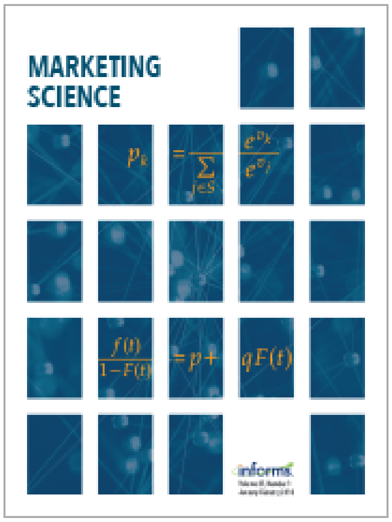


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A Fuzzy Set Model of Search and Consideration with an Application to an Online Market

Jianan Wu, Arvind Rangaswamy,

A Fuzzy Set Model of Search and Consideration with an Application to an Online Market

Jianan Wu • Arvind Rangaswamy
A. B. Freeman School of Business, Tulane University, New Orleans, Louisiana 70118-5669
Smeal College of Business, The Pennsylvania State University, University Park, Pennsylvania 16802-3007
jianan.wu • arvindr@psu.edu

Several two-stage choice models (consideration stage plus choice stage) have been proposed in the marketing literature. We extend this literature by developing a more general model that incorporates how consumer search influences the degree to which they consider various brands. To test the validity and value of our model, we operationalized it with data obtained from Peapod, an online grocer, where we tracked consumers' search processes. We demonstrate that our model performs better than competing models on all the key criteria. New choice models, such as the ones proposed here, are necessary for deriving managerially relevant understanding of choice behavior in online markets. Our empirical results suggest that consumers search both their internal memory and external information at the store to determine the degree to which they consider various brands. Consumers are also heterogeneous with respect to their capability to process external information. For some consumers, external information search dramatically increases the degree to which they consider various brands; but, for others, it has little impact on their consideration. We also find that certain features (e.g., personal lists) reduce consideration set sizes, whereas other features (e.g., sort) increase consideration set sizes.
(*Consideration Sets; Consumer Search; Online Shopping; Fuzzy Set Theory; Choice Models*)

1. Introduction

There is ample evidence to suggest that consumers do not choose products from a universal set of alternatives, but frequently form context-specific consideration sets from which they make their choices (for a review, see Shocker et al. 1991, Roberts and Lattin 1997). In some situations, particularly business-to-business contexts, the consideration set may be explicit and observable (e.g., Hauser and Wernerfelt 1990). In other situations, the consideration set is an unobservable psychological construct and has to be inferred from the choices that consumers make (e.g., Andrews and Srinivasan 1995).

Consideration sets have been the subject of much research in recent years (e.g., Gensch 1987; Swait and Ben-Akiva 1987; Lynch et al. 1988; Roberts and Urban 1988; Finn and Louviere 1990; Hauser and Wernerfelt 1990; Nedungadi 1990; Roberts and Lattin 1991, 1997; Shocker et al. 1991; Kardes et al. 1993; Andrews and Srinivasan 1995; Siddarth et al. 1995; Bronnenberg and Vanhonacker 1996; Desarbo et al. 1996; Chiang et al. 1999). This level of interest in studying consideration sets can be attributed to several factors:

• *Behaviorally*, a consideration set represents a task-simplifying heuristic that consumers use to cope with complex choice problems. Often, we need a multistage model for representing consumer choices (Bettman 1979).

• *Methodologically*, models that ignore consideration set information may have biased parameter estimates (Chiang et al. 1999). For example, in one study, multinomial logit models captured only 22%

0732-2399/03/2203/0411
1526-548X electronic ISSN

of the explainable uncertainty, whereas consideration sets account for the remaining 78% (Hauser 1978). An extensive simulation study by Abramson et al. (2000) shows that, even when four different unobserved effects are simultaneously present (i.e., choice set effects, heterogeneity in preference and market response, state dependence, and serial correlation), a logit model with choice set effects (i.e., incorporating consideration sets) produces the most valid parameter estimates.

• *Managerially*, when a brand enters the consideration set of a consumer, it increases the chances that the consumer will choose that brand even if it is not the most preferred, but exclusion prevents the selection of the brand even if it is likely to be preferred (Andrews and Srinivasan 1995).

Although it is well recognized that consideration set formation is important in understanding and explaining consumer choice behavior, modeling hierarchical choice processes in a parsimonious way is difficult (e.g., Tversky 1972). Empirically testing such models can also be very demanding in terms of data requirements (Gensch 1987). To overcome these drawbacks, researchers have proposed simplified two-stage choice models (Roberts and Lattin 1991, Andrews and Srinivasan 1995, Siddarth et al. 1995, Bronnenberg and Vanhonacker 1996, Chiang et al. 1999), in which consumers form a consideration set in the first stage and choose among the alternatives in their consideration sets in the second stage. The availability of scanner panel data has greatly aided the research on modeling and estimation of choice models that incorporate consideration sets (e.g., Andrews and Srinivasan 1995, Siddarth et al. 1995, Bronnenberg and Vanhonacker 1996, Chiang et al. 1999).

An important limitation of existing consideration set models is that none of them explicitly incorporates the dynamics of consumer search process in consideration set formation. Theoretical research on consumer choice suggests that search processes influence the formation and composition of consideration sets (e.g., Lynch et al. 1988, Shocker et al. 1991, Hauser et al. 1993). Recent papers (e.g., Alba et al. 1997) have argued that empirically understanding the influence of consumer search process on consideration set formation is important from both theoretical and managerial perspectives. Fortunately, online markets generate a considerable amount of observational data about the *process* that consumers go through when making brand choices. Such process data are not available in traditional marketing data sources, such as scanner data. By modeling process data, marketers can understand how search influences both consideration and choice, and use that knowledge to design various search and customization tools to influence consumers' brand consideration (e.g., Haubl and Trifts 2000).

The literature points to two approaches for modeling consideration sets. In the crisp set models (e.g., Roberts and Lattin 1991, Andrews and Srinivasan 1995, Siddarth et al. 1995, Chiang et al. 1999), an alternative is either *considered* or *not considered*. In the fuzzy set approach (e.g., Fortheringham 1988, Bronnenberg and Vanhonacker 1996), each alternative is considered to a *greater or lesser extent* than other alternatives. To incorporate consumer search process in consideration set formation, we use a fuzzy set approach. This allows us to specify how the dynamic search process influences the degrees of consideration of brands, something that other researchers have pointed out is desirable (e.g., Roberts and Lattin 1991). In fact, Roberts and Lattin (1991) suggest that the fuzzy set approach would have higher process validity and flexibility than existing methods.

Fuzzy Consideration Set Models

Fortheringham (1988) derives the following choice probability formula:

$$p(i \text{ is chosen}) = \frac{(i \text{ is considered})e^{v_i}}{\sum_{j=1}^{J} m(j \text{ is considered})e^{v_j}}, \tag{1}$$

where v_i is the utility of store i, m (i is considered) is the degree of membership of store i in the fuzzy consideration set, and J is the number of stores. Equation (1) has some desirable properties. (1) Choice probability is not subject to the assumption of Independence of Irrelevant Alternatives (IIA), a commonly criticized property in Luce-type choice models (Luce 1959). (2) Choice probability will reduce to a conventional multinomial logit formulation when consideration sets are crisp and fixed (if M is a crisp consideration set, then $m(i) = 1$ if $i \in M$ and 0 otherwise). Thus, Equation (1) gives exactly the multinomial logit choice probability conditional on M. (3) The choice probability has a closed formula, which significantly simplifies its estimation. However, Fortheringham's model has several drawbacks. (1) Although he claims that consideration sets are fuzzy, his formulation is not derived from fuzzy set theory. Also, he derives Equation (1) using probability theory, which we show is incorrect (see Technical Appendix). (2) He does not explain the source of fuzziness, or how marketing instruments may influence the extent of fuzziness of consumers' consideration sets. Thus, his arguments have a foundation neither in fuzzy set theory nor in consumer behavior theory.

Bronnenberg and Vanhonacker (1996) take Fortheringham's model a step further. They conduct empirical analyses of scanner data using the model specified in Equation (1). They specify consumers' utility functions for brands at both the *consideration* and *choice* stages in a way that allows them to infer how marketing instruments influence consideration and choice. They also incorporate a latent class segmentation model to capture heterogeneity of consumers' responsiveness to marketing instruments. Although this study provides important insights about choice behavior, it also has several limitations: (1) the authors use Equation (1) without offering any theoretical support; (2) they assume a fixed variance for the consideration utility (i.e., the brand salience) distribution through a Type 1 extreme value distribution for the entire population, an unsound strategy because there is likely to be heterogeneity in the variances of consideration utility distributions among segments (Swait and Stacey 1996); and (3) their segmentation scheme also assumes that the variances of the consideration utility distribution across segments are identical, a condition intrinsic to their model specification; therefore, the segment-level consideration utilities differ only in responsiveness to marketing instruments. However, differences in the distributions of consideration utility between segments may also result from differences in the variances of these subpopulations.

In this paper, our primary objective is to establish a theoretical framework for modeling consideration sets in which we explicitly include the impact of consumer search behavior on consideration set formation. Our model generalizes the existing models and also overcomes several conceptual and methodological limitations of these models. To empirically test our model, we apply it to a data set from an online supermarket, Peapod, Inc. In addition to the types of data available from scanner panels, the Peapod database also contains information about consumers' search processes captured in log files. Thus, our data are particularly suitable for empirically studying how consumer search influences consideration set formation.

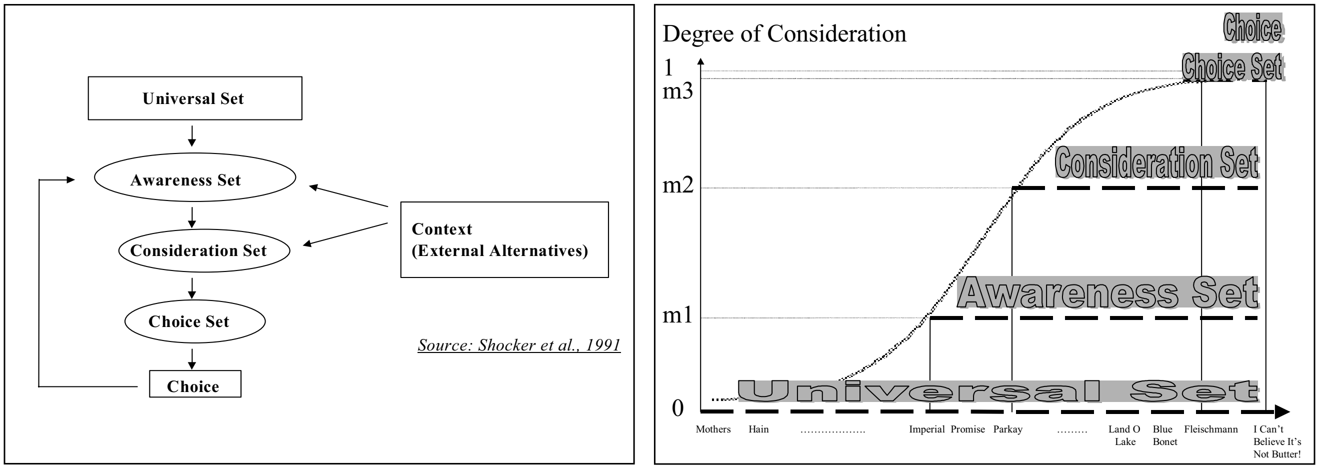
The remainder of the paper is organized as follows. In §2, we develop a general fuzzy-theoretic choice model based on an intuitive set of axioms. In §3, based on the axiomatic framework, we develop an operational model that we calibrate using data from Peapod. In §4, we summarize the empirical evaluation of our model using Peapod data. Finally, in §5, we discuss the contributions and limitations of our research and summarize directions for future research.

2. The Proposed Fuzzy Set Consideration Model

Conceptual Outline of Model

The consumer consideration process is dynamic in nature, which is illustrated by the model of Shocker et al. (1991) model (Figure 1a). Consider a consumer who is choosing a brand of soft margarine. In her search process, she may eliminate brands Mother's and Hain's because she has never heard of them (processing by brand names), which leads to her awareness set. She may then eliminate Imperial and Blue Bonnet because of their fat content (processing by nutrition), which defines her "consideration set." By evaluating prices, this consumer may further eliminate Promise, Parkay, and Land O' Lakes (processing by price), which leads to the choice set in which I Can't Believe It's Not Butter! and Fleischmann's are the only brands left. Finally, by making a trade-off

Figure 1 A Fuzzy Set View of Consideration Set Formation from Dynamic Sequential Choice Model Shocker et al. (1991)



Notes. (a) The model of dynamic choice process from Shocker et al. Consumers are aware of some of the available alternatives that form their awareness set. Furthermore, they seriously consider a subset of the alternatives from the awareness set to form their consideration set, either by searching information stored internally in their memory or information confronted externally, or by evaluating their choice goals along key attributes. Consumers may further select a smaller subset set (i.e., choice set) in which all the alternatives are compared using a compensatory process. Finally they choose a single alternative with the highest utility. (b) The model of Shocker et al. translates into a fuzzy set representation. As consumers proceed from the awareness set to the consideration set, and further to the choice set, and eventually to the final choice, what matters is not which alternative is considered, but rather how much an alternative is considered during the decision process. The consideration membership curve (the dotted curve) represents this dynamic consideration process.

between taste and quality, the consumer may choose I Can't Believe It's Not Butter! from the choice set. The interesting question is, "Which brands did this consumer really consider?" Did the consumer consider Fleischmann? Most of us would say *Yes*. How about Promise and Parkay? Most would say *Maybe*. How about Mothers? Most would say *No*. The vagueness in the answers reflects fuzziness, which makes it difficult to develop a crisp operational definition of the consideration construct.

Many of the past studies (e.g., Hauser and Wernerfelt 1990, Roberts and Lattin 1991, Siddarth et al. 1995) have used fuzzy words—such as, "*seriously* considered," "in the *near* future," and "*immediately prior* to choice"—to characterize the consideration stage in their two-stage models. These fuzzy words (*italicized*) refer to imprecision in meaning—for which fuzzy set theory is well suited—rather than implying uncertainty about events, for which probability theory is well established. A fuzzy set modeling approach is useful both for representing the fuzziness characterizing consideration sets, and for representing how the search process dynamically alters the degree to which consumers consider various brands based on their search processes. Thus, the fuzzy set approach is particularly applicable in online markets, where consumers engage in information search immediately before making their choices. We return to the soft margarine example to provide an intuitive description of how the fuzzy set approach would apply in that context. Suppose that m_i is the number of attributes that a consumer evaluates before brand i is eliminated in the dynamic consideration process. Then, m_i can be viewed as a proxy measure of the degree of consideration (m_i can be normalized to lie between 0 and 1). Some brands are eliminated earlier than others, and thus have a smaller degree of consideration, or, in our fuzzy set terms, have a lower degree of membership in the consideration set. Then, the dotted curve in Figure 1b illustrates the degrees of consideration for our hypothetical example. In the next two subsections, we formalize these intuitive notions.

Apart from basing consideration on fuzzy set theory, the rest of our modeling framework follows the two-stage formalisms currently used in the literature. In particular, in this line of research, researchers have specified different utility functions to represent consumers' consideration and choice decisions.

Bronnenberg and Vanhonacker (1996) propose that consumers may form their consideration and choice utilities based on different sets of information (called brand salience and brand value), and Andrews and Srinivasan (1995) propose that consumers may use the same set of information in both consideration and choice, but with different importance weights (called consideration utility and choice utility). We follow the literature and assume that consumers form consideration sets based on their *consideration utility* and their choices based on *choice utility*. We also allow for the possibility that some information consumers use to form their consideration and choice utilities may be common, whereas other information might uniquely influence either consideration or choice.

In the consideration stage, we posit that consumers initially form their consideration utility of an alternative according to the information they have about that alternative and then dynamically update their consideration utility using both internal search from memory and external search of the shopping environment for that alternative and for other alternatives. Next, they compare their consideration utility to a threshold (i.e., their minimum needs) to determine whether they should consider a particular alternative. The higher the consideration utility relative to threshold, the more they consider that alternative. As consumers get more information, they become more confident about their decisions to include or exclude a brand from their consideration sets. That is, information search reduces the fuzziness of consumers' consideration utilities and thresholds (in our formulation, new information does not shift the mean of the fuzzy consideration utility or threshold, but alters their variability, i.e., fuzziness).

Search reduces the fuzziness associated with consideration utility and, consequently, changes the degree of consideration of an alternative. Let m denote the degree of consideration; m is a function of both the consumers' search activities (S) and the information they encounter during the search (Z). For example, in online markets, S could represent the external search (e.g., sorting products along various attributes in an online market) or the internal search (e.g., recalling from memory prior knowledge about the products) and Z may include the external price and promotion information that consumers may encounter during their search.

We posit that the choice utility of an alternative after search depends on the choice utility for that alternative before the search, *and* on how the search alters the degree of consideration. The choice utility of an alternative before search, denoted as u, is determined from the information available about this alternative (e.g., prior knowledge, intrinsic preferences) before consumers initiate search activities. If new information encountered during the search suggests that a brand actually has more of the desirable characteristics, then consideration utility increases, and as a result, choice utility after the search increases as well. The opposite effects occur if the new information suggests that a brand has fewer desirable characteristics. We refer to the resulting choice utility after search as *search-adjusted choice utility*, which is a function of both u and m. This is analogous to the concept of *risk-adjusted utility* proposed by Roberts and Urban (1988), which we describe later in this section. We assume that consumers maximize search-adjusted choice utility.

We illustrate the concept of search-adjusted choice utility with an example: Suppose that a consumer prefers red cars, but is looking at the available cars through a colored glass (fuzzy glass). This distorts the color of the cars that the consumer actually sees, and consequently, distorts the degree of consideration of each alternative for that consumer. As the fuzzy glass becomes clearer (i.e., the consumer gathers more information), the exact color of the cars becomes more discernable, which changes the degree of consideration for the same car. Consequently, the choice utility of the car after the glass becomes clearer should be adjusted appropriately. If the color of a particular car becomes more clearly red, then that car is more likely to be considered and chosen (i.e., will have higher consideration utility and higher search-adjusted choice utility). On the other hand, if the consumer sees the color as more clearly purple, the car is less likely to be considered and chosen (i.e., lower consideration utility and lower search-adjusted choice utility).

Formal Representation of a Fuzzy Consideration Set Model

We first introduce several definitions.

Definition 1 (Lowen 1997). A fuzzy real number is a fuzzy set $A = \{\Re, \mu_A\}$, where $\Re$ is the set of real numbers, and the membership function μ_A, integrable on $\Re$, satisfies: $\mu_A(x) \geq 0$ and $\int_{\Re} \mu_A(x)\,dx = 1$.

Definition 2 (Lee and Li 1988). Let A be a fuzzy real number as defined in Definition 1. The fuzzy mean and fuzzy spread of A are defined as $\overline{A} = \int_{\Re} x\mu_A(x)\,dx$ and $\delta_A = (\int_{\Re} (x-\overline{A})^2 \mu_A(x)\,dx)^{1/2}$.

Note that the fuzzy mean and spread are defined as akin to the first two moments of a continuous probability distribution.

Definition 3 (Lowen 1997). Let $M(\Re)$ be the set of all fuzzy real numbers. Given two fuzzy real numbers $P = \{\Re, \mu_P\}$, $Q = \{\Re, \mu_Q\} \in M(\Re)$, P and Q are defined on $\Re$ and have membership functions μ_P and μ_Q, respectively. We define a fuzzy order ρ: $M(\Re) \times M(\Re) \to I$ as the degree that P is strictly smaller than Q as

$$\rho(P, Q) = \sup_{t\in\Re} P(-\infty, t) \wedge Q(t, \infty), \tag{2}$$

where $P(-\infty, t) = \int_{-\infty}^{t} \mu_P(x)\,dx$, $Q(t, \infty) = \int_{t}^{\infty} \mu_Q(x)\,dx$, and $P(-\infty, t) \wedge Q(t, \infty)$ the infimum of $P(-\infty, t)$ and $Q(t, \infty)$.

Let a consumer's consideration utility for an alternative i be represented by a fuzzy real number $\nu_i = \{\Re, \mu_{\nu i}\}$ with fuzzy spread $\delta_{\nu i}$. Let the threshold of consideration be represented by a fuzzy real number $T = \{\Re, \mu_T\}$ with fuzzy spread δ_T. Then $m_i = \rho(\nu_i, T)$ is the degree of consideration (i.e., the degree to which $\nu_i > T$ under fuzzy relation ρ given in Definition 3).

Definition 4. Let $\mu_M(i) = m_i$, $\forall i \in U$, where U is the universal set and m_i is the degree of consideration of alternative i. Let

$$M = \{U, \mu_M\}. \tag{3}$$

Then, M defines the consumer's fuzzy consideration set, with μ_M as the membership function of belonging to the fuzzy consideration set M.

Definition 5. Let $u_i \in \Re$ be the choice utility of alternative i before search. Let $f(u_i, m_i)$ be the search-adjusted choice utility (i.e., choice utility after search) of alternative i. f: $\Re \times I \to \Re$, where I is the unit interval.

In the choice stage, we assume that consumers maximize search-adjusted choice utility $f(u_i, m_i)$. However, to account for the inability of the modeler to specify fully a consumer's utility function, we adopt the random utility framework (e.g., McFadden 1974):

$$w_i = f(u_i, m_i) + \epsilon_i, \tag{4}$$

where w_i is the random utility of alternative i with $f(u_i, m_i)$ and ϵ_i as its deterministic and random component, respectively. u_i is the choice utility of alternative i before search, m_i is the degree of consideration of alternative i, and ϵ_i is the random error which is IID Type 1 extreme value distribution. Consumers choose the alternative that maximizes the random utility: $i = \arg\max_j \{w_j\}$. Then, the choice probabilities of the alternatives are given by:

$$p_i = e^{f(u_i, m_i)} \Big/ \sum_j e^{f(u_j, m_j)}. \tag{5}$$

Equation (5) is central to our modeling framework. We assume that consumers dynamically adjust their choice utilities through search and consideration via the functions f and m. Consumers may use different adjustment mechanisms (see §3 for two different adjusting mechanisms). However, the following axioms must apply to all adjustment mechanisms:

Axiom A1. $m_i = 0$ *or* 1 *iff* $\delta_{\nu i} = \delta_T = 0$.

Axiom A2. $m_i = 0.5$ *if* $\delta_{\nu i}$ *or* $\delta_T = \infty$.[1]

Axiom A3. *$f(u_i, m_i)$ is an increasing function of u_i and m_i.*

Axiom A4. *$f(u_i, m_i) = u_i$ for $m_i = m_{i0}$, where m_{i0} is the (unknown) initial degree of consideration of i before search.*

Axiom A5. $f(u_i, m_i) = -\infty$ *for* $m_i = 0$.

We interpret these axioms as follows:

Axiom A1 suggests that an alternative is fully considered or not considered at all if and only if both the consideration utility and threshold are crisp (i.e., not fuzzy). Thus, the crisp consideration set is a special case of the fuzzy consideration set.

Axiom A2 indicates what happens in the fuzziest case: If either the consideration utility is very fuzzy

[1] We could make this axiom more stringent (i.e., iff instead of if) by excluding the possibility that ν_i and T have an identical fuzzy mean.

or the threshold for consideration is very fuzzy, then the degree of consideration equals the degree of non-consideration. Consumers have no clear inclination to either include or exclude an alternative from their consideration sets.

Axiom A3 suggests that, if two alternatives have the same degree of consideration after the search, then the alternative that has the higher choice utility (u_i) before the search will also have the higher search-adjusted choice utility after the search. The search adjustment process does not distort the underlying utility scale. For any alternative, if the degree of consideration increases after the search, the search-adjusted choice utility increases as well.

Axiom A4 asserts that if the search does not change the degree of consideration, then the choice utility remains the same before and after the search.

Axiom A5 asserts that the lowest search-adjusted choice utility value ($-\infty$) is attained when the consumer does not consider the brand at all after the search.

Some Special Cases

There are several important implications of our general model:

(a) If $\delta_\nu = \delta_T = 0$, then by Axiom A1, we have the degree of consideration $m = 0$ or 1. We obtain Roberts and Lattin's (1991) deterministic crisp set model.

(b) If we assume that the initial degree of consideration before search $m_0 = 1$ and choose $f(u, m) = u + \ln m$ (f satisfies Axioms A3–A5), then we have the model proposed by Fortheringham (1988). (Note this assumes that the base state is that all the brands are fully considered, but search will reveal the true degree of consideration of an alternative.) In this case, choice probabilities are given by:

$$p_i = e^{f(u_i, m_i)} \Big/ \sum_j e^{f(u_j, m_j)} = m_i e^{u_i} \Big/ \sum_j m_j e^{u_j}, \quad \text{for } i = 1, \ldots, J. \quad (6)$$

(c) If we assume that the initial degree of consideration before search $m_0 = 1$ and choose $f(u, m) = u + \ln m$, and let m be the CDF of the logistic distribution (f satisfies Axioms A3–A5 and m satisfies Axioms A1 and A2, and the fuzzy consideration utility and threshold are represented by Type 1 extreme value distribution functions), then we have the model of Bronnenberg and Vanhonacker (1996) (i.e., p_i is given by (6) with $m_i = 1/(1 + e^{\theta - s_i})$, where s_i is the mean of the consideration utility of brand i and θ is the mean of the threshold).

(d) There are also similarities between our proposed concept of search-adjusted choice utility $f(u, m)$ and the risk-adjusted utility $g(\nu, r, \sigma^2) = \bar{\nu} - (r/2)\sigma^2$ defined by Roberts and Urban (1988), where $\bar{\nu}$ is the expected choice utility, σ is the dispersion of the choice utility, and r is the consumer's risk-aversion coefficient. For instance, if there is no uncertainty (i.e., $\sigma = 0$), then the risk-adjusted utility is $g = \bar{\nu}$. In our model, if there is no search activity or search does not change the degree of consideration, then by Axiom A4, we have $f = u$. Note also that g is an increasing function of $\bar{\nu}$ (i.e., the higher the mean utility, the higher the risk-adjusted utility, other things being equal). In our model, f is also an increasing function of u, which is guaranteed by Axiom A3.

Special cases (a) and (c) above have *fixed* fuzzy spreads. In (a), $\delta_v = \delta_T = 0$, whereas in (c), $\delta_T = 0$, $\delta_v = \pi/\sqrt{3}$ (the standard deviation of the logistic distribution). In our general model, we do not impose any constraints on δ_ν and δ_T, except that they should be nonnegative, by definition. This distinction is important because, as Swait and Louviere (1993) point out, fixing the variance in logit models may lead to biased estimates of responsiveness coefficients of marketing mix variables included in the utility function.

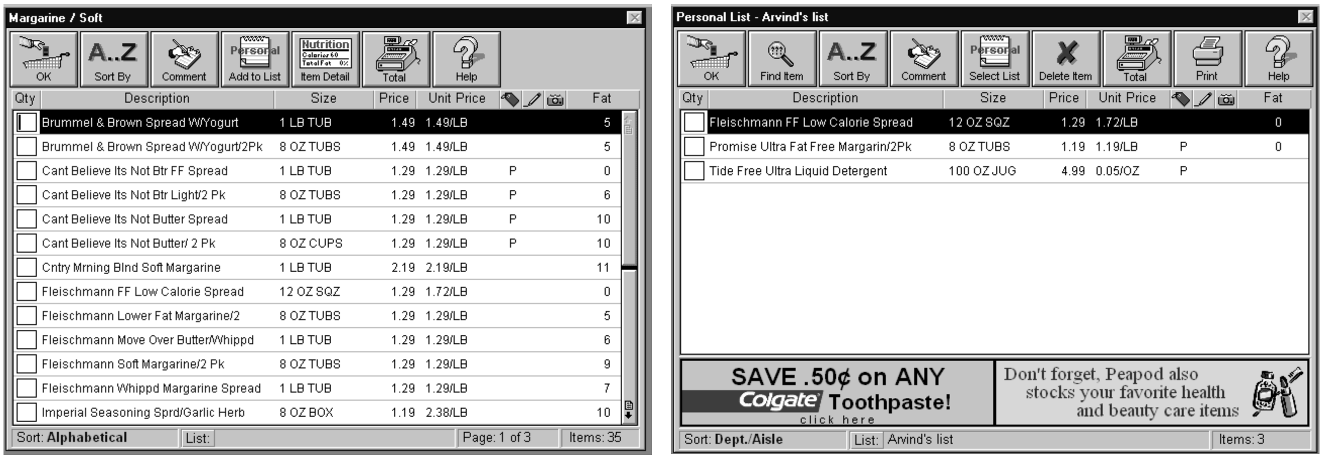
3. An Operational Model

We now describe the structure of the data available from Peapod and customize our general model for Peapod data.

Peapod Data

Peapod is an online grocery store headquartered in Skokie, Illinois. Subscribers to its service log onto the online supermarket via a modem. This system enables consumers to browse the grocery store by "pointing and clicking" icons and menus on the computer screen. Consumers select the products from the electronic aisles and send their "purchase order" to Peapod with a click of a mouse. Basically, there are

Figure 2 Shopping by Category and Personal List at Peapond Inc.



Notes. (a) Default list of search attributes available to consumers in the margarine category. The tag symbol indicates whether the product is on promotion. A "P" in this column indicates that the product is available at a special price to preferred customers who have a store card. A "*" in this column would indicate that the product is on a "Special promotion" available to all customers. The camera icon indicates whether a product picture is available (no pictures are available in the margarine category at this time). The column labeled "Fat" provides information on fat content. The "Sort By" button allows consumers to sort the information presented one criterion at a time. The sort criteria available are Alphabetical, Calories, Carbohydrates, Cholesterol, Dept./Aisle, Dietary Fiber, Fat (shown by default in all food categories), Kosher, Price, Protein, Size, Sodium, Specials, Sugar, and Unit Price. (b) Peapod members can shop for the items listed in their personal lists without having to go to the category aisle. The personal list includes the current price and promotion information.

two ways to shop on Peapod: (1) shop by category, in which all the available items in a product category are shown on the screen (Figure 2a) or (2) shop by using personal lists (created and maintained by consumers), in which the consumer chooses from only the items on the list (Figure 2b). Peapod then fills the orders at a local affiliated grocery store and delivers the products to customers within a specified time slot. The prices and promotions available on Peapod are identical to the in-store prices and promotions. In addition, Peapod sometimes offers its own promotions on behalf of manufacturers. Peapod charges a monthly subscription fee plus 5–10% of the grocery bill for each delivery.

Our Peapod database contains information on 278 subscribers randomly selected from consumers in the Chicago metropolitan area. We have the following information about these panelists for the period from May 1996 to August 1997 (unfortunately, we do not have demographic information about the subscribers):

Purchase Data. ID of person making purchases, date and time of purchase, items purchased, purchase price of items, and promotional status of purchased item (whether on promotion).

Marketing Data. Price and unit price, promotion (Peapod's and the affiliated local store's), product information (picture, nutritional information, etc.), brand name, and product size.

Navigation/Search Data. The sequence of activities completed during purchase (sort, add/delete items to personal list, etc.), whether an item was purchased from the personal list or by shopping the entire category, and whether items in the personal list or category were sorted by price, unit price, promotion, brand name, product size, nutritional information (fat, saturated fat, sugar etc.), or other criteria.

A Single-Segment Operational Model

Consideration Stage. We model consideration utility of brand i as a fuzzy real number $\nu_i = \{\Re, \mu_{\upsilon i}\} \in M(\Re)$, with membership function

$$\mu_{\nu i}(x) = \exp\left(-\frac{(x-\bar{\nu}_i)^2}{2\delta_{\nu i}^2}\right) \Big/ \sqrt{2\pi}\delta_{\nu i}, \quad x \in \Re, \quad \bar{\nu}_i = \sum_q \beta_q Z_{iq}, \quad (7)$$

where Z_{iq} are marketing mix variables (see the utility function specification section for details). For simplicity of notation, we do not include indices h and t to denote consumers and time of purchase, respectively. $\bar{\nu}_i$ is the fuzzy mean of the consideration utility of brand i. $\delta_{\nu i}$ is the fuzzy spread of the consideration utility of brand i. β_q are parameters of the utility function to be estimated.

The threshold level, $T = \{\Re, \mu_T\} \in M(\Re)$, of consideration is also a fuzzy real number with membership function:

$$\mu_T(x) = \exp\left(-\frac{(x-\tau)^2}{2\delta_T^2}\right)\Big/\sqrt{2\pi}\delta_T, \qquad x \in \Re, \tag{8}$$

where τ is a latent value that is the mean of the fuzzy threshold. δ_T is the fuzzy spread of the threshold. We assume the threshold to be constant across alternatives.

Proposition 1. *$m_i = \Phi((\bar{\nu}_i - \tau)/\delta_i)$, where $\delta_i = \delta_{\nu i} + \delta_T$ and Φ is the CDF of the standard normal distribution. $\mu_M(i) = m_i$, $\forall i \in U$ satisfies Axioms* A1 *and* A2.

Proof. See Technical Appendix.

Proposition 1 distinguishes our model from the traditional crisp-set-based probability models. The fuzzy order ρ: $M(\Re) \times M(\Re) \to I$ defined earlier is identical to the probability order if one of the arguments of ρ is crisp; however, ρ is different from the probability order when both of the arguments of ρ are fuzzy. In fact, if Equations (7) and (8) are interpreted within a probability framework, then probability ($\nu_i > T$) is equal to $\Phi((\bar{\nu}_i - \tau)/\sqrt{\delta_{\nu i}^2 + \delta_T^2})$, which is different from $m_i = \Phi((\bar{\nu}_i - \tau)/(\delta_{\nu i} + \delta_T))$ given in Proposition 1.

Consumer Search and Fuzziness Reduction. The ability to specify δ_i as a function of various aspects of consumer search is an important differentiator of our model from existing models of consideration sets. We model the fuzziness measure of the consideration set by specifying:[2]

$$\delta_i = \exp\left(\omega_P \,\text{Plist}_i + \sum_k \omega_k \,\text{Sort}_k + \omega_L \,\text{SI}_i\right), \tag{9}$$

[2] We reckon fuzziness reduction in terms of the combined effect of reductions in $\delta_{\nu i}$ and δ_T, denoted as δ_i. With more fine-grained data sources, we could separately estimate fuzziness reduction in $\delta_{\nu i}$ and δ_T within our model structure.

where Plist_i indicates whether brand i is in the personal list or not, ω_P captures the impact of the personal list on fuzziness reduction and is a parameter to be estimated, Sort_k indicates whether a consumer searches for and processes information about attribute k (e.g., sorting on unit price) during the choice process, ω_k captures the effect that the processing of information on attribute k has on fuzziness reduction and is a parameter to be estimated, SI_i indicates whether brand i was purchased on the previous purchase occasion by the consumer (SI_i is a measure of a consumer's most updated memory or short-term preference in a choice situation; see Bucklin and Lattin 1991, Bucklin and Gupta 1992), and ω_L captures the effect of SI_i on the current consideration set. In this operationalization, personal list and short-term preference capture the impact of internal information on fuzziness reduction, and sorting captures the impact of external information on fuzziness reduction.

Note that whereas Plist_i is brand-specific, Sort_k is a category-wide variable. When the consumer uses any sorting operation, it would "turn on" that variable for all brands (alternatives). Therefore, it is important to see how Sort_k actually affects the degrees of membership of the brands in the consideration set. Plug (9) into m_i in Proposition 1. We have $m_i = \Phi(x_i e^{-\omega_k \text{Sort}_k})$, where $x_i = (\nu_i - \tau)/\exp(\omega_P \,\text{Plist}_i + \sum_{l \neq k} \omega_l \,\text{Sort}_l + \omega_L \,\text{SI})$. If we assume that Sort_k reduces fuzziness, then $-\omega_k$ is positive. For any two brands i and j in the universal set, suppose that brand i has higher membership than brand j before sorting attribute k (i.e., $m_i > m_j$ for $\text{Sort}_k = 0$). Then, $x_i > x_j$ because Φ is an increasing function. After sorting attribute k (i.e., $\text{Sort}_k = 1$), then we have $m_i = \Phi(x_i e^{-\omega_k})$ and $m_j = \Phi(x_j e^{-\omega_k})$. Therefore, $m_i > m_j$ because $x_i e^{-\omega_k} > x_j e^{-\omega_k}$; i.e., sorting does not change the relative order of membership of brands within the consideration set.

Second, the marginal effect of Sort_k on consideration set membership can be determined by taking the partial derivative of m_i with respect to Sort_k:

$$\partial m_i / \partial \,\text{Sort}_k = \phi[(\bar{\nu}_i - \tau)/\delta_i][(\bar{\nu}_i - \tau)/\delta_i](-\omega_k). \tag{10}$$

ϕ is the pdf of the normal distribution. Thus, Sort_k has maximum marginal effect on those brands for which $|\delta_i| = |(\bar{v}_i - \tau)|$. For brands with δ_i close to either 0 or

∞, the marginal effect is close to 0. (Note that when δ_i goes to 0, the first term goes to 0 faster than the second term goes to ∞.) We get the intuitively appealing result that new information (i.e., sorting on an attribute) has minimal impact when a brand's consideration utility is very fuzzy or when it is very crisp. However, sorting has significant impact when the brand's consideration utility is moderately fuzzy (i.e., when the consumer is "on the fence" in deciding whether or not to consider a brand).

Choice Stage. Axioms A3–A5 only specify the outcomes of the search process, and do not indicate how the search process influences choice utility. Consumers may use several different updating mechanisms (Rust et al. 1999). One simple and widely adopted mechanism is that consumers adjust their choice utilities additively at each step of their search based on the relative degree of change in consideration. If we incorporate this assumption, we have:

PROPOSITION 2. *If, in addition to Axioms* A3–A5, *search adjusts choice utility additively, based on the relative change in degree of consideration, then we must have* $f(u, m) = u + \theta \ln m - \theta \ln m_0$, *with* θ *being a fixed positive constant for each alternative.*

PROOF. See Technical Appendix.

We choose $\theta = 1$ (i.e., $f(u, m) = u + \ln m - \ln m_0$) for our empirical analysis because this specification allows us to: (1) compare our operationalized model with previous models (e.g., Bronnenberg and Vanhonacker 1996) and (2) specify a parsimonious model, given the limited number of observations in the Peapod data. To assess the validity of this functional form, we also empirically compare it with an alternative specification, $f(u, m) = u + (1/m_0 - 1/m)$, which also satisfies Axioms A3–A5 but not under the additivity assumption of search effects, as in Proposition 2. Our analyses indicate that the first specification, namely, $f(u, m) = u + \ln m - \ln m_0$, performs significantly better (in a statistical sense) on both goodness-of-fit and prediction, and also converges faster, and hence is computationally more attractive.[3] Under our specification, we have

$$p_i = e^{f(u_i, m_i)} \Big/ \sum_j e^{f(u_j, m_j)}$$
$$= m_i e^{u_i - \ln m_{i0}} \Big/ \sum_j m_j e^{u_j - \ln m_{i0}}, \quad \text{for } i = 1, \ldots, J. \quad (11)$$

Propositions 1 and 2 guarantee that our operationalization of the general model satisfies all the Axioms (A1–A5) we specified earlier.

Utility Function and Variable Specifications. We specify consideration utility $\bar{\nu}_i$ as a linear function of marketing mix variables, namely, price and promotion. We do not specify a constant term because it cannot be separately identified from the threshold τ. Thus, $\bar{\nu}_i = \beta_1 \,\text{Price}_i + \beta_2 \,\text{Promotion}_i$. We specify choice utility as $u_i = k_i + \alpha_1 \,\text{Price}_i + \alpha_2 \,\text{Promotion}_i + \alpha_3 \,\text{LI}_i$. Note that this specification is equivalent to $u_i - \ln m_{i0} = c_i + \alpha_1 \,\text{Price}_i + \alpha_2 \,\text{Promotion}_i + \alpha_3 \,\text{LI}_i$. c_i's are alternative-specific constants. LI_i captures the time-invariant component of preference or long-term preference (Bucklin and Gupta 1992), and is constructed as in Bucklin and Lattin (1991). Specifically, LI_i is the proportion of times that brand i was purchased during the initialization period (i.e., it is a measure of the share of requirements for brand i). We follow Andrews and Srinivasan (1995) for specifying marketing mix variables in both the consideration and choice stages. Following Bucklin and Lattin (1991), we separately model the impact of past purchase behavior as due to a short-term preference variable (SI_i) and long-term preference (LI_i). As in Bronnenberg and Vanhonacker (1996), SI_i impacts consideration and LI_i impacts choice. However, unlike Bronnenberg and Vanhonacker's model, our approach does not require an additional smoothing constant, which makes our model more parsimonious and computationally simpler.

We use the following variables from our data set: price, promotion, price sorting, unit price sorting, promotion sorting, size sorting, SKU name sorting, and use of personal list. Price refers to the actual price seen and paid by the consumers, which is then

[3] We thank the reviewers for suggesting the comparison with an alternative functional form.

adjusted by SKU sizes (i.e., we only use unit prices in the empirical analysis). Promotion is often accompanied by a price cut and is shown as a flag on the screen seen by the consumers. We operationalize promotion as a dichotomous variable.

For representing a consumer's use of a personal list, we define Plist_i, a dichotomous variable with 1 denoting that alternative i is in the personal list and 0 otherwise. In a given choice situation, we only know whether alternative i was purchased from the personal list. When i is purchased, if we specify $\text{Plist}_i = 1$ only for alternative i and $\text{Plist}_i = 0$ for the alternatives not purchased, without knowing what other alternatives were in the personal list, it would mean that we are utilizing endogenous information. To avoid this problem, we dynamically form a personal list for each consumer by updating his/her original personal list *after* each purchase; if a consumer purchases an alternative from the personal list at choice occasion t, then we add the chosen alternative to the personal list of that consumer in all subsequent choice occasions.

The sorting variables are also dichotomous, with 1 representing that the consumer used a particular sorting operation on a given purchase occasion and 0 representing nonuse. Because sorting is sparse, we aggregate all the sorting variables together into a single variable to improve the reliability of our estimates. Specifically, we combine five sorting variables to define an aggregate-level dichotomous sorting variable that takes a value 1 if the consumer uses any sorting operation, and a value 0 if the consumer does not use any sorting operation.

For operationalizing LI_i we require an initialization period (see Bucklin and Lattin 1991, Bucklin and Gupta 1992), which is difficult when we have limited observations for each individual. For each consumer, we initialized this variable by randomly drawing one-fourth of the consumers' observations. Because consumers in our data set are fairly loyal (with a switching rate of less than 20%; see Table 1), this operationalization would represent consumers' long-term preference fairly well.

Based on the above specification, the likelihood function (adding the consumer index h and purchase time index t) is given by:

$$L=\prod_h\prod_t\prod_i (p_{it})^{Y_{it}(h)}, \quad \text{where} \quad p_{it}=m_{it}e^{u_{it}} \Big/ \sum_j m_{jt}e^{u_{jt}}, \tag{12}$$

where $Y_{it}(h) = 1$ if consumer h chooses alternative i at purchase situation t and 0 otherwise. We use Maximum Likelihood Estimation to estimate the parameters.

All the parameters in our models are identified. First, examine membership, m_i, given by Proposition 1. Notice that there is no constant term estimated in $\bar{\nu}_i$ (otherwise $\bar{\nu}_i - \tau$ would have an unidentifiable constant) or in δ_i (otherwise, a scale constant between $\bar{\nu}_i - \tau$ and δ_i cannot be identified). Thus, the parameters in m_i are identified. Note that if we rescale all m_i by multiplying by a constant, the probabilities in (11) remain the same. However, such an unidentifiable constant does not exist because the function Φ is nonlinear. Second, examine the parameters in u_i. By the choice probability formula (11), the parameters in u_i are identifiable up to an additive constant. To ensure this in the estimation, we fix the last alternative-specific constant $c_J = 0$ in (11).

Multiple-Segment Model Specification

Consumers are heterogeneous with respect to: (a) their responsiveness to marketing instruments and (b) their abilities to process information to reduce fuzziness.[4] Bronnenberg and Vanhonacker (1996) examined Type A heterogeneity and found that segments vary in the distribution of consideration set sizes. Alba and Hutchinson (1987) discuss Type B heterogeneity from the perspective of consumer expertise. They hypothesize that consumers with different degrees of knowledge may have significantly different search behavior. To our knowledge, Type B heterogeneity has not been examined in an empirical setting.

[4] Consumers may also be heterogeneous in intrinsic preferences between alternatives (observed) and within alternatives (unobserved) (see, e.g., Gupta and Chintagunta 1994). Because our focus is on consideration rather than choice and because we do not have sufficient consumer switching in this data, we only model *observed* heterogeneity in intrinsic preferences. Indeed, the improvement in goodness-of-fit of our model with the relaxation of alternative-specific constants suggests that we should not increase the number of estimated parameters.

To represent these two types of heterogeneity together, we propose a latent (finite) mixture specification in conjunction with our two-stage model. Suppose there are S segments of consumers. We specify the parameters (i.e., β's in (7), τ's in (8), ω's in (9), and α's in u_i in (11)) at the segment level. Let $m(s)$, $u(s)$, and $\psi(s)$ denote the membership function of the fuzzy consideration set (given by Proposition 1), the choice utility (given by (11)), and the relative size of segment s $(s = 1, \ldots, S)$, respectively. Segment-level parameters can be estimated using the likelihood function:

$$L = \prod_h \sum_s \psi(s) \prod_t \prod_i (p_{it|s})^{Y_{it}(h)}, \quad \text{where}$$

$$p_{it|s} = m_{it}(s) \exp(u_{it}(s)) \Big/ \sum_j m_{jt}(s) \exp(u_{jt}(s)). \tag{13}$$

To ensure $\sum_s \psi(s) = 1$ and $0 \le \psi(s) \le 1$, we reparameterize $\psi(s)$ as (Kamakura and Russell 1989):

$$\psi(s) = \exp(\gamma_s) \Big/ \sum_q \exp(\gamma_q). \tag{14}$$

4. Empirical Application

Sample Selection and Descriptive Statistics

We use the liquid detergent category to evaluate our model's performance and to articulate the potential managerial insights that could be generated from the model. We chose this product category because: (1) it has good penetration in the target market; (2) it is a frequently purchased product category, thereby giving us more observations to calibrate our model; (3) it contains a relatively large number of SKUs so that consumer heterogeneity is substantial; (4) consumers often engage in external information search when they shop this category online (Wu 1998); and, finally, (5) it is a well-studied category in the marketing science literature.

There are 56 SKUs in our data set, which were chosen at least once collectively by our panelists. We selected our final universal set of brands based on the following considerations: (1) the data should have a reasonably large number of alternatives so that the effect of consideration set formation can be measured; and (2) the data should not contain SKUs with very small choice shares—to improve the robustness of our estimates. Our final data set consists of the top 11 SKUs according to their choice shares.[5] All statistics hereafter are based on this data set.

Table 1 summarizes the descriptive statistics for our sample. Fifty-six percent of the choices were made from the top 11 SKUs. The switching rate[6] in this category is 20%, which is lower than for offline customers (see Degeratu et al. 2000) but is fairly high for online grocery shoppers. Fifty-seven percent of the purchases were made using personal lists, and 13% of the purchases were made after an attribute search. Generally, consumers prefer to shop from personal lists instead of shopping from the entire category.

Calibration and Validation of a Single-Segment Model

One hundred fourteen panelists made at least one purchase, which resulted in 707 choices in total. We randomly selected three-fourths of our total panel members as the calibration sample, with the remaining one-fourth being our validation sample. Because of the limited observations in our data set, one draw of the calibration sample may not be reliable. Therefore, we replicate our calibration 10 times by independently drawing calibration samples. The average numbers of panelists and choices in the calibration sample for the 10 replicates are 83.60 and 529.60, respectively.

We use cross-sample validation to assess the predictive validity of our model. To assess our model's (FCSM) calibration and prediction performance, we compare it with models proposed by Guadagni and Little (1983) (G&L) and Bronnenberg and Vanhonacker (1996) (B&V). We assess all three models (FCSM, G&L, and B&V) using two separate test scenarios. First, we calibrate all three models using an identical set of search information. Second, we recalibrate all the models without search information. G&L and B&V operationalize the short-term preference and long-term preference variables differently from FCSM, which may also contribute to differences in model performance. To rule out this possibility, we specify short-term preference and long-term preference variables of

[5] We selected SKUs with a choice share greater than 3%.

[6] A purchase is "switched" if the SKU purchased is different from the SKU purchased on the previous occasion for the panelist.

Table 1 Descriptive Statistics for Liquid Detergent

	Marketing Mix Variables		
Alternative	Price (std)	Promotion (std) Binary Variable (0, 1)	Choice Share (%)
All Free-N-Clear (1 gallon)	5.57 (0.67)	0.14 (0.34)	3.41
All w/Color Safe Bleach (64 oz)	3.15 (0.00)	0.00 (0.00)	3.25
Cheer Ultra w/Colorguard (100 oz)	6.18 (0.60)	0.17 (0.38)	3.96
Dreft Ultra (50 oz)	4.95 (0.14)	0.00 (0.00)	3.88
Tide Free Ultra (100 oz)	6.45 (1.21)	0.38 (0.49)	12.45
Tide (50 oz)	4.06 (0.11)	0.00 (0.00)	3.33
Tide Ultra (100 oz)	6.34 (1.19)	0.44 (0.50)	8.17
Tide Ultra (90 oz refill)	6.56 (0.41)	0.02 (0.13)	4.83
Tide Ultra w/Bleach Alt (100 oz)	6.32 (1.09)	0.36 (0.48)	5.08
Wisk Bleaching Action (100 oz)	7.13 (0.82)	0.15 (0.36)	3.25
Wisk Ultra (100 oz)	7.17 (0.82)	0.15 (0.36)	4.44
Total			56.05
SKU Switching Ratio		19.66%	
	Search Variables		
Personal List (binary variable)	0.57 (0.49)		
Sorting (binary variable)	0.13 (0.33)		
	Panel Description		
	No. of Panelists	No. of Observations	
Total Sample	114	707	

Notes. This table summarizes the descriptive statistics of the sample we selected for our model calibration and prediction. Recall that we selected the top 11 SKUs (in terms of choice share) in the liquid detergent category.

G&L and B&V identical to those used in FCSM.[7] The addition of a variable to represent sorting does not alter the relative choice utilities of alternatives in G&L. As a result, this variable will not be identified within G&L, and we do not include it in that model.

Model Calibration and Prediction with Search Information. Table 2 summarizes the average of the calibration and prediction estimates of the three models across 10 replications. In each replication, we take an independent draw with size equal to three-fourths of the total panelists as our calibration sample, with the rest constituting the prediction sample. All three models perform significantly better than the naïve equal probability model in terms of the goodness-of-fit indices, namely, log-likelihood values, ρ^2 and $\bar{\rho}^2$. Because the competing models are not nested, we use AIC, MAIC (modified AIC), and BIC to assess their relative performance. According to all three indices, FCSM outperforms both G&L and B&V. In cross-sample validation, FCSM is the best performer on all criteria we use: log-likelihood, ρ^2, and predicted choice probabilities.

With respect to the average parameter estimates, the conclusions are generally consistent across the three models. First, both short-term preference and long-term preference are significant (at 0.05 level) in all three models. This is not surprising because consumers' switching rate is low (20%). Second, promotion is significant (at 0.05 level), but price is not significant in G&L. This may be explained by

[7] We have also compared our FCSM model with G&L and B&V models by specifying short-term preference and long-term preference variables as specified in the original papers (i.e., loyalty in G&L, recency and intrinsic preference in B&V) under both test scenarios. There are no qualitative differences between those results and the results reported in Tables 2 and 3. FCSM outperforms B&V there as well. We do not report these results because of space limitations—we will mail these results to interested readers.

Table 2 Model Comparison with Search Information

Variable	Parameter	G&L Stage	G&L Estimate	G&L (t-Stat)	B&V Stage	B&V Estimate	B&V (t-Stat)	FCSM Stage	FCSM Estimate	FCSM (t-Stat)
Threshold	τ	—	—	—	Co	6.36	(4.36)	Co	2.56	(4.76)
Price	β_1	—	—	—	Co	0.00	(0.62)	Co	−0.02	(−0.42)
Promotion	β_2	—	—	—	Co	1.48	(2.03)	Co	0.28	(1.48)
Personal List (Plist)	ω_P	Ch	4.11	(10.53)	Co	4.61	(1.71)	Co	−2.80	(2.30)
Sort	ω_S	—	—	—	Co	2.32	(2.25)	Co	−0.26	(2.84)
Short-term Preference (SI)	ω_L	Ch	0.38	(2.18)	Co	1.00	(3.81)	Co	−0.17	(3.29)
Price	α_1	Ch	−0.05	(−0.43)	Ch	−0.02	(−0.37)	Ch	−0.04	(−0.22)
Promotion	α_2	Ch	0.95	(3.14)	Ch	−0.16	(−0.23)	Ch	0.39	(0.74)
Long-term Preference (LI)	α_3	Ch	2.86	(6.29)	Ch	2.90	(11.31)	Ch	2.80	(12.52)
All Free-N-Clear (1 gallon)	c_1	Ch	−0.17	(−0.41)	Ch	0.00	(−0.01)	Ch	−0.50	(−0.94)
All w/Color Safe Bleach (64 oz)	c_2	Ch	−1.06	(−1.91)	Ch	−0.81	(−1.09)	Ch	−0.94	(−1.34)
Cheer Ultra w/Colorguard (100 oz)	c_3	Ch	−0.28	(−0.73)	Ch	0.03	(0.06)	Ch	−0.17	(−0.41)
Dreft Ultra (50 oz)	c_4	Ch	0.64	(1.76)	Ch	0.73	(1.18)	Ch	0.73	(1.28)
Tide Free Ultra (100 oz)	c_5	Ch	0.68	(2.43)	Ch	0.76	(2.03)	Ch	0.64	(2.67)
Tide (50 oz)	c_6	Ch	0.17	(0.40)	Ch	0.36	(0.81)	Ch	0.40	(0.89)
Tide Ultra (100 oz)	c_7	Ch	0.39	(0.93)	Ch	0.49	(1.27)	Ch	0.42	(1.30)
Tide Ultra (90 oz refill)	c_8	Ch	−1.28	(−2.69)	Ch	−1.20	(−1.97)	Ch	−1.39	(−1.95)
Tide Ultra w/Bleach Alt (100 oz)	c_9	Ch	0.20	(0.62)	Ch	0.27	(0.63)	Ch	0.11	(0.23)
Wisk Bleaching Action (100 oz)	c_{10}	Ch	−0.69	(−1.21)	Ch	−0.47	(−0.78)	Ch	−0.42	(−1.13)
Wisk Ultra (100 oz)	c_{11}	Ch	0	—	Ch	0	—	Ch	0	—

	G&L Calibration	G&L Validation	B&V Calibration	B&V Validation	FCSM Calibration	FCSM Validation
LL	−370.06	−130.80	−331.67	−123.29	−324.69	−114.17
$LL(0)$	−1269.93	−425.39	−1269.93	−425.39	−1269.93	−425.39
ρ^2	0.71	—	0.74	0.71	0.74	0.73
$\bar{\rho}^2$	0.70	—	0.72	—	0.73	—
AIC	770.12	—	701.34	—	687.39	—
MAIC	785.12	—	720.34	—	706.39	—
BIC	834.20	—	783.23	—	768.56	—
No. of Parameters	15	15	19	19	19	19
Average Predicted Choice Probability	—	0.77	—	0.79	—	0.81

Notes. This table presents the average estimates (for both coefficients and t-stats) of 10 independent replicates, sampled with replacement, with average number of panelists = 83.60 and average number of observations = 529.60 in the calibration sample. LL is the final log-likelihood value, $LL(0)$ is the log-likelihood value of the equal probability model. $\rho^2 = 1 - LL/LL(0)$, and $\bar{\rho}^2 = 1 - (LL - K)/LL(0)$, where K = number of parameters estimated. AIC $= -2LL + 2K$, MAIC $= -2LL + 3K$, and BIC $= -2LL + \ln(N)K$, where N = no. of observations. The average predicted choice probability is the average of choice probabilities of the predicted choices. Co = consideration; Ch = choice.

the fact that promotion is almost always accompanied by a temporary price cut in Peapod. Note also that promotion only impacts choice in G&L (G&L does not have a consideration stage) and only impacts consideration in B&V and FCSM. Third, Plist is statistically significant (at 0.05 level) in G&L and FCSM models and marginally significant in B&V.[8] Fourth, the sorting variable has a statistically significant effect (at 0.05 level) on consideration in both B&V and FCSM. These results indicate that, at the consideration stage, consumers' internal and external information searches do reduce the fuzziness of their consideration sets. Finally, the alternative-specific constants for various brands are consistent across the models, with Tide Free Ultra being the most preferred.

Model Calibration and Prediction Without Search Information. Table 3 summarizes the results from

[8] Differences in signs of parameters across the three models are because of differences in their structures.

Table 3 Model Comparison Without Search Information

		G&L			B&V			FCSM		
Variable	Parameter	Stage	Estimate	(t-Stat)	Stage	Estimate	(t-Stat)	Stage	Estimate	(t-Stat)
Threshold	τ	—	—	—	Co	−0.04	(−0.02)	Co	2.64	(6.17)
Price	β_1	—	—	—	Co	−4.07	(−0.45)	Co	0.20	(1.34)
Promotion	β_2	—	—	—	Co	3.08	(0.07)	Co	0.24	(1.87)
Personal List (Plist)	ω_P	—	—	—	—	—	—	—	—	—
Sort	ω_S	—	—	—	—	—	—	—	—	—
Short-term Preference (SI)	ω_L	Ch	1.50	(6.30)	Co	1.24	(7.09)	Co	−5.83	(−2.08)
Price	α_1	Ch	−0.05	(−0.43)	Ch	4.02	(0.43)	Ch	−0.18	(−1.30)
Promotion	α_2	Ch	0.86	(4.33)	Ch	2.25	(2.02)	Ch	0.80	(4.43)
Long-term Preference (LI)	α_3	Ch	3.90	(7.91)	Ch	4.27	(17.77)	Ch	4.06	(13.05)
All Free-N-Clear (1 gallon)	c_1	Ch	0.10	(0.15)	Ch	0.01	(0.02)	Ch	−0.13	(−0.23)
All w/Color Safe Bleach (64 oz)	c_2	Ch	−0.53	(−0.64)	Ch	−0.55	(−0.87)	Ch	−0.57	(−1.16)
Cheer Ultra w/Colorguard (100 oz)	c_3	Ch	0.02	(0.03)	Ch	−0.02	(−0.06)	Ch	0.04	(0.14)
Dreft Ultra (50 oz)	c_4	Ch	1.06	(3.18)	Ch	0.95	(1.78)	Ch	0.70	(1.99)
Tide Free Ultra (100 oz)	c_5	Ch	0.81	(2.07)	Ch	0.67	(2.07)	Ch	0.60	(2.38)
Tide (50 oz)	c_6	Ch	0.33	(0.60)	Ch	0.35	(0.89)	Ch	0.01	(0.03)
Tide Ultra (100 oz)	c_7	Ch	0.44	(1.55)	Ch	0.32	(0.91)	Ch	0.31	(0.74)
Tide Ultra (90 oz refill)	c_8	Ch	−0.12	(−0.31)	Ch	−0.17	(−0.43)	Ch	−0.47	(−1.12)
Tide Ultra w/Bleach Alt (100 oz)	c_9	Ch	0.30	(0.68)	Ch	0.14	(0.36)	Ch	0.10	(0.40)
Wisk Bleaching Action (100 oz)	c_{10}	Ch	−0.57	(−1.31)	Ch	−0.71	(−1.26)	Ch	−1.03	(−1.64)
Wisk Ultra (100 oz)	c_{11}	Ch	0	—	Ch	0	—	Ch	0	—
			Calibration	Validation		Calibration	Validation		Calibration	Validation
LL			−502.46	−175.92		−443.21	−164.60		−433.41	−149.05
$LL(0)$			−1269.93	−425.39		−1269.93	−425.39		−1269.93	−425.39
ρ^2			0.60	—		0.65	0.61		0.66	0.65
$\bar{\rho}^2$			0.59	—		0.64	—		0.65	—
AIC			1032.92	—		924.42	—		900.82	—
MAIC			1046.92	—		943.42	—		917.82	—
BIC			1092.68	—		1006.32	—		973.59	—
No. of Parameters			14	14		17	17		17	17
Average Predicted Choice Probability			—	0.73		—	0.75		—	0.77

Note. This table presents the average estimates using the same 10 calibration and prediction samples as in Table 2.

the calibration and prediction samples for the three models, with *search information excluded*. In general, we obtain similar results as in Table 2. FCSM outperforms G&L and B&V in all the goodness-of-fit measures in model calibration and with respect to all the measures in cross-sample model prediction. Long-term preference in these models has even greater impact (higher parameter values), compared with models in which the search information was incorporated.

Overall, the model comparisons summarized in Tables 2 and 3 indicate that FCSM outperforms the two competing models not only because it incorporates search information, but also because of its model structure. To understand better what contributes to the differences in performance between B&V and FCSM, we examine the consideration set membership functions for the two models. In B&V, the membership function is given by:

$$m = CDF_{\text{logistic}}\big(\beta_0 + \beta_1 \,\text{Price} + \beta_2 \,\text{Promotion} + \beta_3 \,\text{Plist} + \beta_4 \,\text{Sort} + \beta_5 \,\text{SI}\big), \quad (15)$$

whereas in FCSM, it is given by:

$$m = CDF_{\text{normal}}\left(\frac{\beta_0 + \beta_1 \,\text{Price} + \beta_2 \,\text{Promotion}}{\exp(\beta_3 \,\text{Plist} + \beta_4 \,\text{Sort} + \beta_5 \,\text{SI})}\right). \quad (16)$$

If we take the first-order approximation of the exponential function, we can rewrite (16) as

$$m = CDF_{\text{normal}}\Big(\gamma_0 + \gamma_1 \,\text{Price} + \gamma_2 \,\text{Promotion} + \gamma_3 \,\text{Plist} + \gamma_4 \,\text{Sort} + \gamma_5 \,\text{SI} + \sum \text{ two-way interaction terms}\Big). \quad (17)$$

Comparing (15) and (17), we can observe two important differences between B&V and FCSM. First, the underlying distribution assumptions are different. The logistic distribution used in B&V with a fixed variance $\pi/\sqrt{3}$ has flatter tails than the standard normal distribution. Table 2 shows that the fuzzy spread in our model is smaller than or equal to 1, because the estimated coefficients of Equation (9) are all negative. Hence, the logistic distribution may not be ideal for handling choices in which consideration membership is not very fuzzy. Second, FCSM takes into account not only the linear effects of the independent variables as does B&V, but also the interactive effects among the independent variables, which B&V does not. Indeed, in the Peapod data set, such interactions do exist. For instance, on average, consumers who shopped from their personal lists paid \$5.96 per 100 oz, and 24% of their purchases were on promotion, whereas those who shopped from category lists paid \$5.74 per 100 oz, and 37% of these purchases were on promotion. Likewise, on average, consumers who made purchases after sorting paid \$5.09 per 100 oz, and 30% of their purchases were on promotion, whereas those who purchased without sorting paid \$5.94 per 100 oz, and 26% of these purchases were on promotion. Thus, taken separately, the use of personal list is associated with lower price/promotion sensitivity, but the use of sorting is associated with higher price/promotion sensitivity. Jointly, they both reduce the fuzziness of consumers' consideration sets.

Search significantly improves the fit of all three models. For example, using a likelihood ratio test, the FCSM model performs significantly better with search information than without search information ($\chi^2(2) = 217.44$; $p < 0.01$). Similar results hold for the other two models. Therefore, in general, incorporating information about the choice process as an intervening variable between stimulus (choice context) and response (choices) will improve the performance of choice models.

Consideration Set Size. We use the FCSM model and the formula in Propositions 3 and 4 (see Technical Appendix) to investigate consideration set sizes of consumers. More specifically, we want to understand: (1) how consideration set sizes are distributed across consumers, (2) how consumers use their personal lists and sorting in online stores to reduce fuzziness, and (3) how the use of personal lists and sorting influences consideration set formation.

Table 4 represents the normalized membership distribution by consideration set sizes (computed as per Propositions 3 and 4) in our prediction sample, classified by: (1) all consumers, (2) those using personal list only, (3) those using sorting only, and (4) those not using either personal list or sorting. There are two interesting conclusions. First, we note that, although both personal list (Plist) and sorting (Sort) significantly reduce the fuzziness of consideration, their effects on the sizes of the consideration sets are different. Personal lists shrink the consideration sets, whereas sorting enlarges the consideration sets. Second, we also calculated the consideration size distribution from the model with and without using search information. Without the search information, we would substantially overestimate consideration set sizes.

A Multiple-Segment FSCM Model

The single-segment model assumes that consumers are homogeneous with regard to both marketing mix responsiveness and information search behavior. Suppose that there truly are different groups of consumers in our sample, then the differences in their underlying behaviors will be averaged out in the single-segment model. To investigate the heterogeneity of consumers' consideration and choice behaviors, we estimated a multiple-segment latent mixture model.

Calibration and Validation. Table 5 summarizes our findings.[9] To determine whether a two-segment model improves fit over a single-segment model,

[9] We also applied a two-segment B&V model to this data. The two-segment FCSM model performs better in terms of both goodness-of-fit and predicted choice probabilities. To conserve space, we are not reporting the two-segment B&V results here.

Table 4 Normalized Membership Distribution of Consideration Set Sizes

	Model with Search Information				Model Without Search Information			
Consideration Set Size	All	Plist Only	Sort Only	Neither	All	Plist Only	Sort Only	Neither
1	0.48	0.63	0.22	0.26	0.05	0.05	0.05	0.05
2	0.26	0.34	0.13	0.15	0.16	0.16	0.16	0.17
3	0.07	0.03	0.09	0.12	0.23	0.23	0.23	0.24
4	0.05	0.00	0.09	0.11	0.20	0.20	0.20	0.21
5	0.04	0.00	0.10	0.09	0.12	0.12	0.12	0.13
6	0.03	0.00	0.09	0.07	0.06	0.06	0.06	0.06
7	0.02	0.00	0.07	0.05	0.02	0.02	0.02	0.02
8	0.01	0.00	0.05	0.04	0.01	0.01	0.01	0.01
9	0.01	0.00	0.04	0.03	0.01	0.01	0.01	0.01
10	0.02	0.00	0.04	0.04	0.00	0.00	0.00	0.00
11	0.02	0.00	0.08	0.04	0.12	0.12	0.13	0.11
Mean	2.45	1.40	4.71	4.02	4.52	4.54	4.60	4.40

Notes. This table shows the distribution of consideration set sizes. It is calculated from the single-segment FCSM model with 10 independent replicates for each model. The table provides several interesting insights. First, the (inferred) consideration set sizes are smaller using observed search information (2.45) than without the search information (4.52). Second, with the search information, we find that personal lists (Plist) shrink consideration sets (1.40), whereas sorting (Sort) enlarges the consideration sets (4.71). Note that these differential effects of Plist and Sort cannot be determined without including search data in the model. Finally, note also that, on those occasions, consumers did not use either personal list or sorting; their consideration set sizes from both models are close (4.02 versus 4.40), which is what we should expect. (For the FCSM model without search information, we did not use Plist and Sort in model estimation or prediction, but we do know the purchase occasions during which the consumers used personal list or sorting.) All = results based on all observations in the prediction sample; Plist only = results based only on observations in the prediction sample in which consumers used the personal list feature; Sorting only = results based only on those observations in the prediction sample in which consumers used some type of sorting operation; and Neither = results for those observations in the prediction sample in which consumers did not use either personal list or sorting.

the literature suggests different heuristics. For example, Kamakura and Russell (1989) and Bronnenberg and Vanhonacker (1996) use AIC, whereas Bucklin and Gupta (1992) use BIC. We calculated AIC, BIC, and also MAIC (see Wedel and Kamakura 1998, Desarbo and Wu 2001) and obtained mixed results. AIC and MAIC indicate that the two-segment model has a better fit than the single-segment model in model calibration, but BIC indicates the opposite. We conclude that there is some marginal heterogeneity in our data. Because of limited observations in our data, we decided to retain the two-segment solution and not seek solutions with more segments. The relative sizes of the two segments are 0.43 (Segment 1) and 0.57 (Segment 2)[10] in our calibration sample. In terms of the predictive validity of the two-segment model, the predicted log-likelihood increased from −205.65 for the single-segment model to −154.70 for the two-segment model, and the average predicted choice probabilities increased from 0.72 to 0.75.

Parameter Estimates. In the choice stage, the two segments are similar in terms of their marketing mix responsiveness (i.e., no significant differences). However, the segments differ at the consideration stage. First, the personal list has no impact on Segment 1 but reduces fuzziness in Segment 2. Sorting reduces fuzziness of consideration in Segment 1 but has no significant impact on Segment 2. Second, we find that consumers in Segment 1 are also price-sensitive, whereas, interestingly, promotion seems to have a negative effect on purchase probability in Segment 2 (less price-sensitive consumers). It seems that panelists in the two segments search for information in very different ways. Consumers in Segment 1 actively search

[10] We reparameterize the relative sizes as described in the footnote to Table 5.

Table 5 Two-Segment FCSM Model Calibration and Validation Results

			Common		Segment 1 (Seekers)		Segment 2 (Nonseekers)	
Variable	Parameter	Stage	Estimate	(t-Stat)	Estimate	(t-Stat)	Estimate	(t-Stat)
Threshold	τ	Co			1.94	(1.76)	1.64	(1.60)
Price	β_1	Co			−0.34	(−2.93)	−0.09	(−0.64)
Promotion	β_2	Co			0.49	(1.02)	−1.18	(−2.15)
Plist	ω_P	Co			2.06	(1.04)	−1.75	(−2.60)
Sort	ω_S	Co			−0.25	(−2.19)	0.48	(1.73)
Short-term Preference (SI)	ω_L	Co			−0.22	(−2.30)	1.32	(1.98)
Price	α_1	Ch			−1.62	(−1.56)	−0.27	(−0.82)
Promotion	α_2	Ch			1.05	(0.56)	1.53	(1.52)
Long-term Preference (LI)	α_3	Ch			4.26	(3.13)	3.38	(4.33)
All Free-N-Clear (1 gallon)	c_1	Ch	−0.78	(−1.07)				
All w/Color Safe Bleach (64 oz)	c_2	Ch	−2.35	(−2.21)				
Cheer Ultra w/Colorguard (100 oz)	c_3	Ch	−0.09	(−0.18)				
Dreft Ultra (50 oz)	c_4	Ch	0.33	(0.43)				
Tide Free Ultra (100 oz)	c_5	Ch	0.94	(2.39)				
Tide (50 oz)	c_6	Ch	0.11	(0.21)				
Tide Ultra (100 oz)	c_7	Ch	0.64	(1.52)				
Tide Ultra (90 oz refill)	c_8	Ch	−1.84	(−2.74)				
Tide Ultra w/Bleach Alt (100 oz)	c_9	Ch	−0.07	(−0.14)				
Wisk Bleaching Action (100 oz)	c_{10}	Ch	−1.24	(−1.89)				
Wisk Ultra (100 oz)	c_{11}	Ch	0	—				
Relative Size (reparameterized)	γ				0.27	(1.05)	—	—
			Two-Segment Model				Single-Segment Model	
			Calibration	Validation			Calibration	Validation
LL			−313.05	−154.70			−338.21	−205.65
$LL(0)$			−1292.47	−402.85			−1292.47	−402.85
ρ^2			0.76	0.62			0.74	0.49
$\bar{\rho}^2$			0.74				0.72	
AIC			684.10	—			714.41	—
MAIC			713.10	—			733.41	—
BIC			808.50	—			795.92	—
No. of Parameters			29	29			19	19
Average Predicted Choice Probability			—	0.75			—	0.72

Notes. This table shows the calibration and prediction results for our fuzzy set model with a latent mixture specification. We estimated both the two- and single-segment model using the same calibration sample (91 panelists and 539 choices) and prediction sample (23 panelists and 168 choices). The relative sizes of Segments 1 and 2 ($\psi(1) = 1/(1+e^{\gamma})$ and $\psi(2) = 1 - \psi(1)$) are 0.43 and 0.57, respectively. For other notations see Table 2. The statistics for the single-segment model differ from those reported in Table 2, because a completely new random draw was made for calibrating the two-segment model. Because of computational demands, we made only one replication for the two-segment model. The t-stats reported are asymptotic.

information using their internal information (previous purchases) and external information (sorting). This search behavior reduces fuzziness of alternatives and also increases price/promotion sensitivity in consumers' consideration set formation. Consumers in Segment 2 do not actively seek external information but only rely on internal information (primarily personal list) to form their consideration sets. This search behavior also reduces the fuzziness of alternatives but decreases price/promotion sensitivity. We call those in Segment 1 information "seekers" and in Segment 2 the "nonseekers."

Consideration Set Sizes. Seekers and nonseekers have different information search patterns. Consequently, they form their consideration sets differently.

Table 6 Normalized Membership Distribution of Consideration Set Sizes of the Two-Segment FSCM Model

Consideration Set Size	Segment 1 (Seekers)				Segment 2 (Nonseekers)			
	All	Plist Only	Sort Only	Neither	All	Plist Only	Sort Only	Neither
1	0.12	0.17	0.00	0.07	0.43	0.47	0.32	0.45
2	0.29	0.32	0.00	0.31	0.35	0.39	0.33	0.33
3	0.22	0.18	0.00	0.32	0.10	0.09	0.19	0.07
4	0.24	0.22	0.50	0.22	0.03	0.02	0.09	0.02
5	0.11	0.10	0.42	0.06	0.02	0.01	0.04	0.02
6	0.01	0.01	0.07	0.00	0.01	0.01	0.02	0.02
7	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.03
8	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.03
9	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.02
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean	2.97	2.80	4.57	2.90	2.15	1.91	2.26	2.33

Notes. This table shows the consideration set size distribution for Segment 1 ("Seekers") and Segment 2 ("Nonseekers"). On average, seekers consider more alternatives than nonseekers. Seekers also maintain larger personal lists than nonseekers. Seekers use personal list and sorting more efficiently than nonseekers. These results suggest that seekers have a higher information-processing capacity and/or lower information search costs, which allows them to process more external information during the search and, hence maintain larger consideration sets.

Table 6 represents the normalized membership distributions of consideration set sizes for the two segments. There are several interesting conclusions. First, the seekers consider more SKUs (2.97) than the nonseekers (2.15). Second, seekers maintain a bigger personal list[11] (2.80) than do nonseekers (1.91). Third, seekers utilize sorting to expand their consideration sets (4.57 on average versus 2.26 on average for nonseekers). Thus, it appears that seekers are better able to process information than nonseekers, a hypothesis suggested by Alba and Hutchinson (1987), but not tested empirically before.

In summary, our empirical analysis shows that the fuzzy set model (FSCM) outperforms the competing models of Guadagni and Little (1983) and Bronnenberg and Vanhonacker (1996) in two respects. First, our model fits the empirical data better and demonstrates higher cross-sample predictive validity. Second, unlike existing choice models, our model can also help understand the impact of the external information search on fuzziness reduction and, consequently, on consideration set formation. We find that consumers engage in search in different ways (at least in the product category under investigation). Some consumers (seekers) search both internal and external information extensively, and the information search reduces the fuzziness of alternatives and expands consideration sets. Others (nonseekers) simply do not search that much and make choices based more on their intrinsic established preferences.

5. Conclusion

Behavioral scientists have long questioned the adequacy of the simple one-stage logit model as an adequate representation of consumer choice process. At the same time, human decision making is much too complex to be fully captured in parsimonious models. A promising approach for understanding consumer choice is a two-stage model that incorporates how consumers' search processes influence the brands they consider.

The contributions of this paper can be summarized along three dimensions:

(1) Theoretically, we developed a two-stage choice model grounded in fuzzy set theory, which offers a flexible analytical tool for modeling the consideration set formation process. Our axiomatic framework

[11] Recall that Plist is constructed according to consumers' actual personal list usage. The list size is based on this constructed personal list.

generalizes several previous modeling approaches (e.g., Fortheringham 1988, Bronnenberg and Vanhonacker 1996). Our model also bridges the gap between the theoretical construct of consideration and its current operationalizations.

(2) Methodologically, we offered a general framework that can be operationalized in different ways, depending on the choice process that is deemed to be most relevant in a particular situation. Our model allows us to understand how internal and external information acquisition influence the formation of consideration sets. We described how to operationalize the model in the context of an online supermarket and concluded that, for this context, our model had superior performance over competing models. Our framework is attractive because it does not impose the computational burden associated with other two-stage models, such as Andrews and Srinivasan (1995), allowing us to apply the model even in the presence of a large number of alternatives. To assess more fully the validity and generalizability of our modeling effort, we need to test it with data from other product categories and other online markets, tasks we leave for future research.

(3) Managerially, our model can be used to understand the impact of the information search on consideration sets of consumers. This has important implications for online marketers. A website does not necessarily increase or decrease consideration set sizes of its customers. Rather, various features at a website (e.g., profiling, personal lists, search and sort, hyperlinks, etc.) have different effects; some expand consideration sets and others shrink them. In Peapod, the personal list feature shrinks consideration sets, and the sort feature expands them. This suggests that, for encouraging repeat purchases, an online marketer could incorporate features that reduce consideration set sizes, and for identifying cross-sell and up-sell opportunities, the marketer could incorporate features that expand consideration sets. Furthermore, using a latent mixture model, or other approaches to segmentation, managers can identify segments of consumers who may differ in how various marketing mix elements and online store features influence information-processing capabilities of consumers: Which sorting operations most influence consumer choice? Do certain forms of information presentation lead to larger consideration sets for consumers? Our model can provide answers to such questions,[12] which previous models have not been able to address.

Like any other research study of this scope, our study has several limitations that suggest interesting questions for future research.

First, we used longitudinal panel data from online stores to conduct empirical analysis on consideration set formation. As the first study to use this type of data, we faced some limitations. In particular, we had a small number of observations for each panelist and therefore could not use separate data periods for initializing and calibrating the long-term preference variable without significant loss of information. As a consequence, the estimated choice probabilities are partly endogenous (all three models have the same problem, and therefore endogeneity should only minimally affect model comparisons). As online data collection procedures become more routinized (as was the case with scanner panel data), this problem should disappear, and we should be able to explore the generalizability of our findings in other online markets.

Second, we do not explore how our results from electronic markets generalize to traditional markets (see, e.g., Degeratu et al. 2000). Do the insights from this study also apply in offline markets? How do consumers search information to reduce fuzziness of consideration when making purchases in traditional stores? Peapod is a market pioneer, and, therefore, we should expect that its panel members will not necessarily have choice behaviors that are representative of the general population.

Third, it would be worthwhile to explore whether the fuzzy set consideration model offers any advantages in situations in which traditional crisp set models are seemingly adequate (e.g., Roberts and Lattin 1991, Andrews and Srinivasan 1995, Siddarth et al. 1995). Systematic comparisons between these two approaches to consideration set modeling would benefit both academics and practitioners.

[12] For example, online retailers may wish to develop customized shopping interfaces so that consumers in different segments could have different shopping environments.

Finally, if we have individual-level background information (e.g., demographic data) for the panelists, we can conduct a posterior analysis (e.g., Bucklin and Gupta 1992, Desarbo and Wu 2001) or a concomitant variable latent structure analysis (e.g., Gupta and Chintagunta 1994). We can then investigate the impact of each panelist background variable on the prior probability of membership in each market segment (e.g., by linking Equation (13) to the background variables).

Acknowledgments

This paper is based on the doctoral dissertation of the first author. The authors thank the participants of seminar audiences at Boston University, Hong Kong University of Science and Technology, Institut Européen d'Administration des Affaires, Tulane University, University of Texas at Dallas, University of Texas at Austin, and University of Wisconsin for their helpful comments. The authors thank Marci Hughes and Tim Dorgan of Peapod, Inc., for their help in making available the data sets used in this research, and Professor Gary L. Lilien for his encouragement, support, and insightful comments. Finally, the authors thank the editor, area editor, and two reviewers for their helpful comments, which have significantly improved the presentation of their research and their results.

Technical Appendix

Proof of Proposition 1. Suppose that the fuzzy consideration utility ν_i of brand i is given by (7), and the fuzzy threshold T is given by (8), then ν_i and T are two fuzzy real numbers. Therefore,

$$\begin{aligned} m_i &= \text{degree } (i \in \text{ consideration set}) \\ &= \text{degree } (T \text{ is strictly smaller than } v_i) = \rho(T, v_i), \end{aligned} \quad \text{(A1)}$$

where fuzzy order relation ρ is defined in Equation (2). Let $1/(\sqrt{2\pi}) \int_{-\infty}^{t} e^{-x^2/2}\, dx = \Phi(t)$, then we must have

$$\begin{aligned} T(-\infty, t) &= \frac{1}{\sqrt{2\pi}\delta_T} \int_{-\infty}^{t} e^{-((x-\tau)^2/2\delta_T^2)}\, dx \\ &= \Phi\left(\frac{t-\tau}{\delta_T}\right), \\ \nu_i(t, \infty) &= \frac{1}{\sqrt{2\pi}\delta_{\nu i}} \int_{t}^{\infty} e^{-((x-\bar{\nu}_i)^2/2\delta_{\nu i}^2)}\, dx \\ &= 1 - \Phi\left(\frac{t-\bar{\nu}_i}{\delta_{\nu i}}\right). \end{aligned}$$

Thus, $\rho(T, \nu_i) = \sup_{t\in\Re} T(-\infty, t) \wedge \nu_i(t, \infty)$ is attained at that point $t \in \Re$ where $\Phi\big((t-\tau)/\delta_T\big) = 1 - \Phi\big((t - s\bar{\nu}_i)/\delta_{\nu i}\big)$. Since for any $t \in \Re$, $\Phi(-t) = 1 - \Phi(t)$, and $\Phi(t)$ is a strictly increasing function, it follows that $(t-\tau)/\delta_T = -(t-\bar{\nu}_i)/\delta_{\nu i}$, i.e., $t = (\tau\delta_{\nu i} + \bar{\nu}_i\delta_T)/(\delta_{\nu i} + \delta_T)$. Consequently,

$$\rho(T, \nu_i) = \Phi\left(\frac{\bar{\nu}_i - \tau}{\delta_{\nu i} + \delta_T}\right). \quad \text{(A2)}$$

From (A2), we see that if $\delta_{\nu i} + \delta_T \to 0$, then

$$\rho(T, \nu_i) \to \begin{cases} 1 & \text{if } \bar{\nu}_i > \tau \\ 0 & \text{if } \bar{\nu}_i < \tau; \end{cases}$$

i.e., m_i satisfies Axiom A1. Moreover, if $\delta_{\nu i} + \delta_T \to \infty$, then $\rho(T, \nu_i) \to 1/2$, i.e., m_i satisfies Axiom A2 also. This completes the proof. □

Proof of Proposition 2. If consumers adjust their choice utility additively based on the relative change in degree of consideration after obtaining new information, we have (we omit the alternative index i)

$$f(u, m) = g(u) + h(m/m_0), \quad \text{(A3)}$$

where m_0 is initial degree of consideration before search. By Axiom A4, we must have $g(u) = u - h(1)$.

Consider the first two steps in a consumer's search process (e.g., using personal list, followed by using sorting function, etc.), Steps 1 and 2, during a specific shopping occasion. At Step 1, the consumer changes the degree of consideration from m_0 to m_1, and at Step 2, the consumer changes the degree of consideration from m_1 to m_2. By (A3), for Steps 1 and 2 individually, we must have

$$f(u, m_1) = u - h(1) + h\left(\frac{m_1}{m_0}\right). \quad \text{(A4)}$$

and

$$\begin{aligned} f(u, m_2) &= f(f(u, m_1), m_2) \\ &= \left[u - h(1) + h\left(\frac{m_1}{m_0}\right)\right] - h(1) + h\left(\frac{m_2}{m_1}\right). \end{aligned} \quad \text{(A5)}$$

However, for Steps 1 and 2 jointly, we must also have

$$f(u, m_2) = u - h(1) + h\left(\frac{m_2}{m_0}\right). \quad \text{(A6)}$$

Comparing (A5) and (A6), we have:

$$h\left(\frac{m_2}{m_0}\right) = h\left(\frac{m_1}{m_0}\right) + h\left(\frac{m_2}{m_1}\right) - h(1). \quad \text{(A7)}$$

Note that $m_2/m_0 = (m_1/m_0) \times (m_2/m_1)$. Therefore, (A7) can be rewritten as

$$h(pq) = h(p) + h(q) - h(1), \qquad \forall p, q > 0. \quad \text{(A8)}$$

Assume that $h(\cdot)$ is a differentiable function and take first-order derivatives with respect to p in (A8), we have

$$h'(pq) = (1/q)h'(p), \qquad \forall p, q > 0. \quad \text{(A9)}$$

Let $p = 1$, we have (from (A8) and (A9)):

$$h'(q) = (1/q)h'(1), \qquad \forall q > 0. \quad \text{(A10)}$$

Integrating both sides (with respect to q) from 1 to y, we have

$$h(y) - h(1) = h'(1)\ln(y), \qquad \forall y > 0. \quad \text{(A11)}$$

Rename $h'(1)$ as θ, then we have

$$h(y) = h(1) + \theta \ln(y), \qquad \forall y > 0. \tag{A12}$$

Plugging this into (A3), we obtain

$$\begin{aligned} f(u, m) &= [u - h(1)] + [h(1) + \theta \ln(m/m_0)] \\ &= u + \theta \ln m - \theta \ln m_0. \end{aligned} \tag{A13}$$

By Axiom A5, we must also have $\theta > 0$.

This completes the proof. □

Measuring Consideration Set Sizes

We can partition the nonempty subsets in the universal set with J brands into J categories: those subsets of size 1 (Category 1, hereafter), those subsets of size 2 (Category 2, hereafter), …, and those subsets of size J (Category J, hereafter). Let M be the consideration set, then M will be nonempty by definition because at least one alternative (i.e., the chosen alternative) must be in the consideration set. If M is crisp, then M will belong to one of the categories fully (i.e., M will belong to one of these categories with degree 1 and to the others with degree 0). However, If M is fuzzy, then M may belong to these categories partially, or more formally, M belongs to the categories with different degrees. Then the question is: How are these degrees determined?

Recall that m_j represents the degree of consideration of brand j (i.e., m_j is the degree of brand j belonging to the consideration set M). We first take a look at the degree of M belonging to Category 1. There are J subsets with size 1 (i.e., $\{1\}, \{2\}, \ldots, \{J\}$). Therefore, the degree of M belonging to Category 1 is equal to the sum of the degrees of M belonging to each of the J subsets. We define the degree of M belonging to $\{j\}$ as the joint degree[13] of brand j belonging to M and the rest not belonging to M, i.e.,

$$\begin{aligned} &m(M \text{ belonging to } \{j\}) \\ &\quad = (1-m_1)\cdots(1-m_{j-1})m_j(1-m_{j+1})\cdots(1-m_J). \end{aligned} \tag{A14}$$

Thus we have

$$m(M \text{ belonging to Category 1}) = \sum_{j=1}^{J} m(M \text{ belonging to } \{j\}). \tag{A15}$$

In the same way, $\forall\ 1 \le j < k \le J$, we can define

$$\begin{aligned} &m(M \text{ belonging to } \{j, k\}) \\ &\quad = (1-m_1)\cdots(1-m_{j-1})m_j(1-m_{j+1})\cdots(1-m_{k-1})m_k \\ &\qquad \cdot (1-m_{k+1})\cdots(1-m_J) \end{aligned} \tag{A16}$$

[13] The joint degree is formed in the same way as joint probability, because we use measures of fuzzy mean and spread that are the same as the first two moments of a probability distribution (see §2). This definition is intuitively appealing. For instance, if M fully belongs to $\{k\}$, where k is not equal to j, then M does not belong to $\{j\}$ at all.

and

$$\begin{aligned} &m(M \text{ belonging to Category 2}) \\ &\quad = \sum_{1 \le j < k \le J} m(M \text{ belonging to } \{j, k\}). \end{aligned} \tag{A17}$$

Finally, we have

$$m(M \text{ belonging to Category } J) = \prod_{j=1}^{J} m_j. \tag{A18}$$

An immediate property of this definition of the degree to which M belongs to different categories can be stated as follows:

Proposition 3.

$$\sum_{j=1}^{J} m(M \textit{ belonging to category } j) = 1 - \prod_{j=1}^{J} (1-m_j). \tag{A19}$$

Proof. Note that

$$\begin{aligned} 1 &= ((1-m_1)+m_1)((1-m_2)+m_2))\cdots((1-m_J)+m_J)) \\ &= \prod_{j=1}^{J}(1-m_j) + \sum_{j=1}^{J}(1-m_1)\cdots(1-m_{j-1})m_j(1-m_{j+1})\cdots(1-m_J) \\ &\quad + \sum_{j<k}(1-m_1)\cdots(1-m_{j-1})m_j(1-m_{j+1})\cdots(1-m_{k-1}) \\ &\quad \cdot m_k(1-m_{k+1})\cdots(1-m_J) + \cdots + \prod_{j=1}^{J} m_1 m_2 \cdots m_J \\ &= \prod_{j=1}^{J}(1-m_j) + m(\text{C belonging to Category 1}) \\ &\quad + m(\text{C belonging to Category 2}) \\ &\quad + \cdots + m(\text{C belonging to Category } J) \\ &= \prod_{j=1}^{J}(1-m_j) + \sum_{j=1}^{J} m(\text{C belonging to Category } j). \end{aligned}$$

This completes the proof. □

Proposition 4. *If we normalize the membership function of brands belonging to M by $\bar{m}_j = m_j/m_c$, where c is the chosen brand, then the degree of brand c belonging to consideration set M is equal to* 1; *i.e., the chosen brand is always fully considered. For the normalized membership $\bar{m}_j$, the product term in Proposition* 3 *is* 0. *Thus, we have*

$$\sum_{j=1}^{J} \bar{m}(M \textit{ belonging to Category } j) = 1.$$

Proof. Obvious. □

References

Abramson, Charles, Rick L. Andrews, Imran S. Currim, Morgan Jones. 2000. Parameter bias from unobserved effects in the multinomial logit model of consumer choice. *J. Marketing Res.* **XXXVII**(November) 410–426.

Alba, Joseph W., John Lynch, Barton Weitz, Chris Janiszewski, Richard Lutz, Alan Sawyer, Stacy Wood. 1997. Interactive home shopping: Consumers, retailer, and manufacturer incentives to participate in electronic marketplaces. *J. Marketing* **61**(July) 38–53.

——, J. Wesley Hutchinson. 1987. Dimensions of consumer expertise. *J. Consumer Res.* **13**(March) 411–454.

Andrews, Rick L., T. C. Srinivasan. 1995. Studying consideration effects in empirical choice models using scanner panel data. *J. Marketing Res.* **32**(February) 30–41.

Bettman, James R. 1979. *An Information Processing Theory of Consumer Choice*. Addison-Wesley, Reading, MA.

Bronnenberg, Bart J., Wilfried R. Vanhonacker. 1996. Limited choice sets, local price response, and implied measures of price competition. *J. Marketing Res.* **33**(May) 163–173.

Bucklin, Randolph E., Sunil Gupta. 1992. Brand choice, purchase incidence, and segmentation: An integrated modeling approach. *J. Marketing Res.* **XXIX**(May) 201–215.

——, James M. Lattin. 1991. A two-stage model of purchase incidence and brand choice. *Marketing Sci.* **10** 24–39.

Chiang, Jeongwen, Chib Siddhartha, Narasimhan Chakravarthi. 1999. Markov chain Monte Carlo and models of consideration set and parameter heterogeneity. *J. Econometrics* **89**(March/April) 223–248.

Degeratu, Alexandru, Arvind Rangaswamy, Jianan Wu. 2000. Consumer choice behavior in online and regular stores: The effects of brand name, price, and other search attributes. *Internat. J. Res. Marketing* **17**(1) 55–78.

Desarbo, Wayne S., Donald R. Lehmann, Gregory Carpenter, Indrajit Sinha. 1996. A stochastic multidimensional unfolding approach for representing phased decision outcomes. *Psychometrika* **61**(3) 485–508.

——, Jianan Wu. 2001. The joint spatial representation of multiple variable batteries collected in marketing research. *J. Marketing Res.* **38**(2) 244–253.

Finn, Adam, Jordan Louviere. 1990. Shopping-center patronage models: Fashioning a consideration set segmentation solution. *J. Bus. Res.* **21** 259–275.

Fortheringham, A. Stewart. 1988. Consumer store choice and choice set definition. *Marketing Sci.* **7**(3) 299–310.

Gensch, Dennis H. 1987. A two-stage disaggregate attribute choice model. *Marketing Sci.* **6**(3) 223–239.

Guadagni, Peter M., John D. C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Sci.* **2**(Summer) 203–238.

Gupta, S., P. K. Chintagunta. 1994. On using demographic variables to determine membership in logit mixture models. *J. Marketing Res.* **31** 128–136.

Haubl, Gerald, Valerie Trifts. 2000. Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing Sci.* **19**(1) 4–21.

Hauser, John R. 1978. Testing the accuracy, usefulness and significance of probabilistic choice models: An information theoretic approach. *Oper. Res.* **26**(3) 406–421.

——, Birger Wernerfelt. 1990. An evaluation cost model of consideration sets. *J. Consumer Res.* **16**(March) 393–408.

——, Glen L. Urban, Bruce D. Weinberg. 1993. How consumers allocate their time when searching for information. *J. Marketing Res.* **30**(November) 452–466.

Kamakura, Wagner A., Gary J. Russell. 1989. A probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* **26**(November) 379–390.

Kardes, Frank R., Gurumurthy Kalyanaram, Murali Chandrashekaran, Ronald J. Dornoff. 1993. Brand retrieval, consideration set composition, consumer choice, and the pioneering advantage. *J. Consumer Res.* **20**(June) 62–75.

Lee, E. S., R. L. Li. 1988. Comparison of fuzzy numbers on the probability measure of fuzzy events. *Computer Math. Appl.* **15** 887–896.

Lowen, R. 1997. *Fuzzy Set Theory: Basic Concepts, Techniques and Bibliography*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Luce, R.D. 1959. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New York.

Lynch, John G. Jr., Howard Marmorstein, Micheal F. Weigold. 1988. Choice from sets including remembered brands: Use of recalled attributes and prior overall evaluations. *J. Consumer Res.* **15**(September) 169–184.

McFadden, Daniel. 1974. Conditional logit analysis of qualitative choice behavior. P. Zarembka, ed. *Frontiers in Econometrics*. Academic Press, New York, 105–142.

Nedungadi, Prakash. 1990. Recall and consumer consideration sets: Influencing choice without altering brand evaluations. *J. Consumer Res.* **17**(December) 263–276.

Roberts, John H., James M. Lattin. 1991. Development and testing of a model of consideration set composition. *J. Marketing Res.* **28**(November) 429–440.

——, ——. 1997. Consideration: Review of research and prospects for future insights. *J. Marketing Res.* **34**(August) 406–410.

——, Glen Urban. 1988. Modeling multiattribute utility, risk, and belief dynamics for new consumer durable brand choice. *Management Sci.* **34**(February) 167–185.

Rust, T. Roland, J. Jeffrey Inman, Jianmin Jia, Anthony Zahorik. 1999. What you don't know about customer perceived quality: The role of customer expectation distributions. *Marketing Sci.* **18**(1) 77–92.

Shocker, Allan D., Moshe Ben-Akiva, Bruno Boccara, Prakash Nedungadi. 1991. Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing Lett.* **2**(3) 181–197.

Siddarth, S., Randolph E. Bucklin, Donald Morrison. 1995. Making the cut: Modeling and analyzing choice set restriction in scanner panel data. *J. Marketing Res.* **32**(August) 255–266.

Swait, Joffrey, Moshe Ben-Akiva. 1987. Incorporating random constraints in discrete models of choice set generation. *Transportation Res. B* **21B**(2) 91–102.

——, Jordan Louviere. 1993. The role of the scale parameter in the estimation and comparison of multinomial logit model. *J. Marketing Res.* **30**(August) 305–314.

——, Craig Stacey. 1996. The role of consumer brand assessment and assessment confidence in models of longitudinal choice behavior. *1996 INFORMS Marketing Sci. Conf.* University of Florida, Gainesville, FL, March 1996.

Tversky, Amos. 1972. Elimination-by-aspects: A theory of choice. *Psych. Rev.* **79**(July) 281–299.

Wedel, Michel, Wagner A. Kamakura. 1998. *Market Segmentation: Conceptual and Methodological Foundations*. ISQM, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Wu, Jianan. 1998. A fuzzy set model of consideration set formation: Theory, methodology, and calibration using data from an online supermarket. Unpublished doctoral dissertation, The Pennsylvania State University, University Park, PA.

This paper was received March 30, 2001, and was with the authors 4 months for 3 revisions; processed by Michel Wedel.