number to be recalled for each brand.

We find both models somewhat extreme. It is likely that consumers can have weak preferences alongside strong preferences, while not being able to perfectly quantify all brand rankings. On the bright side, both models capture some features of consumer's memory. We therefore develop in this paper both the ordinal and the cardinal variants of the model. We provide the theoretical results for both of them, and study the ordinal one empirically. (An empirical study of the cardinal model involves computational difficulties we have yet to overcome.)

It might be useful to contrast the ordinal and cardinal models in simple examples of "brand loyalty" and of "variety seeking behavior." For simplicity, consider the case of a consumer who is always active in both models. In the ordinal model, a consumer's memory may be in one of $n!$ states, if there are n brands: a state of memory is simply a permutation of the brands, which is interpreted as an ordering, from the most preferred to the least preferred one. The first (top) brand is the one that the consumer will choose. As a result of consuming this brand, the impression the consumer has thereof might change; correspondingly, the consumer's memory might move to a new state, in which the consumed brand is ranked somewhere among the others, while the relative ranking of any two other brands does not change. We will assume that the transition between these memory states is governed by a Markov chain.

If the transition probability from a memory state to itself is one, the best brand according to this state remains the best one as a result of its consumption. Hence the consumer would appear to be "satisficed" with this brand, and maybe even to form a "habit" of consuming it. By contrast, if the consumer's memory always moves to a different state, the consumer will never choose the same brand two consecutive times.

Next consider the cardinal model. It assumes that, at each period, each brand has a "cumulative satisfaction index." A brand which has the maximal index level gets to be chosen, and, as a result of its consumption, its index might change. Specifically, we assume that the index is modified additively: the "instantaneous utility" derived from the brand's consumption is added to the previous value of the index, while the index value of the other brands remains unchanged.

Should a certain brand have a positive expected instantaneous utility, the consumer will, on average, tend to consume it more, the more she has consumed it in the past. We might therefore observe brand loyalty in this case. On the other hand, a negative expected instantaneous utility implies that, on average, the consumer gets more dissatisfied with the brand, the more she has consumed it. If this is the case with all brands, the consumer will exhibit variety seeking behavior.

The cardinal model in which the consumer is always active may be viewed as a stochastic version of "case-based consumer theory" proposed by Gilboa and Schmeidler (1993). As in their paper, the zero level on the instantaneous utility scale may be interpreted as the consumer's "aspiration level": when it is exceeded, the consumer is "satisficed" and tends to choose the same brand again; when it is not obtained, the consumer is "dissatisficed," and tends to alter her choice.

Note, however, that despite the use of Simon's (1957) term "satisficing," in our model, as well as in Gilboa-Schmeidler's, no bounded rationality need be assumed. A "satisficed" consumer may be dubbed "irrational" to the extent that she never tries certain brands, and does not even attempt to maximize the instantaneous utility. However, the consumer's "utility" should be interpreted as the cumulative, not the instantaneous one. Thus, especially if the consumer has had the chance to try out all alternatives, both variety seeking and habit formation are modes of dynamic choice consistent with the basic tenets of "rational" consumer theory.

The rest of this paper is organized as follows. In Section 2 we present the formal ordinal model. We prove that, if all legitimate transition probabilities are positive, any finite sequence of choices has a positive probability according to this model. We also mention that the Markov chain dictates limit frequencies of choice for the various brands.

In Section 3 we spell out the formal cardinal model and prove the corresponding results in this framework. That is, we show that, under mild assumptions on the distributions of instantaneous utility, limit choice frequencies will exist almost surely, and they can be directly computed from the model's parameters. Further, we show that, if these distributions have full support, the cardinal model will also assign a positive probability to any choice sequence. Both Section 2 and 3 deal with potentially-inert consumers as well as with always-active ones.

Section 4 tests the ordinal model empirically. We show that it performs at least as well as, and sometimes better than existing models of similar complexity. Specifically, we test the ordinal model versus the first-order Markov one in a three brand case. We also test the inertia effect for the simplest two-brand choice problem. These tests support the existence of both order and inertia effects. Finally, Section 5 contains the proofs, while Section 6 concludes.

2. The Ordinal Model

Consider an individual consumer confronted with a repeated choice from a set $A = \{1, \cdots, n\}$ of alternatives. In our setting the alternatives will represent different brands of the same product class. We are interested in the sequence of choices (elements of *A*) made by our consumer at times: $t = 1,2,\ldots$.

Let $S = \{s: A \to A \mid s \text{ is a permutation}\}$ be the set of states. A state $s \in S$ is interpreted as follows: $s(1)$ is the "best" product in the consumer's eyes, and $s^{-1}(a) < s^{-1}(b)$ means that the impression the consumer has of alternative *a* is better than that she has of *b*. Correspondingly, we interpret $s^{-1}(a) < s^{-1}(b)$ behaviorally, i.e., as implying that, given the choice, the consumer would consume *a* rather than *b*.

We assume that, at state $s \in S$, the consumer chooses the alternative $a = s(1)$. As a result, she may change her impression of *a*, but not of other alternatives. Hence from state s the consumer will move to a state t, according to which all alternatives $b, c \neq a$ are ranked as in s. In other words, the dynamic path will move along arcs in:

$$E = \left\{ (s,t) \in S \times S \,\middle|\, \begin{array}{l} \forall b,c \in A \text{ such that } s^{-1}(b) \neq 1 \text{ and } s^{-1}(c) \neq 1 \\ s^{-1}(b) < s^{-1}(c) \text{ iff } t^{-1}(b) < t^{-1}(c) \end{array} \right\}$$

Further, we assume a Markov process with state space S and a transition matrix q, such that $q(s,t) = 0$ whenever $(s,t) \notin E$. In particular, $q(s,s)$ is the probability that, if $a = s(1)$ is the best alternative according to s, when it is consumed, it would remain so. If, for instance, $q(s,s) = 1$, the consumer is "satisficed" with $a = s(1)$ and would keep choosing it forever (extreme loyalty). If, on the other hand, $q(s,s) = 0 \;\forall s \in S$, the consumer is always "dissatisficed," and will switch her choice at all periods. (An extreme example of variety seeking.)

We note the following:

Observation 2.1: If $q(s,t) > 0 \;\; \forall (s,t) \in E$, then the Markov chain is irreducible. (Hence, it has a unique stationary distribution.)

(All proofs are relegated to Section 5.)

Observe that, out of $(n!)^2$ pairs of states, only $n \cdot n!$ are in E. This might raise a doubt, whether the model is rich enough. Specifically, are there choice patterns that require the use of arcs outside of E? The following result answers this question in the negative. Namely, any observed sequence of choices can be explained by our model (with positive probability), given an appropriate choice of parameters.

Theorem 2.2: Assume that $q(s,t) > 0 \;\; \forall (s,t) \in E$. Then for every $T \geq 0$ and every sequence $x = (x_1, \cdots, x_T) \in A^T$ there is a sequence of states $(s_0, s_1, \cdots, s_T) \in S^{T+1}$ such that:

(i) $\quad q(s_\tau, s_{\tau+1}) > 0 \qquad \forall\; 0 \leq \tau \leq T-1$

(ii) $\quad s_\tau(1) = x_{\tau+1} \qquad \forall\; 0 \leq \tau \leq T-1$

That is, there is an initial state $s_0 \in S$ and a positive-probability path in *S*, which dictates the choice sequence x.

We now turn to extend the model to incorporate inertia. There are several ways in which this can be done. For instance, one might consider a state space $\widetilde{S} = S \times (\mathrm{N} \cup \{0\})$, where (s,n) is interpreted as implying that the consumer's memory is in state s, and that she has been dormant for the last n periods. Such a model would allow the transition probabilities to depend on the length of the dormancy period. By contrast, one may assume that the next state is independent of the dormancy period, and employ a more concise model. In particular, for each $s \in S$, let $\underset{\sim}{Q}^S$ be a distribution over the natural numbers, with finite mean and variance. Assume that, once the consumer is at state s, she will move to the next state according to the transition probabilities q as above, but that she will do so only after a number of periods Y which is a random variable with distribution $\underset{\sim}{Q}^S$.

We observe that the counterparts of Observation 2.1 and Theorem 2.2 hold in this model. Specifically, the limit relative frequencies of the brand choices can be computed from the Markov chain stationary distribution and the expectations of the distributions $\underset{\sim}{Q}^S$. Also, with positive transition probabilities, every sequence has a positive probability of being selected. (Note, however, that if $q(s,s) > 0$ for some states s, there will typically be more than one way to obtain it from the Markov process and the associated dormancy times.)

One might argue that allowing the distributions $\underset{\sim}{Q}^S$ to vary across states compromises the ordinal nature of the model. Indeed, the expected time during which a consumer appears to be "dormant" while consuming a certain brand might be viewed as a numerical measure of the consumer's satisfaction with this brand. But this interpretation is not the only one possible. For instance, a consumer may become active only if there is an increase in the price of the brand she chooses. Assuming a certain probability for price changes in each brand, the distribution of dormancy time becomes dependent on the chosen brand, $s(1)$, and thus on the state s. Yet, the consumer is only assumed to remember the ordering of the brands, whereas the distributions Q^S describe a process which is external to memory.

3. The Cardinal Model

We will now extend the model to deal with a cardinal memory. A consumer chooses from the set $A = \{1,2,\cdots,n\}$ of alternatives. We suppose that if she chooses alternative $a \in A$, she will get an instantaneous "utility" of $\tilde{u}(a)$, where $\tilde{u}(a)$ is a random variable distributed according to G_a, with mean $\mu_a < 0$ and variance $\sigma_a^2 < \infty$.

Let the history at time t, H_t, be a sequence of pairs, denoting the choices the consumer made in the past and her satisfaction with them. That is, $H_t = ((x_\tau, r_\tau))_{\tau=1}^t$, where $x_\tau \in A$ is the alternative chosen at time τ, and r_τ is the instantaneous utility realization.

Should the consumer be active at stage t, she would evaluate every alternative $a \in A$ by the functional:

$$U(a,t) = \sum_{\{\tau \le t: x_\tau = a\}} r_t .$$

Two comments regarding the cognitive interpretation of U-maximization are in order. First, the consumer is not assumed to consciously calculate U. Rather, this functional attempts to capture the consumer's "overall satisfaction" with the various brands. Second, that calculation of U requires to retain only n values. Specifically,

$$U(a,t) = \begin{cases} U(a,t-1) + r_t & \text{if} \quad x_t = a \\ U(a,t) & \text{otherwise} \end{cases}$$

Thus, the U-maximization model does not assume that the consumer's memory is unbounded.

Let I_τ be an indicator function, indicating whether the consumer is active at stage τ. That is, $I_\tau = 1$ indicates that the consumer chooses a *U*-maximizing alternative. On the other hand, if $I_\tau = 0$, the consumer simply chooses the previous alternative again $(x_\tau = x_{\tau-1})$.

A state of the world ω is a sequence of triples of the form $(x_t^\omega, r_t^\omega, I_t^\omega)$, where $x_t^\omega \in A$ denotes the consumer's choice at time *t*, $r_t^\omega \in \Re$ denotes its utility realization, and $I_t^\omega \in \{0,1\}$ indicates whether the choice was made in a "conscious" manner. A state ω can therefore be written as $\omega = ((x_1^\omega, r_1^\omega, I_1^\omega),(x_2^\omega, r_2^\omega, I_2^\omega),\cdots)$.

The set of all states of the world is $\Omega \equiv (A \times \Re \times \{0,1\})^N$. We endow Ω with the σ-algebra, Σ, generated by the random variables (x_t, r_t, I_t) (where r_t is Borel measurable.)

For a state ω, alternative *a*, and time $t \geq 1$, define the number of appearances of *a* in the sequence ω up to stage *t* to be:

$$F(\omega,a,t) = \#\{1 \leq \tau \leq t \mid x_\tau^\omega = a\}$$

and let $U(\omega,a,t)$ denote the *U*-value of brand $a \in A$ at that stage, according to the history dictated by ω. That is, $U(\omega,a,t) = \sum_{\tau \leq t} r_\tau^\omega \cdot 1_{\{x_\tau^\omega = a\}}$.

Of special interest will be the relative frequencies of the brands chosen, denoted by

$$f(\omega,a,t) = \frac{F(\omega,a,t)}{t}$$

and their limit

$$f(\omega,a)=\lim_{t\to\infty}\frac{F(\omega,a,t)}{t}.$$

(We will use this notation even if the limit is not guaranteed to exist.)

Denote the vector of limit frequencies by $f(\omega)=(f(\omega,1),\cdots,f(\omega,n))$.

Let us first consider the case where the consumer is active at every stage ($I_\tau=1$ with probability 1). We define a probability measure Pr on (Ω,Σ) that will reflect the consumer choices as well as the distribution of the random variables I,r. At the initial stage the consumer will randomize uniformly among all her alternatives:

$$\Pr[x_1=a]=\frac{1}{n} \qquad \text{for all } a\in A.$$

At stage $t>1$, let $C(t)$ be the set of alternatives with the highest *U*-value:

$$C(t)=\{a\in A\; U(a,t)\geq U(b,t) \quad \forall b\in A\}.$$

The consumer's choice is assumed to be uniform over this set:

$$\Pr[a_{t+1}=a]=\frac{1}{|C(t)|} \qquad \text{for all } a\in C(t).$$

To complete the definition of the probability measure we set $\Pr[I_t=1]$[2] i.e., the consumer is active at every period, and we assume that Pr agrees with the distributions $\{G_a\}_a$ with respect to the variables $\{x_t\}$.

---

[2] For the statement of our first result, I_t may be dropped from the definition of Ω. Yet we find it more convenient to define a single measure space (Ω,Σ).

Our first result states that, with Pr measure 1, limit frequencies exist, and they are inversely proportional to the expectations μ. That is, given μ_i $(\mu_i < 0)$, define

$$\left(f_\mu\right)_a = \frac{\prod_{b \neq a} \mu_b}{\sum_{c \in A} \prod_{b \neq c} \mu_b} .$$

Alternatively, $f_\mu \in \Re^n$ is defined by: (i) $\frac{\left(f_\mu\right)_a}{\left(f_\mu\right)_b} = \frac{\mu_b}{\mu_a}$;

(ii) $\sum_{a \in A} \left(f_\mu\right)_a = 1$.

We can now state:

Theorem 3.1: $\Pr\left[\left\{\omega \in \Omega \quad \exists f(\omega) = f_\mu\right\}\right] = 1$.

Theorem 3.1 suggests an interpretation of the instantaneous utility that the consumer derives from the brands. Consider two alternatives $a,b \in A$. The relative frequencies with which they will be consumed are in inverse proportion to their mean utility levels, regardless of the specific utility distribution:

$$\frac{f(x,a)}{f(x,b)} = \frac{\mu_b}{\mu_a} .$$

For example, if $\mu_a = -2$ and $\mu_b = -4$, brand a will be consumed twice as frequently as brand b. If these are the only two brands, their consumption frequencies will be $\frac{2}{3}$ and $\frac{1}{3}$, respectively.

This example illustrates the difference between our "instantaneous utility" and the neo-classical notion of utility. In our model, even if there is no randomness in $\tilde{u}$, $\mu_b < \mu_a$ does not imply that brand b will never be chosen. It will be selected less often than brand *a*, but if $\mu_a < 0$, $U(a)$ will eventually be lower than $U(b)$. Correspondingly, the ratios μ_a/μ_b determine the ratios of choice frequencies. Since the former can also be computed from the latter, $\{\mu_a\}_a$ are (asymptotically) observable up to a multiplicative factor.

Let us turn to the more general case where the consumer need not be active at every stage. In this case, if the consumer is active at time *t*, she may become dormant for L_t periods, where L_t is a random variable. Specifically, for $\omega \in \Omega$, let

$$L_t^{\omega} = \min\left\{k > 0 \,\middle|\, I_{t-k}^{\omega} = 1\right\}.$$

Thus, L_t is a random variable taking values in $\{1,2,\cdots\}\cup\{\infty\}$. We assume that, if $I_t = 1$, and $x_t = a$, L_t follows a distribution Q^a, with a finite mean, $\upsilon_a \geq 1$, and finite variance.

Given $Q = (Q^1,\ldots,Q^n)$ let $\Pr_Q$ be the probability measure on (Ω,Σ) that reflects the consumer choices as well as the distribution of the random variables r, L. Define

$$\left(f_{\mu\cdot\upsilon}\right)_a = \frac{\left(f_\mu\right)_a \cdot \upsilon_a}{\sum_{b\in A}\left(f_\mu\right)_b \cdot \upsilon_b} \qquad \text{for every } a \in A.$$

Theorem 3.2: $\Pr_Q\left[\left\{\omega \in \Omega \,\middle|\, \exists f(\omega) = f_{\mu\cdot\upsilon}\right\}\right] = 1$.

In other words, if the consumer follows the specified decision rule, the relative frequencies with which she will consume any two brands is inversely proportional to their average utility levels times their average inertia times. In particular, if the inertia period distribution is independent of the brand chosen, we get the same limit frequencies as in Theorem 3.1:

Corollary 3.3: If $Q^a = Q^b$ for every $a, b \in A$ then $\Pr_Q\left[\left\{\omega \in \Omega \mid \exists f(\omega) = f_\mu\right\}\right] = 1$.

Next we turn to the cardinal counterpart of Theorem 2.2. That is, we address the question, which choice sequences have a positive probability according to this model? The answer will obviously depend on the distribution of the instantaneous payoffs. For instance, if each brand has a positive probability to yield zero payoff, any sequence will have a positive probability. However, we will be interested in more robust results. In particular, let us consider the case in which each distribution is continuous, with full support.

It turns out that, if the consumer starts out with zero initial *U*-values, not every sequence of choices will have a positive probability. Consider the following example: the consumer has three alternatives $A = \{a, b, c\}$. If she starts with zero initial *U*-values, a pattern of choices like a, b, a will *not* have a positive probability with continuous instantaneous utility distributions. Indeed, if b was chosen after a , a's first utility realization must have been negative. For the same reason, so was b's realization. Therefore c should have been chosen at the third stage, rather than a again. However, there is no reason to assume that all initial *U*-values are zero. On the contrary, observing the consumer's choices starting at an arbitrary period, we are likely to find effects of past consumption. For instance, if brand c in the example above has a low initial *U*-value, the pattern a, b, a may be a result of *U*-maximization. The question we address is, therefore, given a sequence of brands, are there initial *U*-values, $U(\cdot, 0)$, and realizations r_t, such that sequence follows from strict *U*-maximization ? A naïve calculation of degrees of freedom is not very encouraging: for a given brand to be optimal at a given period, we need $(n-1)$ inequalities to hold. Thus we have $T \cdot (n-1)$ inequalities, but only $n + (T-1)$ parameters to choose (n initial *U*-values and the realizations $(r_t)_{t=1}^{T-1}$). Yet, the following shows that such parameters do exist.

Theorem 3.4: For every sequence $(x_1,...x_t) \in A^t$ of choices there are n initial values $U(i,0)$, and a sequence of realizations r_τ for $1 \le \tau \le t-1$, such that the sequence is determined by U-maximization with strict inequalities throughout. (That is, such that x_τ is the unique U-maximizer at stage τ, for every $1 \le \tau \le t$.)

Since the theorem guarantees strict U-maximization, there is an open neighborhood of $(r_\tau)_{\tau=1}^t$ for which U-maximization dictates the choices $(x_\tau)_{\tau=1}^t$. Hence, if G_a has full support for every $a \in A$, the sequence x has a positive probability. Finally, we observe that a similar result holds also in the case that the consumer might be dormant.

4. Empirical Evidence

For the empirical analysis, we used data from three panels of households. The data for the first panel were collected by a marketing research firm, Information Research Incorporated (IRI), and pertained to the purchases of saltine crackers in the Rome, Georgia, market. The second and third data sets consisted of purchases of catsups and yogurts by a panel of households in Springfield, Missouri. These data sets were provided by A. C. Nielsen. All three are optical scanner panel data sets and contain information on all purchases made by the same household over the data collection period (about two years). Each panelist is provided with an identification card that is presented at the checkout counter at the time of the purchase. (The consumers have an incentive to present the card at each purchase.) Purchases are scanned and recorded under the consumer's identification number. The data provided constitute a reasonably complete record of the households' purchases over time as all participants indicated that they shop at least 90% of the time in the participating retailers.

Since we focus on the temporal aspect of a single consumer's choice, long choice sequences provide a more reliable test of our model than do short ones. Thus, the data were screened to eliminate households that did not make at least <u>fifteen</u> purchases. (This approach is often used for analyzing choice data. For instance, Bawa (1990) restricts attention only to households with more than 24 purchases).

<u>Crackers</u>

The data on the saltine crackers consist of 125 households with 3138 purchases. There are three major national brands — Sunshine, Keebler, and Nabisco with market shares of 7.4%, 6.8%, and 54.1%, respectively. The other 31.7% of the market are distributed among several local brands that we grouped together under "private label" in the analysis. For all of the four brands the 16-once package is the one analyzed.

<u>Yogurt</u>

The yogurt data include of four brands — Yoplait (6 oz.) , Dannon (8 oz.), Weight Watchers (6 oz.), and Hiland (8 oz.) with market shares of 32.6%, 39.8%, 25.5%, and 2.1% respectively. We analyze 51 households that made 1871 combined purchases.

<u>Catsup</u>

The catsup data set consists of 38 households making 771 purchases. Heinz is the major brand with three different sizes, Heinz 40, Heinz 32, and Heinz 28, with market shares of 2.9%, 71.1%, and 19.7% respectively. The other 6.3% are captured by Hunt's 32. We treat those products as four different brands in the analysis of the catsup product category.

Due to computational difficulties, we tested only restricted versions of our model. Thus, we do not attempt to estimate a "universal" model that would encompass all the features discussed above. Rather, our goal here is to empirically validate the order and inertia effects, while their interaction calls for further analysis.

4.1. Test of the Order Effect

We start by testing the model of Section 2, without the inertia assumption, i.e., assuming that the consumer is active at all periods. For the case of two brands, say *a* and *b*, this model has two states, corresponding to the orderings *ab* and *ba*. Thus, it is equivalent to the well known "first-order Markov process" (See Jeuland (1979)). In this model, there is a state corresponding to each brand, and transition probabilities that govern the consumer's current choice, given her most recent one. In the presence of only two brands, a state *a* in the "first order" model corresponds to the state *ab* in ours. Therefore, we need at least three brands to distinguish between the models.

Assume, then, that $A = \{a, b, c\}$. The first order Markov model would include three states, one per brand. In our model, by contrast, there are six states:

$$S = \{abc, acb, bac, bca, cab, cba\}$$

Note that not all transition probabilities may be positive. For n brands, out of the $(n!)^2$ pairs of states, only $n \cdot n!$ are in the set *E*. In the case $n = 3$, 18 of the 36 pairs are in *E*, and 18 are not. A schema of the transition matrix is depicted in table 1.

Table 1:

Destination State						
Origin Sate	*abc*	*acb*	*bac*	*bca*	*cab*	*cba*
abc	+	0	+	+	0	0
acb	0	+	0	0	+	+
bac	+	+	+	0	0	0
bca	0	0	0	+	+	+
cab	+	+	0	0	+	0
cba	0	0	+	+	0	+

"+" denotes potentially positive entries. The entries in each row must sum to 1.

Observe that the first order Markov model is a special case of ours. Indeed, given Markov transition probabilities $(p_{i,j})_{i,j\in A}$, one may compute the corresponding probabilities $(q_{s,t})_{s,t\in S}$ in our model. For instance, $q_{(abc),(abc)}$ is the probability that a consumer, who at time t has memory *abc* will retain it at time $t+1$. This is obviously $p_{a,a}$, i.e., the probability that *a*, once consumed, would remain the top choice. For a more interesting case, let us consider $q_{(abc),(bac)}$. This is the probability that *a* will not be consumed next, i.e., $(1-p_{a,a})$, and furthermore, that it will be consumed again before *c*. (Note that, if *a* is not consumed next, brand *b* is the next choice by definition of the memory state (abc).) According to the first order Markov model, once the consumer chooses *b*, *a* will be chosen before *c* with probability $\frac{p_{b,a}}{p_{b,a}+p_{b,c}}$, and *c* before *a* with probability $\frac{p_{b,c}}{p_{b,a}+p_{b,c}}$. Thus,

$$q_{(abc),(bac)} = (1-p_{a,a})\cdot\frac{p_{b,a}}{p_{b,a}+p_{b,c}}$$

$$q_{(abc),(bca)} = (1-p_{a,a})\cdot\frac{p_{b,c}}{p_{b,a}+p_{b,c}}$$

The rest of the probabilities $(q_{s,t})$ are computed similarly.

It is easy to see that a general transition matrix $(q_{s,t})$ is derived from the first order Markov model iff:

(i) $$q_{(abc),(abc)} = q_{(acb),(acb)}$$

(ii) $$q_{(bac),(bac)} = q_{(bca),(bca)}$$

(iii) $$q_{(cab),(cab)} = q_{(cba),(cba)}$$

(iv) $$\frac{q_{(abc),(bac)}}{q_{(abc),(bca)}} = \frac{q_{(cba),(bac)}}{q_{(cba),(bca)}}$$

(v) $$\frac{q_{(acb),(cab)}}{q_{(acb),(cba)}} = \frac{q_{(bca),(cab)}}{q_{(bca),(cba)}}$$

(vi) $$\frac{q_{(bac),(abc)}}{q_{(bac),(acb)}} = \frac{q_{(cab),(abc)}}{q_{(cab),(acb)}}$$

In the present study we analyzed data sets with four brands each. In order to obtain three-brand choice sequences, we conducted two types of analysis. First, we restricted the analysis to consumers who chose only the three leading brands. This generated the "selective" data set. Second, we lumped together the two least popular brands per consumer (as was first done in Massy, Montgomery, and Morrison (1970) and has been common practice ever since), generating the "combined" data set. For each consumer, at each stage t, the state s_t is defined as in Section 2. Note that the last states are not well defined, since we observe only the top-ranked brand in the permutation. To render them well-defined, we "doubled" each sequence by concatenating it with itself.

Thus we obtained, for each consumer, an occurrence transition matrix. We focus on the hypothesis that, for each consumer $q_{(abc),(abc)} = q_{(acb),(acb)}$ where a is the consumer's most frequently chosen brand. Observe that this is an implication of the hypothesis that the transition matrix $(q_{s,t})$ is derived from a first-order Markov model. (In fact, this is but one of the six equalities specified above.) This hypothesis was tested using a standard goodness-of-fit χ^2- test with maximum likelihood estimators for $p_{a,a}$ for each consumer.

Due to the unavailability of long choice sequences, the occurrence matrices tend to have small numbers in some entries. Hence the theoretical validity of the χ^2- test is somewhat arguable, and we therefore present the following with some diffidence.

Table 2 presents the p-values of the χ^2 tests for the six data sets:

Table 2:

Product Category	**Version**	**p-value**
Cracker	Selective	0.1636
	Combined	0.0025
Yogurt	Selective	0.0033
	Combined	0.0001
Catsup	Selective	0.4489
	Combined	0.2615

As Table 2 shows, there is very strong evidence that the ordinal model provides explanatory power beyond the first-order Markov process, that is, that order effects exist, in the crackers and yogurt categories. Interestingly, the catsup data are inconclusive. This might be due to the fact that three of the four catsup brands are manufactured by Heinz and are similar enough to be confounded in the consumer's memory[3]. At any rate, we find the above results indicative of the existence of order effects.

[3] Indeed, when the two least popular Heinz brands are considered as one, the p-value drops to 0.1520.

4.2. Test of the Inertia Effect

Following Massy, Montgomery, and Morrison (1970), we lumped together the three least popular brands for each consumer. Thus the problem was reduced to a two-brand choice, "most preferred" (Brand 1) versus "other" (Brand 2).

For each consumer we computed maximum likelihood estimators using the first-order standard Markov process, and the first-order standard Markov process with inertia. The comparison of the last two directly tests the inertia assumption in the context of our ordinal model for $n = 2$.

The specific model we estimated allows inertia (dormancy) periods only for brand 1. Indeed, repeated choice of "brand" 2 may be an artifact of our lumping process. We assumed that, at every period in which Brand 1 was chosen, there is a probability r of the consumer becoming dormant for L periods, with $L \sim N(\mu, \sigma^2)$. Thus, the inertia model has three additional parameters, namely, (r, μ, σ) as compared to the standard first-order Markov one.

Note that, for a consumer who bought a single brand, both models can be calibrated to obtain a likelihood value of 1. Thus these consumers cannot help us distinguish between the models. The percentage of consumers that bought a single brand were 27.5%, 24.5%, and 39.4% for the cracker, yogurt and catsup data sets respectively. We therefore restrict attention to the other consumers.

The results of our analysis are presented in Table 3. For each data set, we provide the percentage of consumers that bought more than one brand, for whom (i) the inertia model performs better than the first-order Markov one; (ii) the above holds with statistical significance at various significance levels (using a $\chi^2_{(3)}$ test for the log-likelihood ratios).

Table 3:

	Higher Likelihood	At least 50% Significant	At least 70% Significant	At least 90% Significant
Cracker	80.2%	59.4%	45.1%	15.4%
Yogurt	87.2%	76.9%	36.0%	17.9%
Catsup	82.6%	52.2%	30.4%	13.1%

For most consumers, the number of purchases available in the data sets is not very large. Thus standard significance levels applied to single consumers in this literature range between 70% and 90%. (See Jeuland (1979) and Givon (1984).) According to this measure, the results reported above seem to indicate that the inertia effect does exist [4].

However, one may wonder how much additional explanation is provided by our model as compared to the first-order Markov one. The standard measure for goodness-of-fit for these models was suggested by McFadden(1974), and it is computed as follows. Given a proposed model, one computes the maximum of the log-likelihood function according to this odel, to be denoted $L(X)$. As a basis for comparison, one computes the log-likelihood function of the independent, equi-probable choice model, according to which every brand is chosen with probability $\frac{1}{n}$ at every stage. Denoting the latter by $L(0)$, the goodness of fit measure is

$$\rho^2 = 1 - \frac{L(X)}{L(0)} \ .$$

Note that ρ^2 is in [0,1], with $\rho^2 = 0$ denoting no improvement of the model over the independent equi-probable choice model, while $\rho^2 = 1$ corresponds to a perfect fit. Hauser (1978) has shown that, for time-independent models, ρ^2 can be derived from

[4] The data as well as the estimation program are available upon request.

information-theoretic arguments, as the fraction of explained entropy. However, this measure is commonly used for time-dependent models as well.

For each consumer in each data set, we computed the value of ρ^2 for the two models compared above: the first-order Markov chain, and our model with inertia. To aggregate over consumers, we calculated two averages per data set: a simple average, and a weighted one, where we used the number of purchases as weights. As expected, both models perform better according to the weighted average than according to the simple one, because consumers with longer choice sequences tend to support the dependence on past choices. Since ρ^2 compares both models to the time-independent, equi-probable choice model, we get higher ρ^2 values when more weight is put on longer choice sequences. At any rate, according to both averages, the explanatory power of our model is substantially greater than that of the first-order Markov model, as shown in Table 4.

Table 4:

		First-order Markov	**Inertia**
Cracker	Simple Average ρ^2	0.3367	0.4360
	Weighted Average ρ^2	0.3781	0.4793
Yogurt	Simple Average ρ^2	0.4083	0.5681
	Weighted Average ρ^2	0.6274	0.6895
Catsup	Simple Average ρ^2	0.3110	0.3672
	Weighted Average ρ^2	0.3319	0.3959

To sum, the data indicate that the inertia effect is not only statistically significant for a non-negligible set of consumers, but also that it improves the explanatory power of the model.

5. Proofs

Proof of Observation 2.1: It is sufficient to observe that for any $s, t \in S$ there is a path of length n (or less) leading from s to t, using only arcs in *E*. Q.E.D.

Proof of Theorem 2.2:

Extend $x = (x_1, \cdots, x_T)$ to an infinite sequence $x = (x_1, \cdots, x_T, x_{T+1}, \cdots,) \in A^\infty$ such that every $a \in A$ appears in x infinitely often. We will find a sequence $(s_0, s_1, \cdots, s_T, s_{T+1}, \cdots,) \in S^\infty$ such that conditions (i) and (ii) hold for all $\tau \geq 0$. In particular, $(s_0, s_1, \cdots, s_T)$ will be the desired sequence.

Let $d(x,t,a) = \min\{i \geq 0 \mid x_{t+i} = a\}$ be the number of periods (from time t on) until the first choice of *a*. Note that for all t, *a*, $d(x,t,a) < \infty$ by our choice of (the extension of) x. For any time $t \geq 0$ define a binary relation $\succ_t$ on A as follows:

$$a \succ_t b \;\; \textit{iff} \;\; d(x,t,a) < d(x,t,b)$$

(*a* is "preferred" to *b* iff after time t, the first time *a* is chosen precedes the first time *b* is chosen.)

It is easy to verify that for all $t \geq 0$, $\succ_t$ is a linear ordering of A (i.e., it is transitive, irreflexive, and connected in the sense $(a \succ_t b)$ or $(b \succ_t a)$ for all $a \neq b$). Hence, $\succ_t$ corresponds to a permutation $s_t \in S$. Furthermore, x_t is obviously the $\succ_t$-maximal element in A. That is, $s_t(1) = x_t$. Finally, by definition of $\succ_t$, if $b, c \in A$ are <u>not</u> $\succ_t$-maximal, then $b \succ_t c$ *iff* $b \succ_{t+1} c$. Therefore $(s_t, s_{t+1}) \in E$ and the proof is complete.

Q.E.D.

Proof of Theorem 3.1:

The proof consists of three stages. We will first prove that with a very high probability every alternative will be chosen infinitely many times. Next we will show that with a probability which is close to one, for every two alternatives the ratio of the number of times they were chosen is (almost) the inverse of the ratio of the means of their distributions. Finally we use standard arguments to prove the existence of the limit.

Lemma 1: For every $a \in A$, $K \geq 0$, and $\varepsilon > 0$, there exists T such that

$$\Pr\left[\left\{\omega \in \Omega \mid F(\omega, a, T) \geq K\right\}\right] \geq 1 - \varepsilon .$$

Proof: For $T \geq 0$, let B_T be the set of states of the world at which a was chosen at most K times until time T:

$$B_T = \left\{\omega \in \Omega \mid F(\omega, a, T) < K\right\}$$

We are looking for $T \geq 0$ such that $\Pr[B_T] \leq \varepsilon$. The idea of the proof is as follows. If alternative a has been chosen only K times, with a very high probability its U-value is not too low. Moreover, there has to be another alternative that was chosen many times. By a law of large numbers, this other alternative had a lower U-value than a (with a large probability), in contradiction to U-maximization.

Let $L < 0$ be such that $\Pr[\tilde{x} > L] > \left(1 - \frac{\varepsilon}{2}\right)^{1/K}$ where $\tilde{x} \sim G_a$. Notice that such a number exists since G_a has a finite variance. Let D_T be the subset of B_T on which alternative a has had a very low U-value by time T:

$$D_T = \{\omega \in B_T \quad \exists t \le T \quad U(\omega, a, t) < K \cdot L\}$$

The choice of L guarantees that $\Pr[D_T] < \frac{\varepsilon}{2}$ for every T. We now wish to show that, for large T, $(B_T \setminus D_T)$ has small probability. Define

$$B_T^b = \left\{\omega \in B_T \setminus D_T \quad b \in \arg\max_{a \in A} F(\omega, a, T)\right\},$$

and note that for $\omega \in B_T^b$, $F(\omega, b, T) \ge \frac{T-K}{n-1}$.

Fix $b \ne a$. Let $F_b \ge 1$ be large enough such that for any sequence $\widetilde{Y}_1, \widetilde{Y}_2, \ldots$ of i.i.d random variables on Ω, each distributed according to G_b,

$$\Pr\left[\left\{\omega \in \Omega \,\middle|\, \sum_{j=1}^{m} \widetilde{Y}_j < K \cdot L \quad \forall m \ge F_b\right\}\right] \ge 1 - \frac{\varepsilon}{2n}$$

Such a number $F_b \ge 1$ exists by a law of large numbers (see, for instance, Halmos (1950)), since $\mu_b < 0$. Let $T_b = K + (n-1) \cdot F_b$. For every $T \ge T_b$ and $\omega \in B_T^b$ we are guaranteed that up to time T, b was chosen at least F_b times. Next define $t^*(\omega, b, T)$ to be the last time that b was chosen at ω up to time T. Note that for $\omega \in B_T^b$ and $T \ge T_b$, $F(\omega, b, t^*(\omega, b, T)) \ge F_b$. Let us focus on period $t^* = t^*(\omega, b, T)$ for $\omega \in B_T^b$ and $T \ge T_b$. We can now show that $\Pr[B_T^b] < \frac{\varepsilon}{2n}$. Indeed,

$$B_T^b = \{\omega \in B_T^b \mid U(\omega, b, t^*) \ge K \cdot L\} \cup \{\omega \in B_T^b \quad U(\omega, b, t^*) < K \cdot L\} \quad .$$

Note that, by definition, $B_T^b \subseteq (D_T)^c$. Thus for $\omega \in B_T^b$ we get $U(\omega,a,t) \geq K \cdot L$ for all $t \leq T$. Since b was chosen at $t^* = t^*(\omega,b,T)$, $U(\omega,b,t^*) \geq U(\omega,a,t^*)$. That is,

$$\Pr\left[\left\{\omega \in B_T^b \quad U(\omega,b,t^*) < K \cdot L\right\}\right] = 0 \quad .$$

Hence for $T \geq T_b$, $\Pr\left[B_T^b\right] = \Pr\left[\left\{\omega \in B_T^b \quad U(\omega,b,t^*) \geq K \cdot L\right\}\right] \leq \frac{\varepsilon}{2n}$. Let $T = \max_{b \neq a} T_b$ and observe that $B_T \subseteq \left(\bigcup_{b \neq a} B_T^b\right) \cup D_T$. Hence $\Pr[B_T] \leq \sum_{b \neq a} \Pr\left[B_T^b\right] + \Pr[D_T] \leq \frac{\varepsilon}{2n} \cdot n + \frac{\varepsilon}{2} = \varepsilon$. Q.E.D.

Lemma 2: For every $\varepsilon > 0$ and $a,b \in A$, there exists T such that

$$\Pr\left[\left\{\omega \in \Omega \quad F(\omega,a,t) \geq \left(\frac{\mu_b}{\mu_a} - \varepsilon\right) \cdot F(\omega,b,t) \quad \forall t \geq T\right\}\right] \geq 1 - \varepsilon$$

Proof: Let there be given $\varepsilon > 0$. Let $\eta > 0$ satisfy $\eta < \min\left(-\mu_a, \frac{-\varepsilon \cdot \mu_a^2}{(1-\varepsilon) \cdot \mu_a + \mu_b}\right)$. Let F_a be large enough such that for every sequence $\widetilde{Y}_1, \widetilde{Y}_2, \ldots$ of i.i.d random variables distributed according to G_a,

$$\Pr\left[\left|\frac{1}{m}\sum_{j=1}^{m} \widetilde{Y}_j - \mu_a\right| > \eta \quad \forall m \geq F_a\right] < \frac{\varepsilon}{6}$$

Notice that such F_a exists by the strong law of large numbers. Define F_b similarly (with G_b and μ_b replacing G_a and μ_a, respectively). Using Lemma 1, find T_a such that the probability that a was chosen at least F_a times before T_a is greater than $1 - \frac{\varepsilon}{3}$.

Using lemma 1 again find a time T such that up to it alternative b was chosen at least $T_a + F_b$ times with probability $1-\frac{\varepsilon}{3}$ or higher.

Let E_a be the event that a was not chosen frequently enough by time T_a, or that, after that time, the average utility realization of a was far from μ_a:

$$E_a = \{\omega \in \Omega \mid F(\omega,a,T_a) < F_a\} \cup$$
$$\left\{\omega \in \Omega \;\middle|\; F(\omega,a,T_a) \geq F_a \;\; and \;\; \exists t > T_a \;\; s.t \;\; \left|\frac{U(\omega,a,t)}{F(\omega,a,t)} - \mu_a\right| > \eta\right\}$$

Clearly $\Pr[E_a] < \frac{\varepsilon}{6} + \frac{\varepsilon}{3} = \frac{\varepsilon}{2}$. Define E_b similarly, with b, F_b, and T instead of a, F_a, and T_a, respectively.

From now on we will focus on $\Omega_1 = \Omega \setminus (E_a \cup E_b)$ whose measure is at least $1-\varepsilon$. Let B be the set of states at which a was not chosen often enough relative to b:

$$B = \left\{\omega \in \Omega_1 \;\middle|\; \exists t_\omega > T \quad F(\omega,a,t_\omega) < \left(\frac{\mu_b}{\mu_a} - \varepsilon\right) \cdot F(\omega,b,t_\omega)\right\}$$

For $\omega \in B$ consider $t_\omega > T$, and let $t^* = t^*(\omega,b,t_\omega)$ be the last time that b was chosen at ω up to time t_ω. Clearly:

$$F(\omega,a,t^*) \leq F(\omega,a,t_\omega) \text{ and } F(\omega,b,t_\omega) = F(\omega,b,t^*) \quad .$$

Since $F_b \geq T_a$, we must have $t^* \geq T_a$. Hence on Ω_1, at t^*, a was chosen enough times to guarantee that $\left|\frac{U(\omega,a,t^*)}{F(\omega,a,t^*)} - \mu_a\right| < \eta$. Thus, for $\omega \in B$,

$$F(\omega,a,t^*) \le F(\omega,a,t_\omega) < \left(\frac{\mu_b}{\mu_a} - \varepsilon\right) \cdot F(\omega,b,t_\omega) = \left(\frac{\mu_b}{\mu_a} - \varepsilon\right) \cdot F(\omega,b,t^*)$$

Since b was chosen at $t^* = t^*(\omega,b,T)$, $U(\omega,b,t^*) \ge U(\omega,a,t^*)$. But for $\omega \in B$ $(\subseteq \Omega_1)$,

$$U(\omega,b,t^*) < (\mu_b + \eta) \cdot F(\omega,b,t^*) \qquad \text{and} \qquad U(\omega,a,t^*) > (\mu_a - \eta) \cdot F(\omega,a,t^*).$$

Combining the above inequalities we get

$$(\mu_b + \eta) \cdot F(\omega,b,t^*) > (\mu_a - \eta) \cdot F(\omega,a,t^*)$$

or

$$\frac{\mu_b + \eta}{\mu_a - \eta} \cdot F(\omega,b,t^*) < F(\omega,a,t^*) \ .$$

However, η, was chosen to be small enough so that, $\frac{\mu_b + \eta}{\mu_a - \eta} > \frac{\mu_a}{\mu_b} - \varepsilon$, in contradiction to $\omega \in B$. Thus $\Pr[B] = 0$ and the required inequality holds with probability $\Pr[\Omega_1] = (1 - \varepsilon)$. Q.E.D.

We can now prove the theorem: consider a state of the world $\omega \in \Omega$ at which the limit does not exist or it exists but is different from f_μ. Then there must exist an integer $k > 0$, two alternatives $a, b \in A$, and an infinite sequence of stages S_k, such that

$$\left|\frac{F(x,a,t)}{F(x,b,t)} - \frac{\mu_b}{\mu_a}\right| > \frac{1}{k} \qquad \forall t \in S_k \ .$$

Given *k*, let B_k be the set of states at which there exists such S_k. Clearly, if for some *k*, $\Pr[B_k] \geq \eta > 0$ we get a contradiction to lemma 2 with $\varepsilon < \min(\eta, \frac{1}{k})$. Thus $\Pr[B_k] = 0$ for all *k* and therefore $\Pr\left[\bigcup_k B_k\right] = 0$ and the measure of the set on which there is no convergence to f_μ is zero. Q.E.D.

Proof of Theorem 3.2:

For a state ω, alternative a and time $t \geq 1$, define the number of times the consumer chose a when she was active in the sequence ω up to stage *t* to be:

$$C(\omega, a, t) = \#\left\{1 \leq \tau \leq t \;\; x_\tau^\omega = a \;\; and \;\; I_\tau^\omega = 1\right\}$$

where I_τ^ω is the choice indicator function. Let $U(\omega, a, t)$ denote the *U*-value of outcome $a \in A$ at that stage, according to the history dictated by ω. Of special interest will be the relative frequencies of the brands actively chosen, denoted by

$$c(\omega, a, t) = \frac{C(\omega, a, t)}{\sum_{b \in A} C(\omega, b, t)}$$

and their limit $c(\omega, a) = \lim_{t \to \infty} c(\omega, a, t)$. Denote the vector of limit frequencies by $c(\omega) = \left(c(\omega, 1), \cdots, c(\omega, n)\right)$.

If we restrict attention to the sequence of stages at which the consumer actually chose an alternative, we are back in a situation where the consumer chooses a brand at every period and gets a utility of:

$$\tilde{v}(a)=\sum_{j=1}^{L}\tilde{u}_j(a)$$

where L is a random variable with distribution Q^a, and $\tilde{u}_j(a)$ are i.i.d random variables, each distributed according to G_a. Hence $\tilde{v}(a)$ is a random variable with a negative mean and finite variance (see, for instance, Ross (1972), p. 75) and by applying Theorem 3.1 to $c(\omega)$ we get the limit frequencies of the choices at active periods.

Let $G=\{\omega\in\Omega \mid \forall a\in A,\ C(\omega,a,t)\to\infty\}$. Since the distributions $\{Q^a\}$ have finite variances, they are bound to be infinitely many active periods. Formally,

$$\Pr_Q\left[\left\{\omega\in\Omega \ \sum_{a\in A}C(\omega,a,t)\underset{t\to\infty}{\to}\infty\right\}\right]=1 \ .$$

By theorem 1, we get $\Pr_Q[G]=1$.

For $\omega\in G$ and $k\geq 1$, let $Y_k(\omega,a)$ be the length of the inertia period after the kth active choice of a by the consumer and prior to her (k+1)st such choice. Let

$$\overline{Y}_n(\omega,a)=\tfrac{1}{n}\sum_{k-1}^{n}Y_k(\omega,a) \qquad \text{and} \qquad G_1=\left\{\omega\in G \ \forall a\in A,\ \overline{Y}(\omega,a,)\to\upsilon_a\right\}.$$

For every a, $Y_1(\omega,a),Y_2(\omega,a),\ldots$ is a sequence of i.i.d random variables on G and according to the law of large numbers $\Pr_Q[G_1]=1$. For every time t between the kth and (k+1)st choices of brand a, the following inequality must hold:

$$\tfrac{k-1}{k}\cdot\overline{Y}_{k-1}\leq\frac{F(\omega,a,t)}{C(\omega,a,t)}\leq\frac{F(\omega,a,t)}{k}\leq\overline{Y}_k \ .$$

Thus, for every $\omega \in G_1$ and every $a \in A$: $\lim_{t\to\infty} \frac{F(\omega,a,t)}{C(\omega,a,t)} = \lim_{t\to\infty} \bar{Y}(\omega,a) = \upsilon_a$, which proves our desired result. Q.E.D.

Proof of Theorem 3.4:

We adopt the notation of Section 2. By Theorem 2.2, there is a sequence $(s_0, s_1, \cdots, s_T) \in S^{T+1}$ such that $s_\tau(1) = x_{\tau+1}$ and $(s_t, s_{t+1}) \in E \quad \forall\, 0 \le \tau \le T-1$. Let $U(a,0) = s_0^{-1}(a)$. For $\tau \ge 0$, let $a_\tau = s_\tau(1)$ and find $r_\tau \in \Re$ such that

$$U(a_\tau, \tau) + r_\tau < (>)\, U(b,\tau) \qquad \text{iff} \qquad s_{\tau+1}^{-1}(a_\tau) < (>)\, s_{\tau+1}^{-1}(b)$$

for all $b \ne a_\tau$. For instance, set $r_\tau = 0$ if $s_{\tau+1}^{-1}(a_\tau) = 1$, $r_\tau = U(s_\tau(n), \tau) - U(a_\tau, \tau) - 1$ if $s_{\tau+1}^{-1}(a_\tau) = n$, and

$$r_\tau = \frac{1}{2}\left[\min\left\{ U(b,\tau) \mid s_{\tau+1}^{-1}(a_\tau) > s_{\tau+1}^{-1}(b) \right\} + \max\left\{ U(b,\tau) \mid s_{\tau+1}^{-1}(a_\tau) < s_{\tau+1}^{-1}(b) \right\}\right] - U(a_\tau, \tau)$$

otherwise. Q.E.D.

6. Concluding Remarks

6.1 The models presented here do not explicitly deal with marketing factors such as prices, advertisement, and so forth. Rather, as in the early works on random utility models, these factors are incorporated into the distribution of the instantaneous utility (in the cardinal model), or into the transition probabilities (in the ordinal one). One direction in which these models should be generalized is to explicitly model the effects of such factors, which are both relevant and measurable.

6.2 Both the ordinal and the cardinal models may be extended to deal with situations where one brand's consumption affects the consumer's impression of other brands as well. This has been referred to as "act similarity" in Gilboa and Schmeidler (1993). In the present set-up, act similarity may be modeled by allowing all possible transitions in the ordinal model, or, in the cardinal one, by allowing $U(a,t)$ to change as a result of consuming brand $b \neq a$.

References

Bass, F. M., E. Pessemier, and D. R. Lehmann (1972), "An Experimental Study of Relationships Between Attitudes, Brand Preferences, and Choice," Behavioral Science, 17 (November): 532-541.

Ben-Akiva, M., and S. R. Lerman (1985), Discrete Choice Analysis. Cambridge, MA, MIT Press.

Carpenter, G. S., and D. R. Lehmann (1985), "A Model of Marketing Mix, Brand Switching, and Competition," Journal of Marketing Research, 22 : 318-329.

Chintagunta, P. K., S. Halder, and R. Roy (1994), "A Framework for Investigating Habit, 'The Hand of the Past,' and Heterogeneity in Dynamic Brand Choice," forthcoming in Marketing Science.

Chintagunta, P. K., C. D. Jain, and N. J. Vilcassim (1991), "Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data," Journal of Marketing Research, 28 : 417-428.

Chintagunta, P. K., C. D. Jain, and N. J. Vilcassim (1994), "A Random Coefficient Logit Brand Choice Model Applied to Panel Data," Journal of Business & Economic Statistics, 12 (3): 317-328.

Gilboa, I., and D. Schmeidler (1993), Case-Based Consumer Theory. Northwestern University.

Givon, M. (1984), "Variety Seeking through Brand Switching," Marketing Science, 3 (Winter): 1-22.

Guadagni, P. M., and J. D. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," Marketing Science, 2 (Summer): 203-238.

Halmos, P. R. (1950), Measure Theory. Princeton, NJ, Van Nostrand.

Hauser, J.R. (1978), "Testing the Accuracy, Usefulness, and Significance of Probabilistic Choice Models: An Information Theoretic Approach," Operations Research, 26 (May): 406-421.

Jeuland, A. (1979), "Brand Choice Inertia as One Aspect of the Notion of Brand Loyalty," Management Science, 25 (7): 671-682.

Massy, W. F., D. B., Montgomery, and D. G. Morisson (1970), Stochastic Models of buying Behavior The MIT Press, Cambridge, Mass.

McAlister, L. (1982), "A Dynamic Attribute Satiation Model of Variety Seeking Behavior," Journal of Consumer Research, 9 (September): 141-150.

McAlister, L., and E. Pessemier (1982), "Variety Seeking Behavior: An Interdisciplinary Review," Journal of Consumer Research, 9 (December): 311-322.

McFadden, D. (1981). Econometric Models of Probabilistic Choice. Structural Analysis of Discrete Data with Econometric Applications. Cambridge, MA, MIT Press.

Ross, S. M. (1972), Introduction to Probability Models. New York, Academic Press.

Simon, H. A. (1957), Models of Man. New York, Wiley.

Trivadi, M., F. M. Bass, and C. M. Rao (1994), "A Model of Stochastic Variety-Seeking," Marketing Science, 3 (Summer): 274-301.