

Elsevier required licence: © 2016

This manuscript version is made available under the CC-BY-NC-ND 4.0 license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

The definitive publisher version is available online at

<https://doi.org/10.1016/j.jretai.2016.07.002>

A Model for Inferring Market Preferences from Online Retail Product Information Matrices

Timothy J. Gilbride

Imran S. Currim

Ofer Mintz

S. Siddarth*

July 2016

Forthcoming in Journal of Retailing

* Timothy J. Gilbride (tgilbrid@nd.edu) is Associate Professor and Notre Dame Chair in Marketing, Mendoza College of Business, University of Notre Dame, Notre Dame, IN 46556. Imran S. Currim (iscurrim@uci.edu) is Chancellor's Professor and Professor of Marketing, Paul Merage School of Business, University of California, Irvine, CA 92697. Ofer Mintz (omintz@lsu.edu) is Assistant Professor of Marketing, E. J. Ourso College of Business, Louisiana State University, Baton Rouge, LA 70803. S. Siddarth (siddarth@usc.edu) is Associate Professor of Marketing, Marshall School of Business, University of Southern California, Los Angeles, CA 90089. The authors thank Internet Technology Group, Inc. (ITGi) for providing the data; seminar participants at the University of Houston, Texas A&M University, INSEAD, INFORMS Marketing Science Conference in Atlanta, Georgia, Direct/Interactive Marketing Research Summit in San Diego, California, and INFORMS International Conference in Waikoloa, Hawaii for their constructive comments; and the UCI Paul Merage School of Business Dean's Office for financial support.

A Model for Inferring Market Preferences from Online Retail Product Information Matrices

Abstract

This research extends information display board methods, currently employed to study information processing patterns in laboratory settings, to a field based setting that also yields managerially useful estimates of market preferences. A new model is proposed based on statistical, behavioral, and economic theories, which integrates three decisions consumers must make in this context: which product-attribute to inspect next, when to stop processing, and which, if any, product to purchase. Several theoretical options are considered on how to model product attribute selection and how to treat uninspected attributes. The modeling options are empirically tested employing datasets collected at a popular e-tailer's website, while customers were making product evaluation and purchase decisions. Subsequent to identifying the best model, we show how the resulting attribute preference estimates can be managerially employed to improve customer targeting of abandoned shopping carts, for follow up communications aimed at improving sales conversions.

Keywords: Choice Models; Information Processing; E-Commerce; Sequential Search; Expected Utility.

A large number of retailer websites (i.e., Apple, Nikon, Dell, Ford, Best Buy, etc.) organize or have a feature that allows customers to organize product-attribute information at the point-of-purchase in the matrix form reminiscent of information display boards (IDBs) widely employed in lab-based information processing research (see Figure 1). In this research, we update the IDB methodology to an online format and propose and estimate an econometric model that infers attribute importance weights from consumers making durable goods purchases at a vertically-integrated e-tailer's website. Product information matrices require shoppers to make a sequence of decisions: (i) which product-attribute to inspect next, (ii) when to stop processing information, and (iii) which, if any, product to purchase. Our model-based, multi-disciplinary approach incorporates statistical, behavioral, and economic theories to estimate attribute importance weights using data from this sequence of three decisions.

Determining the relative importance of product attributes is one of the canonical problems in marketing management. However, as Feit, Beltramo, and Feinberg (2010) discuss, both market and lab based techniques to measure preferences have their shortcomings. Specifically, while market data comes from consumers making actual decisions, it often lacks the variability in attribute levels needed to estimate attribute preference weights. On the other hand, survey and lab based methods such as conjoint analysis provide the necessary variation in product attributes, but produce inconsistencies between predictions and actual market outcomes, suggesting that respondents may not make "hypothetical survey choices exactly as they make purchase decisions" (pp. 785-786). Our research attempts to overcome these shortcomings by showing how IDB data collected by e-tailers outside the lab and with relatively little attribute level variation, can still yield meaningful market preferences for attributes provided consumer decisions are modeled in an integrated manner. We anticipate that the proposed method will be

of most interest to e-tailers vertically integrated with manufacturers. However, the ease of implementing the online IDB make it attractive for other e-tailers to collect this information to collaborate with their suppliers on product design, pricing, and advertising

IDBs have been extensively used to study how consumers process product attribute information (e.g., see Payne, Bettman, and Johnson 1993 and Bettman, Luce, and Payne 1998 for reviews) but have not been used to infer product attribute weights. Yang, Toubia, and De Jong (2015) and Meißner, Musalem, and Huber (2016) have shown how using eye-tracking technology and modeling the information processing of subjects can be used to improve holdout sample predictions in a laboratory based conjoint study. However, they rely on specialized equipment, experimental methods, and do not test their model's predictions with actual market outcomes. The current research uses data from shoppers who made a single visit to an e-tailer's website and models their information processing and choices on an online IDB. The proposed method is not necessarily a substitute for other lab based studies, but it does leverage data that can uniquely be collected at the point-of-purchase by e-tailers.

Consumers can learn about products and their attribute-levels from a variety of sources through both active and passive learning over a period of time. We assume that the cumulative effect of all past learning and search can be summarized at a point in time by a consumer's current preferences for particular attributes and current expected level and range of attribute values in the market. So for instance, someone shopping for a car would know the relative importance (to them) of horsepower, fuel economy, price, etc. and might expect that the fuel economy for subcompact cars to range from 30 to 35 miles per gallon, say. We do not model the consumer's learning process over time or the source of their preferences; we assume that the data to model this would be unavailable to most e-tailers. Rather, we propose a model in which

consumer information processing is driven by their ex-ante expectations of the level and range of attribute values and the costs and benefits of various actions.

The data collected for this work comes from the retail website of a firm that showed shoppers the names of products and attributes, as in the first row and column of Figure 1, but concealed the product-attribute values in the remaining cells. Shoppers then clicked on cells to reveal the product-attribute levels and were given the option of purchasing one of the products. We develop an integrated, utility-based econometric model for the three required IDB decisions, i.e., which cell to open, whether or not to continue acquiring additional information, and which product to buy. To model the decision of which cell to open we must quantify the benefit of hitherto unopened cells. In our base model, the benefit is quantified using a standard expected value approach, which integrates over the potential values that an attribute level might take. We term this as the Expected Value model. However, some researchers (e.g., Miller 1993) have questioned the validity of representing consumer beliefs via parametric distributions, which are essential for expected value calculations. Therefore, we also develop a new, non-parametric approach, which we term as the Max-Min model. This model draws on the work of Jaccard, Knox, and Brinberg (1979) and Jaccard, Brinberg, and Ackerman (1986) by relying only on estimating the end points of the belief distribution. Finally, we propose a third model, termed as the Hybrid model, in which the proximity of an unopened cell also influences the benefit of opening it. These model elements are also incorporated into the decision of when to stop opening cells in the IDB and to move on and make a final product selection.

As repeatedly demonstrated in behavioral work, consumers typically do not inspect all product-attributes available to them in an IDB (e.g., Sheluga, Jaccard, and Jacoby 1979). Are these product-attributes unimportant to the consumer and ignored? Does a consumer rely on a

prior value of this product-attribute? Using the previous literature as a starting point (e.g., Bettman, Luce, and Payne 1998; Branco, Sun, and Villas-Boas 2012; Meyer 1982), we test three plausible approaches to assigning values of uninspected cells: a) using the mean of the actual attribute levels (Actual Mean), b) setting the unseen attribute level to zero, i.e., unseen attributes do not impact the decision (Not Used), and c) estimating the inferred attribute level via a model parameter (Inferred Mean). In Table 1, we summarize the differing model variants for determining product-attribute selection and values of unseen product-attributes.

Over all datasets we find that the combination of the Inferred Mean method for handling unseen attribute levels and the Max-Min method for cell opening works best. Our results include attribute importance weights which would typically only be available from laboratory based conjoint analysis for consumer durable products as well as information on the market's expected value and range of attributes. These parameters can help e-tailers address issues related to new product design, advertising, promotion, and follow up communications. For instance, when the inferred maximum and minimum values of an attribute exceed those actually found in the marketplace this indicates greater uncertainty about this attribute: highlighting this attribute on websites or in follow-up advertisements may help to educate consumers. In addition, we illustrate how the model can be used to target customers who visited the IDB enabled website who were most likely to make a purchase and which item they are more likely to purchase; this provides a method of following up on "abandoned shopping carts," a problem familiar to most e-tailers. We know of no other methods that provide these insights using revealed preference data.

BACKGROUND

The literature on information processing is vast and, in this section, we briefly review only those studies that have a direct bearing on our research. The economic modeling literature takes a

formal cost-benefit approach wherein decision makers compare the expected value of additional information to the cost of acquiring it. Hagerty and Aaker (1984) model how consumers process information by operationalizing the marginal benefit of further search via the expected value of sample information (EVSI) and proposing a stopping rule based on comparing the EVSI to the cost of obtaining additional information. The EVSI is dependent on prior beliefs and the sequence in which information is inspected. Hagerty and Aaker require that consumers' prior expectations for the value of each product's attribute level be described by a multivariate normal distribution. Hagerty and Aaker test their model using parameter estimates obtained from an initial survey and evaluate how well the EVSI model predicts the subsequent IDB information acquisition activity of their subjects. We develop our cost-benefit models based on Hagerty and Aaker, but significantly expand upon their model by permitting our models and empirical implementation to handle both nominal and continuous attributes, and to estimate market level attribute importance from observed behavior alone, without the need to augment with lab data.

Yang, Toubia, and De Jong (2015) use eye-tracking software in an experimental conjoint analysis with repeated observations per person to infer consumer preferences. They find that incorporating the sequence of eye movements, which represent consumer information acquisition, improves hold-out sample fit and reduces the number of profiles that respondents need to rate. Implementing their model on websites will be problematic for e-tailers because 1) it requires all attributes to be categorical, 2) requires consumers to believe that all attribute levels are equally likely, and 3) requires shoppers to use computers with eye-tracking cameras/software installed. Of course their work has the usual benefits of lab-based methods, i.e., the ability to incorporate hypothetical attributes/levels and to capture heterogeneity in attribute preferences. Overall, we see our approach as complementary but distinct from Yang, Toubia, and De Jong

(2015). Our work is distinctive because it relies on a different information acquisition method (IDB vs. eye-tracking), uses product data with very different characteristics (duplicate attribute levels vs. experimental design) that is more applicable to real-world purchasing scenarios, and requires a novel modeling approach. Our work is complementary because it also shows that consumer attribute preferences can be estimated more accurately by tracking and modeling consumers' endogenous information acquisition. We are the first to show that the benefits of attribute preference measurement can be achieved in real-world retail setting using IDB data.

The behavioral information processing literature based on laboratory studies using IDBs primarily focuses on *describing* consumers' information processing and choice behaviors. Among others, Payne, Bettman, and Johnson (1993), Bettman, Luce, and Payne (1998), and Dhar and Nowlis (2004) find that consumers' process information in three basic ways; (1) by alternative; (2) by attribute and (3) a combination of alternative- and attribute-processing. Behavioral researchers have found these patterns to be pervasive, supporting the notion of constructive choice processes, and to vary systematically depending on the stage and type of choice task (Bettman and Park 1980; Biehal and Chakravarti 1986). For example, Biehal and Chakravarti (1986) argue that if memory is "brand-organized" then the initial processing of information will be brand or alternative (column) based as opposed to attribute (row) based. Our models advance this literature by incorporating attribute and alternative based processing into an overall cost-benefit approach.

Other behavioral approaches to information processing have a bearing on our proposed models. Simonson, Huber, and Payne (1988) collect data similar to the method of Hagerty and Aaker (1984) and find that prior certainty, attribute importance, and an overall measure of brand attractiveness have a significant impact on the sequence of information processed. Meyer (1982)

models the probability that information on an alternative will be accessed at a given time, and how expectations with respect to attribute values are updated over the information acquisition sequence. Meyer (1982) tests his descriptive model in two controlled experiments, and unlike our research, does not model consumers' purchasing decision. However, he notes the importance of "assess(ing) the degree to which it (the model) can predict individual behavior in complex settings" (p. 120).

Other research on the same dataset employed for the current study highlights the differences in methodology and the contribution of this research. Mintz, Currim, and Jeliaskov (2013) use a process framework and categorize the degree to which each shopper uses an overall pattern of "attribute" or "alternative" based information acquisition. Mintz, Currim, and Jeliaskov (2013) relate the "pattern" of information acquisition to whether consumers' buy or don't buy one of the products. By comparison, the model primitives in this research are shoppers' preferences and expected level and range of attribute levels, which are revealed in each step of the information acquisition process. While Mintz, Currim, and Jeliaskov (2013) reduce the process to a descriptive scalar measure (e.g. degree of attribute based processing) and does not provide attribute importance weights, this research proposes a cost-benefit/behavioral structure used to infer preferences and market expectations. Currim, Mintz, and Siddarth (2015) describe product choice as a function of the revealed product attributes as opposed to all the information available. They demonstrate that it is important to consider what information is accessed in an IDB when predicting choice. However, since shoppers do not inspect product-attribute levels at random, the product attribute-weights measured by Currim, Mintz, and Siddarth (2015) suffer an endogeneity bias and cannot be relied upon for management decision making. In contrast, the method proposed in this research overcomes the endogeneity bias by

jointly modeling information acquisition and product choice. In summary, neither Mintz, Currim, and Jeliaskov (2013) nor Currim, Mintz, and Siddarth (2015) consider the full range of decisions made by shoppers at the point of purchase and consequently do not provide useful measures of attribute importance.

Our model infers preferences and elements of consumers' prior information using an (updated) IDB methodology. In that prior research has focused almost exclusively on lab based studies, how do we know the IDB method is valid in actual choice situations? First, past research by Johnson, Meyer, and Ghose (1989) and Johnson, Payne, and Bettman (1988) show that subjects display similar information processing, choices, and eye movements compared to subjects who are not using IDB computerized decision process tracers. Thus, there is evidence that requiring shoppers to "click" on information in the IDB does not alter their shopping behavior. Second, our research uses hold-out samples of consumers in actual buying situations and test the predictions of the model against actual shopping behavior. Third, Lehmann and Moore (1980) demonstrated in a longitudinal study using actual purchases (healthy bread) that information acquisition via an IDB conformed to theoretical predictions of how much information would be accessed as well as which specific product-attributes would be inspected. Thus, as argued by Aschemann-Witzel and Hamm (2011), consumers' greater familiarity with these IDB formats and the strong theoretical and laboratory based results suggest field based research using IDB's will yield useful managerial insights.

In summary, our study differs from past research in that it yields attribute importance weights by formulating an integrated model of which attribute to inspect next, when to stop, and which alternative to choose in an actual point-of-purchase situation where product attributes are not experimentally manipulated and prior information about consumers is incomplete.

DATA

We introduce the data collection method at this point as a template for how e-tailers can collect the data and to motivate the econometric model in the next section. The data was collected by Internet Technology Group, Inc. (ITGi) for an online study of consumer behavior at a well-known electronic retail-manufacturer's website. Due to confidentiality agreements, the name or exact nature of the product cannot be revealed. However, the product is a consumer durable with the types and number of features similar to what might be found in a computer, tablet, e-reader, or camera. The data collection procedure is discussed next followed by descriptive statistics.

Data Collection

Unlike previous IDB studies, the data does not come from a lab-based study but represents information processing and purchases of real shoppers. The retailer installed the Decision Board Platform (Mintz et al. 1997) on its website for a consecutive 50 hour period over a weekend. Similar to the classic Mouselab IDB used in many previous lab studies (e.g., Payne, Bettman, and Johnson 1993), this platform recorded shoppers' sequence of product-attribute information acquisition and their final choice. As shown in Figure 2, three products were available in column format with each product's model number and price shown in the first row.¹ Cells containing information on eleven other product attributes for each alternative appeared in corresponding rows below, but these values were concealed until the consumer clicked and revealed the value of that attribute level. Similar to the "Choose" command button in the mock-up, a prominent "Customize and Buy" command button was located at the bottom of each column.² If a customer clicked on the "Customize and Buy" button, they were taken to a secure server where they entered shipping location and credit card information, which for legal reasons is unavailable to

us. No other data on consumers' prior knowledge or demographics was collected by the company, typical of internet-based retail settings.

The data was collected from 895 shoppers who selected one of three distinct price tiers (i.e., high, medium, or low) and were directed to different websites that featured three products and eleven attributes. To control for heterogeneity between these market segments, we analyze these datasets separately. We first present a detailed description and results for the high priced market segment. To underscore the generalizability of the method and results, we also present selected model based results for the other two segments when reviewing the empirical application. Table 2 indicates which attributes were continuous and discrete, and the ordinal ranking of their attribute levels. Every shopper saw the same product-attribute matrix in which products 1 and 3 were the lower-priced alternatives and product 2 had a higher price. Attributes 10a and 10b were revealed together when a cell in row 10 was clicked; these were elements of the physical dimensions of the product (e.g., height and weight), which were not perfectly correlated across the alternatives (i.e., the lightest product was not the shortest). The products shown were real, not hypothetical, and, as a result, attribute-levels for six of the attributes (A2, A4, A6, A7, A8, and A9) were the same for *all* products; further, for several other attributes these levels were the same for two of the three products. Despite these overlaps, a careful inspection reveals that the highest priced alternative, product 2, does not dominate on all the attributes because it is missing the desirable discrete attributes 3 and 11.

Descriptive Statistics

Data from the higher priced market segment consists of point-of-purchase information processing data from 136 shoppers. As shown in the right hand side of Figure 2, the data from each shopper provides the following information: the sequence of product-attribute levels

inspected, the last cell accessed, and which one of the three products or the “no purchase” alternative was chosen. We only include those observations in which the customer accessed more than one cell. The products, attributes, and website layout were chosen by the e-tailer.

As reported in panel A of Table 3, 58 of the 136 shoppers (43%) “customized and bought,” with 27 customers purchasing product 2 (20%), the most expensive product, and 13 and 18 shoppers, respectively, purchasing the two equally priced products 1 (10%) and 3 (13%). Only 7% of shoppers accessed all the information in the Decision Board (panel B lower right corner), with the average shopper accessing 11.54 of the 33 cells (panel A). Those who “customized and bought” accessed more cells than those who did not (13.62 vs. 9.99), however inter-shopper variation was high (standard deviation 9.24). 68% of shoppers accessed information on all three products, 9% for two products and 23% for a single product (not listed in table).

As expected and as reported in panel B of Table 3, shoppers accessed attributes in the top rows more often than those in the bottom rows; A1 was accessed most often (89% of shoppers accessed A1 for at least one of the products, see column labeled “Overall”), A3 second most (70%), and A10 least (40%). 50% of the shoppers accessed five or fewer rows, 15% accessed 6-10 rows, and only 35% of shoppers accessed all 11 rows (not listed in table). Panel B also reports the percentage of shoppers who clicked on each individual cell in the IDB and shows that most shoppers only accessed a subset of the information available to them in what would be considered a relatively high price, high involvement type of purchase.

MODEL

Our model is intended to provide managerially useful estimates of attribute importance from consumers in an actual purchasing situation. We assume that shoppers learn about products from

a variety of sources at different points in time and that this knowledge can be represented by shoppers' attribute importance and their expected level and range of product-attribute values. The product attribute weights $\{\beta\}$ and the expected distribution of product attribute levels and ranges (represented by $f(x_k)$ and $\{Max_k, Min_k\}$) are model primitives that we infer from the data by modeling the information acquisition process including which product-attributes to inspect, when to stop acquiring additional information, and which product, if any, to select.

Base Model

Modeling IDB data presents two major challenges: (1) how to represent the value of acquiring new information and (2) how to handle attribute levels that were not inspected. We begin by describing a model that incorporates an expected value calculation of processing benefits and uses the average values of the actual attribute levels to represent the ex-ante values of product-attributes not inspected. An outline of the estimation procedure is included and full details are provided in the Online Technical Appendix. We then describe alternative models that incorporate other behavioral mechanisms for the two processes.

Consumers process information for and reveal product-attribute levels by clicking on cells in an online point-of-purchase IDB; products are in columns and attributes are in rows. We assume that consumers have ex-ante expectations for the levels of the product-attributes, represented by \bar{x}_{jk} for product j and attribute k . As a consumer inspects product-attributes, she discovers x_{jk} , the actual value of attribute k for product j . Let the vector \tilde{x}_{ij} represent the combination of revealed and ex-ante expected values for product j for person i . Before inspecting any product-attributes $\tilde{x}_{ij} = \bar{x}_j$. When a product-attribute level is revealed, x_{jk} replaces \bar{x}_{jk} in the vector \tilde{x}_{ij} and this vector therefore corresponds to consumer i 's unique information processing

pattern. The vector \tilde{x}_{ij} includes an intercept term that represents the product name or any other additional information that is revealed prior to search, in our case the product name and price.

The first component of the model specifies how consumers make the final choice between product alternatives. As is standard in the discrete choice literature, we assume that the final product decision is based on a linear compensatory indirect utility function. The indirect utility that consumer i expects from product j is given as:

$$V_{ij} = \beta' \tilde{x}_{ij}^f + \varepsilon_{ij}. \quad (1)$$

Here \tilde{x}_{ij}^f represents the final vector of product-attribute levels for product j that consumer i has revealed. The error term ε represents additional uncertainty about any remaining attributes and/or other factors related to the product, e.g., the performance of the product in particular usage situations, durability, quality, etc. If the consumer makes a product choice without inspecting any product-attribute levels at the point-of-purchase, his choice is based on $\beta' \bar{x}_j + \varepsilon_{ij}$. As the consumer sequentially inspects product-attribute levels, indirect utility is given by $\beta' \tilde{x}_{ij} + \varepsilon_{ij}$ at each step of the process.

Let \tilde{X}_i^f represent the final matrix of product-attributes across alternatives, in a particular choice set used by a consumer and ε_i the stacked vector of ε_{ij} . The consumer's ultimate decision problem is to choose the alternative corresponding to $Max(\beta' \tilde{X}_i^f + \varepsilon_i)$. When ε_{ij} is distributed i.i.d. standard extreme value, the expected maximum utility across alternatives is given as (see Anderson, de Palma, and Thisse 1992):

$$E_{\varepsilon} \left[Max(\beta' \tilde{X}_i^f + \varepsilon_i) \right] = \ln \left(\iota' \exp \left[\beta' \tilde{X}_i^f \right] \right) + \lambda, \quad (2)$$

where ι is a vector of ones and λ is Euler's constant.

We now consider the benefit to the consumer of inspecting or processing more information as part of their decision process. A consumer will acquire additional information if it leads to a “better” final decision. In this case, we operationalize “better” as the change in the expected maximum utility, equation (2). At any step of the process, let \tilde{X}_i^n represent the matrix of product-attribute levels currently being used by the consumer after n cells have been opened³ and \tilde{X}_i^{n+} the state of the matrix if one additional cell is opened. The benefit of revealing the contents of this new cell, x_{jk} , is given by $E_\varepsilon [\text{Max}(\beta' \tilde{X}_i^{n+} + \varepsilon_i)] - E_\varepsilon [\text{Max}(\beta' \tilde{X}_i^n + \varepsilon_i)]$, which in turn can be written as:

$$\chi = \ln\left(t' \exp\left[\beta' \tilde{X}_i^{n+}\right]\right) - \ln\left(t' \exp\left[\beta' \tilde{X}_i^n\right]\right). \quad (3)$$

Per equation (3), opening a new cell in the IDB will yield an incremental expected maximum utility, χ . Note that this value can be positive or negative depending on the value of x_{jk} in the opened cell; x_{jk} can be less than that what was expected resulting in a negative value for χ .

Consumers acquire information about products and attribute levels from many different sources at different points in time and bring that information with them to the point-of-purchase. While we do not model that extended search process, if their prior information is complete and known with certainty, we would not see any additional information acquisition at the point-of-purchase. Consistent with economic models of product search (e.g., Moorthy, Ratchford, and Talukdar 1997), we assume that consumers prior knowledge and uncertainty about the values of the product-attribute levels x_{jk} can be represented via the parametric distribution $f(x_{jk})$. Then, the expected benefit of inspecting the cell with attribute k for product j is:

$$\gamma_{ijk}^{n+} = \int \ln\left(t' \exp\left[\beta' \tilde{X}_i^{n+}\right]\right) f(x_{jk}) dx_{jk} - \ln\left(t' \exp\left[\beta' \tilde{X}_i^n\right]\right). \quad (4)$$

The quantity γ_{ijk}^{n+} can be calculated for every product-attribute level which has not already been inspected by the consumer and this quantity is used to decide if any additional information processing will be undertaken, and if so, which product-attribute level will be inspected next. Numerical analysis shows that the value in (4) is positive when the ex-ante value of x_{jk} is the expected value or average \bar{x}_{jk} ; under other assumptions one could calculate the absolute value. We assume the cost of processing information in the point-of-purchase IDB is constant for each inspected cell and is not included in equation (4), but influences the decision to continue processing, as discussed below. The quantity in equation (4) plays the same role as the EVSI of Hagerty and Aaker (1984), but can accommodate any parametric distribution, not just the multivariate normal, thus accommodating both continuous and discrete attributes.

We finalize the model specification by formally stating the decisions faced by the consumer. Because, all consumers in our study inspected at least one product attribute level, the first decision made is which product attribute level to inspect. Mathematically, the consumer inspects x_{jk} if γ_{ijk}^{n+} is $\max\{\gamma_{ijk}^{n+}\}$ for all j and k values. Like Hagerty and Aaker (1984), we assume that consumers will continue to inspect product-attributes as long as the expected gain exceeds some threshold, τ , which represents the cognitive cost of acquiring and processing additional information. In other words, consumers will continue to inspect product-attributes as long as:

$$\text{Max}\{\gamma_{ijk}^{n+}\} > \tau, \quad (5)$$

where the set $\{\gamma_{ijk}^{n+}\}$ excludes cells that have already been inspected. We use the superscript “ $n+$ ” and the subscript i to emphasize that the value of γ_{ijk}^{n+} depends on which other product-attribute levels have already been inspected, i.e., the n previous selections made by individual i . Finally, letting \tilde{X}_i^f represent the final attribute matrix containing all the inspected attribute levels and a

column representing the “none” or outside good, then the chosen alternative is the one which corresponds to $Max(\beta' \tilde{X}_i^f + \varepsilon_i)$.

Equation (4) calculates the benefit of processing additional information as the expected value of revealing a particular product-attribute compared to the current state of knowledge. This formulation is similar to the “directed cognition model” of Gabaix et al. (2006). Unseen product-attributes, i.e., the product-attribute levels not revealed by the consumer in the IDB, are represented by the ex-ante expected values \bar{x}_{jk} .

Model Estimation. The data contains three pieces of information corresponding to the model elements described above: (i) which product-attribute level to inspect next, (ii) whether or not to continue inspecting product-attribute levels, and (iii) which product to choose. We assume that idiosyncratic shocks attend to not only the final product choice, but also to information processing decisions. This provides a straightforward way to model the multinomial or binary outcomes of the information processing model. The actual observations of these events allows for the specification and estimation of the statistical models.

Let x_{ijk}^{n+} be a candidate product-attribute level (i.e., the cell in the IDB). It is assumed the consumer makes a stochastic choice based on $\gamma_{ijk}^{n+} + \eta_{ijkn}$ where η is an i.i.d. extreme value error term with scale 1 and the subscript n indicates the n^{th} product-attribute level accessed. The consumer chooses which attribute level to inspect next based on $max\{\gamma_{ijk}^{n+} + \eta_{ijkn}\}$ for all j and k ; and the probability of this multinomial outcome is given by:

$$\Pr(x_{ijk}^{n+} = 1) = \frac{\exp[\gamma_{ijk}^{n+}]}{\sum_{\substack{j=1 \\ j \notin M}}^J \sum_{\substack{k=1 \\ k \notin M}}^K \exp[\gamma_{ijk}^{n+}]}, \quad (6)$$

where M is the set of product-attribute levels already inspected.

Let $c_i^n = 1$ if individual i continues to inspect product-attribute information at step n .

Then $c_i^n = 1$ if the largest value of γ_{ijk}^{n+} is greater than some threshold. Specifically, if

$\max_{j,k \in M} [\gamma_{ijk}^{n+}] + v_{in} > \tau$ then information processing continues. Assume the idiosyncratic shock v_{in} is

distributed i.i.d. extreme value with scale 1 then:

$$\Pr(c_i^n = 1) = \frac{\exp\left(\max_{j,k \in M} [\gamma_{ijk}^{n+}] - \tau\right)}{1 + \exp\left(\max_{j,k \in M} [\gamma_{ijk}^{n+}] - \tau\right)}. \quad (7)$$

Given the previous assumptions about ε , the final choice probability is given by:

$$\Pr(y_{ij} = 1) = \frac{\exp[\beta' \tilde{x}_{ij}^f]}{\sum_{j=1}^{J+1} \exp[\beta' \tilde{x}_{ij}^f]}, \quad (8)$$

where the $J+1^{\text{th}}$ product corresponds to the “no-purchase” option. The deterministic component of utility for the outside good is set to zero.

For each individual, N_i attribute levels are inspected. Let $[x_i^n]$ correspond to the cell inspection probability given in (6) for the n^{th} cell opened, $[c_i^n]$ correspond to the probability of continuing to inspect cells given in (7), $[1 - c_i^n]$ the probability to quit, and $[y_i^j]$ the final choice probability, given in (8). The likelihood function for an individual is given by:

$$\ell_i = \left[\prod_{n=1}^{N_i-1} [x_i^n] [c_i^n] \right] [x_i^{N_i}] [1 - c_i^{N_i}] [y_i^j]. \quad (9)$$

Assume a shopper inspected 11 product-attribute levels. Then the terms $\prod_{n=1}^{N_i-1} [x_i^n] [c_i^n]$ refer to the

first 10 cells opened and the decision to keep opening cells, and the terms $[x_i^{N_i}] [1 - c_i^{N_i}]$ refer to

the 11th cell opened and the decision to stop opening cells. Bayesian Markov Chain Monte Carlo

(MCMC) methods are used to obtain draws from the posteriors of all model parameters. In addition to the vector of importance weights β and the scalar stopping threshold τ , the parameters for $f(x_{jk})$ must be estimated or specified. Because all the products in the current study represent the same brand in a particular price category, we assume $f(x_{jk}) = f(x_k)$, i.e., the *a priori* distribution of the likely levels of an attribute are the same across products. However, this assumption may need to be revised in situations in which the products represent different brands or are from different price tiers.

Our data contain both continuous and discrete attributes. For continuous attributes, we assume x_k is distributed log-normal (μ_k, σ_k^2) . Following the example of De los Santos, Hortaçsu, and Wildenbeest (2012), we use the actual distribution of product-attribute levels in the choice set and calculate $E[x_k] = \bar{x}_k$, the arithmetic mean. The value of σ_k^2 is estimated from the data. For discrete attributes, $f(x_k)$ is assumed to be Bernoulli with parameter $\theta_k = 0.5$. The MCMC chain then produces a posterior distribution of the discrete values of x_k . Orthogonal coding is used for discrete attributes; i.e., for a discrete attribute with two levels, its presence is coded by 0.5 and its absence is coded by -0.5 . The product name and price are revealed to all consumers at the start of the data collection process and dummy variable coding is used to represent these model intercepts. The integration in equation (4) is implemented numerically.

Details of the estimation procedure appear in the Online Technical Appendix. Priors are diffuse but proper, and a standard Metropolis-Hastings algorithm is used to draw parameter values from their posteriors. The algorithm follows procedures detailed in Rossi, Allenby, and McCulloch (2005). Estimation of the remaining models are variations on the components outlined here and will therefore not be repeated. The most computationally challenging part of

the estimation procedure is calculating the set $\{\gamma_{ijk}^{n+}\}$ for each individual, for each attribute-level revealed. In a product-attribute matrix that has 3 products and 11 attributes (discussed in previous section), at time 0 the set $\{\gamma_{ijk}^{n+}\}$ contains 33 items, and after the first attribute level is revealed it contains 32 items, and so on. Having a computationally tractable expression for the value of revealing a particular attribute-level (e.g., equation (4)) is essential to estimating this class of models with real data involving multiple products and attributes and hundreds of consumers. Yang, Toubia, and De Jong (2015) have a similar expression in their model but it is made tractable by assuming all attributes are discrete, Hagerty and Aaker (1984) simplify their analysis by assuming all attributes are continuous; with data on real products we must consider both discrete and continuous attributes.

The intuition for model identification lies in the fact that although consumers were presented the same IDB, each consumer had a unique sequence in opening product-attributes, opened a different number of product attributes, and used a unique set of product-attributes to make the final product selection. Consumers revealed different subsets of the product-attribute levels creating variation in the \tilde{X}_i^f matrix; because of this variation across consumers, the attribute importance weights β can be identified using just the final product choice portion of the model, as in Currim, Mintz, and Siddarth (2015). Recall that after the first product-attribute level is revealed, each consumer has a unique set of $\{\gamma_{ijk}^{n+}\}$. For a given cell position specified by j and k , the value of γ_{ijk}^{n+} will change as additional product-attribute levels are revealed, so that $\gamma_{ijk}^{n+} \neq \gamma_{ijk}^{n+(+1)}$. Because of this, the $\{\gamma_{ijk}^{n+}\}$ are not confounded with the row or column order of the product-attributes in the IDB. Fixing \bar{x}_k and changing σ_k^2 implies a different sequence of

revealed product-attributes for each respondent. The parameters β and $\{\sigma_k^2\}$ cannot be simultaneously changed in equation (6), keeping the likelihood constant because only β contributes to the likelihood in equation (8). Changing the stopping parameter τ implies a different number of inspected attributes and again since β and $\{\sigma_k^2\}$ cannot be simultaneously changed without changing the other components of the likelihood, τ is uniquely identified by equation (7). Simulation experiments available from the authors demonstrate that the parameter values are identified and can be recovered using data sets analogous to those used in the empirical example.

Product-Attribute Selection

In the base model, the benefit of continuing to process information or to inspect a particular product-attribute level is the difference between the expected maximum utility of inspecting candidate cell x_{ijk}^{n+} and the expected maximum utility of making the final choice using the current set of revealed product-attributes. This is captured in equation (4) which requires specifying the distribution of x_k and integrating over the possible values of x_{jk} ; we refer to this as the “Expected Value” approach to product-attribute selection. In addition to statistical and economic theory, Meyer (1982) shows in an experimental setting that the probability that an alternative will be inspected is a function of both the expected utility and the uncertainty about the alternative. However, Miller (1993) suggests that the assumption that consumers necessarily manifest that information in the form of a probability distribution function is “less tenable.”

We look to the behavioral literature for an alternative method of computing the benefit of processing additional information that retains the level and dispersion of expected utility, but does not rely on parametric assumptions and is computationally less demanding. Jaccard,

Brinberg, and Ackerman (1986) propose a “Subjective Probability Measure” for attribute importance. Using their example, a subject is asked to rate on an 11 point scale how willing they would be to consider a car that was “inexpensive.” They are then asked how willing they would be to consider a car that was “expensive.” An index of importance is formed by taking the difference between the willingness to consider the car when the attribute “price” is at its highest level versus its lowest level. See Jaccard, Knox, and Brinberg (1979) for the logic and relation of this measure to different behavioral theories. This difference between the focal measure (consideration) when the attribute is at its highest versus lowest level motivates the following change to the model.

Let Min_k represent the lowest value that attribute k is expected to take and Max_k represent the corresponding highest value. Using the Max_k and Min_k , equation (4) is recast as:

$$\gamma_{ijk}^{n+} = \left[\ln \left(t' \exp \left[\beta' \tilde{X}_i^{n+(Max_k)} \right] \right) - \ln \left(t' \exp \left[\beta' \tilde{X}_i^{n+(Min_k)} \right] \right) \right]. \quad (10)$$

Thus, instead of integrating over the unknown values of x_{jk} , the benefit of inspecting a cell is given by the difference in expected maximum utility when the attribute level is at its *a priori* expected highest value and when it is at its lowest. Instead of “consideration” from the Jaccard, Brinberg, and Ackerman (1986) example, our focal measure is “expected maximum utility.” The level of expected utility is reflected in the values of Max_k and Min_k while the range, $Max_k - Min_k$, represents the uncertainty in the attribute levels. When a consumer is more uncertain about an attribute, there will be a relatively larger difference between Max_k and Min_k , resulting in a higher value for γ_{ijk}^{n+} which increases the probability that cell j,k will be inspected. In other words, when a consumer is uncertain about what she is going to get, there is a lot of benefit for her to acquire additional information. In contrast, when a consumer is relatively

certain about an attribute, Max_k and Min_k will be relatively close, γ_{ijk}^{n+} will be relatively small, and the probability that cell j,k is inspected will decrease. We refer to this as the “Max – Min” method of product-attribute selection.

For discrete attributes, Max_{jk} is set equal to the attribute being “present” and Min_{jk} is set equal to the attribute being “not present” in product j . When higher values of an attribute are expected to decrease indirect utility, i.e., $\beta_k < 0$, then the first two terms in equation (10) are reversed. When the analyst does not know whether higher or lower values will be preferred, γ_{ijk}^{n+} can be based on the absolute value of the difference. Equation (10) has a closed form which facilitates model estimation.

In terms of the model likelihood function, equation (10) simply replaces equation (4) in the calculation of γ_{ijk}^{n+} and the remainder of the model is unchanged. Instead of specifying a particular $f(x_k)$ such as the log-normal and estimating the coefficient of variation for continuous variables, we estimate Max_k and Min_k . The Online Technical Appendix contains full details. To facilitate estimation, we take Max_k and Min_k to be symmetric around \bar{x}_k but find that within a broad range, the results are not sensitive to different assumed \bar{x}_k . We restrict $Min_k \geq 0$.

Theory, past research, and empirical patterns in the data suggest that a consumer may select a particular product-attribute level based on its proximity to the previously opened cell, i.e., whether it is in the same row (attribute based processing in our data), column (alternative based processing), or diagonal (mixed processing) to the previous selection. Gabaix et al.'s (2006) experimental data was analyzed by Sanjurjo (2014) and, consistent with the behavioral literature, he found strong tendencies for row, column, and “typewriter” processing in an IDB; he referred to these patterns as “spatial biases.” One way to capture this information acquisition

process is via a distance metric that permits product-attributes closer to the last revealed cell to be preferred. Let $Dist_{ijk}^r$ and $Dist_{ijk}^c$ represent the row and column distance from the last item revealed; this is a “city block” type distance measure.⁴ The next product-attribute accessed is that j,k combination that satisfies:

$$Max(\gamma_{ijk}^{n+} - \phi_r Dist_{ijk}^r - \phi_c Dist_{ijk}^c). \quad (11)$$

In equation (11), holding γ_{ijk}^{n+} constant, cells which are closer to the currently opened cell are preferred to those that are further away. The relative magnitudes of ϕ_r and ϕ_c determine whether row (attribute) or column (alternative) proximity is more important. In our empirical analysis, we will combine the Max - Min calculation of equation (10) with the distance metric in equation (11) and refer to this as the “Hybrid” method of determining product-attribute selection.

Unseen Product-Attributes

In the base model, if a particular product-attribute was not inspected by a consumer, it is assumed that the mean \bar{x}_k for that product-attribute is used by the consumer. An analogous assumption is used in Yang, Toubia, and De Jong’s (2015) conjoint analysis. Meyer (1982) provides some experimental support for this by showing that consumers treat completely unknown alternatives as if they had the average utility of products in the market. We refer to this assumption as using the “Actual Mean” to represent unseen product-attributes.

We calculate \bar{x}_k as the arithmetic mean of the actual values x_{jk} for continuous attributes in the choice set. While this has precedence in the information processing literature and it seems to be a reasonable assumption, there is no guarantee in any setting that the actual mean matches consumers’ expectations. We therefore explore two different options for representing unseen product-attributes. The first is that consumers simply do not use product-attributes that they do

not inspect in the IDB. This is consistent with the finding that consumers limit cognitive effort and focus on only a subset of information to make decisions (see Bettman, Luce, and Payne 1998). Using our notation from earlier, before any product-cells are inspected, the vector $\tilde{x}_{ij} = 0$. When a product attribute level is revealed, x_{jk} replaces the appropriate 0 in the vector \tilde{x}_{ij} . Unseen attributes are simply not included in the calculation of the indirect utility. We will refer to unseen product-attributes being “Not Used” when the appropriate elements of $\tilde{x}_{ij} = 0$ when product-attributes are not inspected.⁵

The “Actual Mean” method of dealing with unseen product-attributes relies quite heavily on the specified value of \bar{x}_k since it appears in the indirect utility function. The “Not Used” method requires the assumption that consumers use attribute levels for deciding the benefit of inspecting product-attributes, but then ignore the attributes in the final product choice; we would like to avoid these assumptions. Therefore, we build on Branco, Sun, and Villas-Boas (2012) and propose a, third, expectation deviation method of handling unseen product-attributes. Thus far, the intercept term in the model simply represents the product label or information listed in each column of the IDB that is seen by every shopper at the beginning of the task; however, we now want to redefine the meaning of the intercept to include the “expected value” of the attributes. Consider a product class with only two attributes, with indirect utility specified as:

$$V_{ij} = \beta_0^{\bar{x}} + \beta_1(x_{j1} - \bar{x}_1) + \beta_2(x_{j2} - \bar{x}_2) + \varepsilon_{ij}. \quad (12)$$

Here $\beta_0^{\bar{x}}$ is the indirect utility when the product-attribute levels all equal their expected values. This suggests a new parameterization of the indirect utility function. Let $d_{ijk} = 1$ when product-attribute j,k has not been inspected by consumer i and $d_{ijk} = 0$ if it has. Then:

$$V_{ij} = \beta_0^* + \beta_1 x_{j1}(1 - d_{ij1}) + \beta_2 x_{j2}(1 - d_{ij2}) + \delta_1 d_{ij1} + \delta_2 d_{ij2} + \varepsilon_{ij}. \quad (13)$$

In this model δ_k theoretically equals $\beta_k \bar{x}_k$, but there are no restrictions on δ_k or β_k and both are estimated simultaneously from the data. Here we can again assume that if a product-attribute is not inspected that $x_{jk} = \bar{x}_k$ but we do not need to specify \bar{x}_k , it is estimated from the data via the parameter δ_k . To the extent that δ_k / β_k yields plausible values of \bar{x}_k , it supports the veracity of the model. We will refer to this method of handling unseen product-attributes as “Inferred Mean.” Additional details on the derivation are contained in the technical appendix. Table 1 summarizes the main modeling elements and the different variations that will be tested using our empirical data.

RESULTS

We have data from three different market segments: high priced, medium priced, and low priced. In order to control for parameter heterogeneity, separate models were estimated for each market segment. As in the Data section, we will focus on the high priced market segment and then show that similar results were obtained in the other two segments. A total of 110 consumer responses were used to calibrate the models listed in the top of Table 4; 26 responses were reserved for hold-out testing. For all models, the MCMC chains were run for 50,000 iterations with a thinned sample of every 10th from the last 25,000 used to estimate the posterior moments of the parameters. The β 's associated with Attributes A1 – A9 and A11 were expected to be positive and equation (10) was used to calculate γ_{ijk}^{n+} for the Max-Min and Hybrid models. Attributes 10a and 10b describe physical aspects of the product and it is unclear if larger or smaller values would be desired for this class of product; therefore, the absolute value of equation (10) was used for these two attributes. For the Expected Value models, in order to ensure comparability to the Max-Min and Hybrid models, the β 's associated with Attributes A1 – A9 and A11 were

restricted to be >0 . Table 4 describes the models and contains each model's fit statistics, while Table 5 provides the posterior means and posterior standard deviations for parameters from selected models.

Model Comparisons

A broad set of model combinations from Table 1 were estimated with choices guided by the logical consistency of the model alternatives and preliminary results. In-sample fit is measured by the log marginal density (LMD), calculated using the Gelfand and Dey (1994) importance sampler and includes a penalty for the number of parameters. The LMD favors the model with the largest value, which is model 8, the Hybrid product-attribute selection process and the Inferred Mean method of handling unseen attributes. The out-of-sample fit is measured by computing the log likelihood of the observed data for the 26 consumers in the hold-out sample using a random sample of 2,500 draws of the parameters from the posterior distribution, and then calculating the average. This measure favors the model with the largest value, which again is model 8.

Focusing on the first four rows of Table 4, we can contrast the performance of the Expected Value versus the Max-Min method of calculating the benefits of additional information processing. In each comparable model, the Max-Min methodology has a better in-sample and out-of-sample fit than the Expected Value methodology. The Max-Min model fits the data better and offers significant computational savings. The Hybrid model modifies Max-Min (see equation (11)) by penalizing cells in the IDB which are farther away from the last inspected cell. Results in the upper panel of Table 4 show that for otherwise comparable models, the Hybrid model has both better in-sample and out-of-sample fit than the Max-Min model (Models 3 vs. 6, 4 vs. 7,

and 5 vs. 8). These results suggest that the proximity of cells has an important role in determining the information processing of consumers in point-of-purchase IDBs.

In the three comparable models on how to handle unseen product-attributes, models with the Not Used assumption fit the in-sample and out-of-sample data better than models with the Actual Mean assumption. However, the Inferred Mean method, which includes parameters that account for the product category attribute means, fit the data the best. These results offer support for the substantive conclusion that consumers rely on prior information and provide a mechanism for modeling this when data on consumers' prior beliefs is not available.

In summary, when deciding which product-attribute levels to inspect, both the change in expected maximum utility and the location of the cell in relation to the last cell inspected are important. Also, when evaluating the overall products, consumers use a prior value for product-attributes which are not inspected in the IDB. Methodologically, the results show using a Max-Min type of calculation for the anticipated benefit of information processing is effective compared to the computationally more demanding Expected Value calculation. Further, the proposed Inferred Mean method of representing consumers' prior values for product-attributes is viable in situations with limited consumer data. This analysis was replicated for the medium priced (177 consumers; 149 estimation and 28 validation sample) and low priced (582 consumers; 522 estimation and 60 validation sample) market segments for the best fitting models (models 7 and 8), and the results match the pattern seen in the high priced market segment. Model fit statistics are reported at the bottom of Table 4.

Parameter Estimates

Table 5 contains parameter estimates for selected models for the high priced market segment. In addition to the models already described, a model which used only the final product choices to

calibrate the β 's was also estimated. In this model, the decision of which product-attribute to inspect next and the when to stop inspecting attributes is random⁶ and only the opened product-attribute levels were used to calibrate the final choice model. Table 4 shows that the in-sample and out-of-sample fit for the model which did not consider consumers' information processing (Model 9) was worse than all other models considered. Table 5 shows that all the posterior estimates of β for the "random" model included 0 in their 95% highest posterior densities (i.e., were not statistically significant) except for the product intercepts. By contrast, virtually all the estimates of the β 's in models 7 and 8 which model consumers' information processing are statistically significant and have the expected sign (i.e., greater values of each attributes should be preferred except attributes A10a and A10b, which were height and weight dimensions, and it is unknown whether consumers preferred larger and heavier products). Only the β for A10a in Model 8 is not statistically significant. Nine of the eleven estimates of δ from Model 8 are statistically significant; because attributes A10a and A10b are opened at the same time, only a single value of δ could be estimated. Managerial applications of these parameter estimates will be discussed in the next section.

In models 7 and 8, both the row ϕ_r and column ϕ_c coefficients of the Hybrid model are significant at the 0.05 level. The value for the row coefficient is greater than the value for the column coefficient (0.85 vs. 0.56 in model 7, 0.85 vs. 0.40 in model 8), which indicates that consumers were more likely to process by alternative (column) than by attribute (row). This contrasts with the empirical findings of Simonson, Huber, and Payne (1988) and Yang, Toubia, and De Jong (2015) who found a greater propensity for attribute level processing in their laboratory based choice tasks; but supports the empirical findings in the two-stage literature (e.g., Bettman and Park 1980, Gensch 1987, Hauser and Wernerfelt 1990, Payne 1976), which suggest

that alternative-based processing is expected over attribute-based processing just before a choice. As noted earlier, Biehal and Chakravarti (1986) suggest that consumers may switch from brand or alternative (column) to attribute (row) based processing in the midst of accessing information from the IDB; this would imply dynamic values of ϕ_r and ϕ_c , a topic we leave for further research.⁷

The posterior means of the Max_k and Min_k are listed for models 7 and 8. In estimating the Max-Min model, the Min_k value was constrained to be greater than or equal to 0 because it would not make sense to have negative values for these attributes. The results show that for A1, A10a, and A10b that Min_k was not significantly different from 0. Importantly, the upper limit Max_k is not constrained and none of these parameter estimates are unreasonable given the actual value of the attributes.

In summary, these results show that the information processing models (models 7 and 8) produce better parameter estimates than models which just use the final choice data (model 9) to estimate market level preferences. These results also show that the models can be estimated using just the revealed sequence of product-attribute levels and final choices from an actual market setting.

Managerial Application of Models

Here we briefly explore two managerially relevant applications of model 8 which incorporate the Hybrid product-attribute selection and the Inferred Mean approach to handling unseen product-attributes. One of the interesting aspects of the Inferred Mean method is that theoretically,

$\delta_k = \beta_k \bar{x}_k$ where \bar{x}_k is some prior value for attribute k that consumers use when product-attribute k is not opened for a particular product. This implies that $\bar{x}_k = \delta_k / \beta_k$ and we can estimate the market's expected value for attribute k . The last column of Table 5 performs this calculation

using the posterior means of β_k and δ_k from the data. Since orthogonal coding was used for the discrete attributes, $\delta_k/\beta_k = 0.5$ indicates that the market expected that attribute to be part of the product offering while $\delta_k/\beta_k = -0.5$ indicates the opposite. However, since no constraints were placed on the estimates of δ it should not be surprising that δ_k/β_k does not exactly equal 0.5 or -0.5 ; we will interpret the sign of δ_k/β_k as indicating the market's expectations. Attributes A2 and A7 – A9 were all expected to be part of the product offering in the high priced market. Attribute A3 was not expected to be included. Attribute A11 was a special one-time promotion that would have been unknown to consumers, this may explain why δ_{11} was not statistically significant, but directionally, the market did not expect this attribute. For the three continuous attributes that we can measure unambiguously, the implied values of \bar{x}_k (0.73, 65.70, and 13.50) are plausible given the product and are within the range of Min_k and Max_k estimated for model 8. These results show that the model can provide estimates of the market expectations for the different attributes. Together with the values of $\{Max_k, Min_k\}$, these results can highlight where market expectations are different from actual product offerings or consumers are especially uncertain of the values of product attributes, leading to specific advertising messages or point-of-purchase education on e-tailers' websites. For instance, since our results also indicated that the proximity of cells matter, e-tailers may want to “cluster” attributes with high uncertainty closer together in a product information matrix in order to simplify the consumer's task and improve the probability that the information is processed.

Second, our method addresses a common problem in online marketing, poor conversion rates. Our model provides recommendations for prioritizing who should receive follow-up communications via either e-mail or retargeting display ads and which product should be

featured for each individual, based on actual information processing on the firm's IDB enabled website. In our hold-out sample of 26 shoppers, 16 did not purchase one of the three products. Table 6 provides the purchase probabilities from model 8 for each of the three products for each of these shoppers, with results sorted in terms of those most likely to buy one of the three products. This provides managers diagnostics for which shoppers had greater probabilities of purchasing and for which products they are most interested. Despite the common model parameters, consumers have different purchase probabilities as a result of their unique information acquisition patterns. Managers can compute a cut-off via an expected value calculation to determine who should receive follow-up information; for instance, in our example it may not be profitable for managers to follow-up with shoppers 9 – 16 because their forecasted purchase probability is relatively low.⁸ It is important to note that because our method computes attribute importance weights, it gives different recommendations than one which simply counts the number of product-attributes revealed and targets shoppers based on that metric. Specifically, both shoppers 2 and 4 revealed 11 product-attribute levels, but they have very different purchase probabilities. Among the 10 consumers in the hold-out sample who did choose one of the three products, using the maximum predicted probability correctly identified the chosen product in 9 out of the 10 cases.

CONCLUSION

This research models information processing of an IDB at the point-of-purchase from consumers in actual purchasing situations while they were shopping on a popular manufacturer's retail website. Unlike previous studies using IDB data, by building an integrated model of the consumer information processing decisions, we can infer attribute importance weights, β_k , the ex-ante expected values of attributes, δ_k/β_k , and a measure of uncertainty in the values of the

attributes Max_k and Min_k , all in the face of realistically small variation in attribute levels. We find that our behaviorally inspired ($Max_k - Min_k$) method of measuring the benefit of information acquisition does better in our data set than an analogous expected value calculation. Using the actual average value of attribute levels to represent unseen product attributes does worse than assuming those attributes are not used in product choice; however, inferring the ex-ante value dominates both these approaches. Similar to past research, we find that the row and column proximity of attributes in the IDB is important in determining which product-attribute will be inspected next.

This research contributes to management practice since the proposed method has minimal information requirements and can be implemented by e-tailers in actual purchase situations, as demonstrated in this research. Online shoppers can be randomly selected and diverted to websites where product-attribute information is revealed on-demand. Estimating attribute importance weights that account for the inherent endogeneity in information processing is an important managerial contribution of this research. This method will be of most interest to vertically integrated e-tailers or e-tailers looking to collaborate with their suppliers on new product design, advertising, promotion, etc. This method also provides diagnostics to help managers prioritize follow-up communications towards consumers who were most likely to purchase based on their current visit to a retailer's IDB enabled website (see Table 6). Because consumers have different sequences of information processing as well as different final sets of revealed product-attributes, the model can overcome the difficulties of not having an experimentally designed product matrix. Across consumers, not everyone reveals all the information in the IDB, and the differences create variation that allows for parameter estimates, even for attributes with identical levels across attributes.

Compared to previous eye-tracking and conjoint academic studies, the consumers in this research were making actual purchase decisions and the data collection technique is uniquely suited to e-tailers. This research suggests several other avenues of additional research. First, we do not model the “upstream” information search and processing (ISP) that occurs prior to arriving at the e-tailer’s IDB. We argue that at a given moment in time, the cumulative effect of consumers’ “upstream” activities can be represented by their current preferences (attribute importance weights), their current expected values of attributes, and their expected range of those attributes and we use the data to recover these parameters. But how do consumers’ form their expectations and how much time do they spend searching for information across different media? This extended search process will be a function of shopper specific characteristics (opportunity cost of time, prior purchase experience, risk tolerance, etc.) and the mix of “search” and “experience” attributes in the product. The decision context must be expanded to include the decision to buy a product, not buy a product now but continue to search, or exit the market altogether, while accounting for different types of interactions across a variety of platforms.⁹ Bronnenberg, Kim, and Mela (2016) provide some initial findings about the online search process using aggregated product purchase data. Whereas our model is appropriate for inferring market level values of $\{\beta\}$ and $\{Max_k, Min_k\}$ at a point in time, it does not capture the full ISP process. Such models will have to use tracking data from multiple sources, online and off-line, but this is increasingly possible, subject to the limitation of consumers’ privacy concerns.

Another important area for additional research is modeling heterogeneity. In infrequently purchased categories it is rare to observe multiple purchases from the same consumer in point-of-purchase data. This makes it difficult to accurately estimate distributions of heterogeneity. One option would be to make parameters such as β_0^* or δ functions of demographic information.

Google uses browser history and IP addresses to infer a range of demographics from web browsers; this information could be incorporated without altering the current data collection methodology. In addition, finite mixture models or Dirichlet process priors could be used with the current data collection procedure. Incorporating heterogeneity into the Expected Value model may not be computationally feasible but should be practical in the Max-Min and Hybrid models: we leave the testing of these models to future research.

Two other extensions are worth noting. First, as posited by Biehal and Chakravarti (1986), consumers may change their decision processing procedure within a decision task and/or different situational factors may give rise to different choice procedures. As noted earlier, it may be worthwhile to investigate a dynamic parameterization of ϕ_r and ϕ_c or different measures of brand or alternative (column) and attribute (row) processing. Second, although we have argued that the current data collection method is well suited to e-tailers and that experimental evidence suggests it does not change shopping and decision patterns, the proposed models could be used and validated with services such as YouEye that offer online eye-tracking via panelists webcams.

Marketing research has always been interdisciplinary and this study extends that tradition by developing and applying new models with roots in economics and psychology. Importantly, this research shows that theoretical and lab based results can be inferred and tested in an actual purchase situation. The new models can be used by management in applied situations as demonstrated and hopefully will be used as a basis for additional academic and commercial research.

Footnotes

¹ In the actual implementation, additional information such as lens type shown in Figure 2 was not displayed in the first row of the IDB.

² The retailer decided to implement the “Customize and Buy” option rather than go with the Decision Board’s default “Choose” button.

³ This includes the product-attribute levels already revealed and the ex-ante values for product-attributes which have not been opened.

⁴ We also tested a “Euclidean” distance metric but the city block metric produced better in-sample and out-of-sample fit.

⁵ If a consumer is using a non-compensatory choice process, such as eliminations by aspects, then there may be no reason to open a particular cell. We thank an anonymous reviewer for this alternative explanation of why consumers might not inspect a particular product-attribute.

⁶ The probability of inspecting a product-attribute was equal to $1/(N-n_i)$ where N is the total number of product-attributes ($N=33$) and n_i is the number of cells opened so far by consumer i ; the probability is one divided by the number of product-attributes currently unopened in the IDB. The probability of stopping was $\frac{1}{2}$ after each cell was inspected.

⁷ It is interesting to note that in Figure 2, the consumer processes the first 15 cells by row (attributes) and then switches to processing by column (brand) for the last three cells, selecting the product associated with those last three cells. As noted by an anonymous reviewer, this is consistent with the process suggested by Biehal and Chakravarti (1986).

⁸ Shoppers 13 – 16 each opened the same 3 cells in the first row of the information display board.

⁹ We thank an anonymous reviewer for drawing this distinction between the extended ISP process and our modeling efforts.

References

- Anderson, Simon P., André De Palma, and Jacques François Thisse (1992), *Discrete Choice Theory of Product Differentiation*, Cambridge, Massachusetts: MIT Press.
- Aschemann-Witzel, Jessica and Ulrich Hamm (2011), “Measuring Consumers’ Information Acquisition and Decision Behavior with the Computer-based Information-Display-Matrix,” *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 7 (1), 1–10.
- Bettman, James R., Mary Frances Luce, and John W. Payne (1998), “Constructive Consumer Choice Processes,” *Journal of Consumer Research*, 25 (3), 187–217.
- and C. Whan Park (1980), “Effects of Prior Knowledge and Experience and Phase of the Choice Process on Consumer Decision Processes: A Protocol Analysis,” *Journal of Consumer Research*, 7 (3), 234–48.
- Biehal, Gabriel and Dipankar Chakravarti (1986), “Consumers’ Use of Memory and External Information in Choice: Macro and Micro Perspectives,” *Journal of Consumer Research*, 12 (4), 382–405.
- Branco, Fernando, Monic Sun, and J. Miguel Villas-Boas (2012), “Optimal Search for Product Information,” *Management Science*, 58 (11), 2037–56.
- Bronnenberg, Bart J., Jun B. Kim, and Carl F. Mela (2016), “Zooming In on Choice: How Do Consumers Search for Cameras Online?” *Marketing Science*, Forthcoming.
- Currim, Imran S., Ofer Mintz, and S. Siddarth (2015), “Information Accessed or Information Available? The Impact on Consumer Preferences Inferred at a Durable Product E-commerce Website,” *Journal of Interactive Marketing*, 29 (1), 11–25.
- De los Santos, Babur, Ali Hortaçsu, and Matthijs R. Wildenbeest (2012), “Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior,” *American Economic Review*, 102 (6), 2955–80.
- Dhar, Ravi and Stephen M. Nowlis (2004), “To Buy or Not to Buy: Response Mode Effects on Consumer Choice,” *Journal of Marketing Research*, 41 (4), 423–32.
- Feit, Eleanor McDonnell, Mark A. Beltramo, and Fred M. Feinberg (2010), “Reality Check: Combining Choice Experiments with Market Data to Estimate the Importance of Product Attributes,” *Management Science*, 56 (5), 785–800.
- Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg (2006), “Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model,” *American Economic Review*, 96 (4), 1043–68.

- Gelfand, A. E. and D. K. Dey (1994), "Bayesian Model Choice: Asymptotics and Exact Calculations," *Journal of the Royal Statistical Society. Series B (Methodological)*, 56 (3), 501–14.
- Gensch, Dennis H. (1987), "A Two-Stage Disaggregate Attribute Choice Model," *Marketing Science*, 6 (3), 223–39.
- Hagerty, Michael R. and David A. Aaker (1984), "A Normative Model of Consumer Information Processing," *Marketing Science*, 3 (3), 227.
- Hauser, John H. and Birger Wernerfelt (1990), "An Evaluation Cost Model of Consideration Sets," *Journal of Consumer Research*, 16 (4), 393–408.
- Jaccard, James J., David Brinberg, and Lee J. Ackerman (1986), "Assessing Attribute Importance: A Comparison of Six Methods," *Journal of Consumer Research*, 12 (4), 463–68.
- , Richard Knox, and David Brinberg (1979), "Prediction of Behavior from Beliefs: An Extension and Test of a Subjective Probability Model," *Journal of Personality and Social Psychology*, 37 (7), 1239–48.
- Johnson, Eric J., Robert J. Meyer, and Sanjoy Ghose (1989), "When Choice Models Fail: Compensatory Models in Negatively Correlated Environments," *Journal of Marketing Research*, 26 (3), 255–70.
- , John W. Payne, and James R. Bettman (1988), "Information Displays and Preference Reversals," *Organizational Behavior & Human Decision Processes*, 42 (1), 1.
- Lehmann, Donald R. and William L. Moore (1980), "Validity of Information Display Boards: An Assessment Using Longitudinal Data," *Journal of Marketing Research*, 17 (4), 450–59.
- Meißner, Martin, Andres Musalem, and Joel Huber (2016), "Eye Tracking Reveals Processes That Enable Conjoint Choices to Become Increasingly Efficient with Practice," *Journal of Marketing Research*, 53 (1), 1–17.
- Meyer, Robert J. (1982), "A Descriptive Model of Consumer Information Search Behavior," *Marketing Science*, 1 (1), 93–121.
- Miller, Harvey J. (1993), "Consumer Search and Retail Analysis," *Journal of Retailing*, 69 (2), 160–92.
- Mintz, Alex, Nehemia Geva, Steven B. Redd, and Amy Carnes (1997), "The Effect of Dynamic and Static Choice Sets on Political Decision Making: An Analysis Using the Decision Board Platform," *American Political Science Review*, 91 (3), 553–66.

- Mintz, Ofer, Imran S. Currim, and Ivan Jeliazkov (2013), "Information Processing Pattern and Propensity to Buy: An Investigation of Online Point-of-Purchase Behavior," *Marketing Science*, 32 (5), 716–32.
- Moorthy, Sridhar, Brian T. Ratchford, and Debabrata Talukdar (1997), "Consumer Information Search Revisited: Theory and Empirical Analysis," *Journal of Consumer Research*, 23 (4), 263–77.
- Payne, John W. (1976), "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis," *Organizational Behavior & Human Performance*, 16 (2), 366–87.
- , James R. Bettman, and Eric J. Johnson (1993), *The Adaptive Decision Maker*, Cambridge, UK: Cambridge University Press.
- Rossi, Peter E., Greg M. Allenby, and Rob McCulloch (2005), *Bayesian Statistics and Marketing*, Chichester, West Sussex, England: John Wiley & Sons.
- Sanjurjo, Adam (2014), "The Role of Memory in Search and Choice," Working Paper, Universidad de Alicante.
- Sheluga, David A., James Jaccard, and Jacob Jacoby (1979), "Preference, Search, and Choice: An Integrative Approach," *Journal of Consumer Research*, 6 (2), 166–76.
- Simonson, Itamar, Joel Huber, and John W. Payne (1988), "The Relationship Between Prior Brand Knowledge and Information Acquisition Order," *Journal of Consumer Research*, 14 (4), 566–78.
- Yang, Liu (Cathy), Olivier Toubia, and Martijn G. De Jong (2015), "A Bounded Rationality Model of Information Search and Choice in Preference Measurement," *Journal of Marketing Research*, 52 (2), 166–83.

Figure 1. Online Shopping Example at Best Buy

Español | My Account
Store Locator Weekly Deals

Q

PRODUCTS
SERVICES
SHOPS & DEALS
GIFTS

Please create an account

HOT JULY BLACK FRIDAY
Happening now. Plus free shipping. Shop now >

COMPARE

Back to results

PRODUCTS	X	X	X
	<p style="font-size: 10px; margin: 5px 0;"> Nikon - 1 J3 Compact System Camera with 10-100mm VR Lens Model: 27657 SKU: 7807042 Sale: \$849.99 Reg. Price: \$1,049.99 Add to Cart </p>	<p style="font-size: 10px; margin: 5px 0;"> Nikon - D5200 Digital SLR Camera with 18-55mm VR Lens Model: 1503 SKU: 7525218 Sale: \$799.99 Reg. Price: \$899.99 Add to Cart </p>	<p style="font-size: 10px; margin: 5px 0;"> Nikon - 1 V2 Compact System Camera with 10-30mm VR Lens Model: 27604 SKU: 7039094 Sale: \$799.99 Reg. Price: \$899.99 Add to Cart </p>
Megapixels (Effective)	14.2	24.1	14.2
Megapixels (Total)	15.13	24.71	15.13
Image Resolution	Up to 4608 x 3072	Up to 6000 x 4000	Up to 4608 x 3072
Optical Zoom	10x	3x	3x
Digital Zoom	None	None	None
Lens Features	1 NIKKOR VR telephoto zoom lens; 19 elements in 12 groups; 7 diaphragm blades; 3 aspherical elements; 2 ED glass elements; internal focusing; 55mm filter; 77° maximum angle of view; 10x zoom ratio	DX-format lens; 11 elements in 8 groups (1 aspherical); 7 diaphragm blades; super integrated coating; Silent Wave Motor	CX-format with 12 lens elements in 9 groups, including 3 aspherical elements; 7 aperture blades; Nikon Super Integrated Coating; silent-stepping AF motor; 40.5mm filter size
Lens Focal Length(s)	10-100mm (35mm equivalent 27-270mm)	18-55mm	10-30mm (35mm equivalent 27-81mm)
LCD Screen Size	3"	3"	3"
Product Height	Camera: 2.4"; lens: 2.4"	3.9"	Body: 3.2"; 10-30mm lens: 2.3"
Product Width	Camera: 4"; lens: 2.4"	5.1"	Body: 4.2"; 10-30mm lens: 2.3"
Product Weight	Camera: 7.1 oz.; lens: 10.5 oz.	1.2 lbs.	Body: 9.8 oz.; 10-30mm lens: 4.1 oz.
Product Depth	Camera: 1.1"; lens: 2.8"	3.1"	Body: 1.8"; 10-30mm lens: 1.7"
AVAILABILITY			
Shipping:	Shipping: Usually leaves our warehouse in 1 business day Estimate Arrival Time	Shipping: Usually leaves our warehouse in 1 business day Estimate Arrival Time	Shipping: Usually leaves our warehouse in 1 business day Estimate Arrival Time
Store Pickup	Store Pickup: Available at most stores Check Stores	Store Pickup: Available at most stores Check Stores	Store Pickup: Available at most stores Check Stores
PRODUCTS	X	X	X
	<p style="font-size: 10px; margin: 5px 0;"> Nikon - 1 J3 Compact System Camera with 10-100mm VR Lens Model: 27657 SKU: 7807042 Sale: \$849.99 Reg. Price: \$1,049.99 Add to Cart </p>	<p style="font-size: 10px; margin: 5px 0;"> Nikon - D5200 Digital SLR Camera with 18-55mm VR Lens Model: 1503 SKU: 7525218 Sale: \$799.99 Reg. Price: \$899.99 Add to Cart </p>	<p style="font-size: 10px; margin: 5px 0;"> Nikon - 1 V2 Compact System Camera with 10-30mm VR Lens Model: 27604 SKU: 7039094 Sale: \$799.99 Reg. Price: \$899.99 Add to Cart </p>

Figure 2. Decision Board Example

Sequence of Cells Accessed

Features	 Nikon - 1 J3 Compact System Camera with 10-100mm VR Lens Model: 27657 SKU: 7807042 Sale: \$849.99 Reg. Price: \$1,049.99	 Nikon - D5200 Digital SLR Camera with 18-55mm VR Lens Model: 1503 SKU: 7525218 Sale: \$799.99 Reg. Price: \$899.99	 Nikon - 1 V2 Compact System Camera with 10-30mm VR Lens Model: 27604 SKU: 7039094 Sale: \$799.99 Reg. Price: \$899.99
Megapixels (Effective)	14.2	24.1	14.2
Megapixels (Total)	▶ Select	▶ Select	▶ Select
Image Resolution	Up to 4808 x 3072	Up to 8000 x 4000	Up to 4808 x 3072
Optical Zoom	10x	3x	3x
Digital Zoom	▶ Select	▶ Select	▶ Select
Lens Features	▶ Select	▶ Select	▶ Select
Lens Focal Length(s)	10-100mm (35mm equivalent 27-270mm)	18-55mm	10-30mm (35mm equivalent 27-81mm)
LCD Screen Size	3"	3"	3"
Dimensions	▶ Select	Height: 3.9" Weight: 1.2 lbs.	▶ Select
Shipping	▶ Select	Usually leaves our warehouse in 1 business day	▶ Select
Store Pickup	▶ Select	Available at most stores	▶ Select
Your Choice:	▶ Choose	▶ Choose	▶ Choose

		
1	2	3
4	5	6
7	8	9
13	14	15
10	11	12
	16	
	17	
	18	
	X	

Table 1. Summary of Model and Alternatives

Product-Attribute Selection	
<i>Which cell in the IDB will be opened next?</i>	
Expected Value	The benefit to acquire product-attribute information is the change in the expected maximum utility. Need to integrate over the potential values of the candidate product-attribute.
Max - Min	The benefit to acquire product-attribute information is the difference between the expected maximum utility, when the candidate product-attribute is at its maximum versus at its minimum expected value.
Hybrid	The benefit to acquire product-attribute information includes the proximity of cells in the IDB: closer cells are preferred to cells which are farther away.
Unseen Product-Attributes	
<i>If a cell in the IDB is not opened, does the consumer use that product-attribute? If so, what value is used for that product-attribute?</i>	
Actual Mean	If a product-attribute is not seen by a consumer, she uses a prior value: \bar{x}_k calculated from the actual product attributes.
Not Used	If a product-attribute is not seen by a consumer, she simply does not use that product-attribute in the final product choice decision.
Inferred Mean	If a product-attribute is not seen by a consumer, she uses a prior value: $\delta_k = \beta_k \bar{x}_k$ are estimated as part of the model.

Table 2. Matrix of Information Presented to Shoppers

Attribute	Row	Continuous or Discrete Variable	Product 1 (Level)	Product 2 (Level)	Product 3 (Level)
Price	N/A	Continuous	1	2	1
A1	1	Continuous	1	2	1
A2	2	Discrete	1	1	1
A3	3	Discrete	2	1	2
A4	4	Continuous	1	1	1
A5	5	Continuous	2	3	1
A6	6	Discrete	1	1	1
A7	7	Discrete	1	1	1
A8	8	Discrete	1	1	1
A9	9	Discrete	1	1	1
A10a	10	Continuous	1	3	2
A10b		Continuous	2	3	1
A11	11	Discrete	1	1	2

Prices and alternative (model) names were available to shoppers without clicking a cell; Attribute 10a and 10b were accessed by clicking the same cell; Level 1 indicates lowest values and level 3 indicates highest values for the attributes, for example, level 1 for price indicates a lower price, level 1 for attribute 10a indicates a shorter alternative, and level 1 for attribute 10b indicates a lighter alternative.

Table 3. Descriptive Statistics – High Priced Market Segment

Panel A. Number of Cells Accessed by Shoppers based on Customize and Buy (C&B) Decision

Number of Cells Accessed	Total Shoppers		No C&B		C&B Product 1		C&B Product 2		C&B Product 3	
	Count	%	Count	%	Count	%	Count	%	Count	%
Average	11.54		9.99		16.08		13.33		12.28	
St. Dev.	9.24		8.36		11.49		9.97		9.19	
Median	9		7		14		11		11	
2-4 Cells	38	28%	26	33%	2	15%	6	22%	4	22%
5-9 Cells	33	24%	19	24%	2	15%	3	11%	3	17%
10-15 Cells	28	21%	15	19%	3	23%	10	37%	6	33%
16+ Cells	37	27%	18	23%	6	46%	8	30%	5	28%
Total	136	---	78	57%	13	10%	27	20%	18	13%

Panel B. Percentage of Shoppers Accessing Different Cells

Attribute	Product 1	Product 2	Product 3	Overall
A1	60%	73%	68%	89%
A2	35%	44%	38%	68%
A3	41%	46%	40%	70%
A4	32%	39%	36%	60%
A5	28%	35%	32%	54%
A6	22%	32%	27%	49%
A7	21%	32%	26%	47%
A8	27%	32%	32%	48%
A9	24%	34%	29%	48%
A10	21%	26%	23%	40%
A11	30%	35%	32%	48%
Overall	79%	85%	82%	7%

In Panel B, the “Overall” column should be interpreted as follows: 89% of shoppers inspected attribute A1 for one of the three products (upper right-hand corner); the “Overall” row indicates that 79% of shoppers inspected at least one attribute for Product 1 (lower left-hand corner). 7% of shoppers inspected all the product-attribute levels (lower right-hand corner).

Table 4. Model Fit Statistics**High Price Market Segment, In-sample N=110, Out-of-sample N=26**

<u>Model</u>	<u>Product-Attribute Selection</u>	<u>Unseen Product-Attributes</u>	<u>LMD</u>	<u>Out-of-sample</u>
1	Expected Value	Actual Mean	-4385.9	-956.5
2	Expected Value	Not Used	-4317.6	-949.4
3	Max-Min	Actual Mean	-4350.3	-920.0
4	Max-Min	Not Used	-4037.0	-801.3
5	Max-Min	Inferred Mean	-3924.3	-800.1
6	Hybrid	Actual Mean	-3225.1	-662.9
7	Hybrid	Not Used	-3126.7	-607.8
8	Hybrid	Inferred Mean	-3055.0	-598.7
9	Random	Not Used	-5025.9	-1060.1

Medium Price Market Segment, In-sample N= 149, Out-sample N=28

<u>Model</u>	<u>Product-Attribute Selection</u>	<u>Unseen Product-Attributes</u>	<u>LMD</u>	<u>Out-of-sample</u>
7	Hybrid	Not Used	-4388.6	-876.7
8	Hybrid	Inferred Mean	-4258.5	-857.1

Low Price Market Segment, In-sample N=522, Out-of-sample N=60

<u>Model</u>	<u>Product-Attribute Selection</u>	<u>Unseen Product-Attributes</u>	<u>LMD</u>	<u>Out-of-sample</u>
7	Hybrid	Not Used	-14192.9	-1690.6
8	Hybrid	Inferred Mean	-13380.0	-1622.7

Table 5. Selected Parameter Estimates

β 's for Product Attributes	Model 9 – Random, Not Used		Model 7 – Hybrid, Not Used		Model 8 – Hybrid, Inferred Mean				
	$\bar{\beta}$	(std. dev.)	$\bar{\beta}$	(std. dev.)	$\bar{\beta}$	(std. dev.)	$\bar{\delta}$	(std. dev.)	$\bar{\delta}/\bar{\beta}$
A1	0.241	(0.36)	2.112	(0.13)	1.475	(0.38)	1.078	(0.65)	0.73
A2	1.672	(1.26)	1.651	(0.04)	4.984	(1.07)	1.810	(0.61)	0.36
A3	0.266	(0.81)	1.561	(0.03)	1.317	(0.17)	-0.893	(0.23)	-0.67
A4	-0.015	(0.01)	0.018	(0.00)	0.112	(0.03)	7.334	(2.16)	65.70
A5	0.035	(0.05)	0.045	(0.00)	0.337	(0.13)	4.554	(1.92)	13.50
A6	-0.028	(1.26)	1.489	(0.03)	4.352	(1.22)	1.739	(0.75)	0.40
A7	2.098	(1.92)	1.585	(0.03)	4.476	(1.47)	1.687	(0.88)	0.37
A8	1.627	(1.57)	1.689	(0.03)	6.190	(1.63)	3.034	(0.87)	0.49
A9	-0.452	(1.82)	1.568	(0.03)	4.966	(1.63)	2.702	(0.95)	0.54
A10a	1.940	(2.39)	-3.390	(0.52)	-0.208	(0.93)	1.504	(1.30)	
A10b	-0.349	(0.54)	0.586	(0.02)	0.294	(0.04)			
A11	-1.025	(0.83)	1.699	(0.05)	1.383	(0.29)	-0.721	(0.42)	-0.52
Product 1	-3.024	(0.54)	-6.628	(0.43)	-27.510	(1.81)	---	---	---
Product 2	-2.753	(0.61)	-6.535	(0.46)	-27.282	(1.81)	---	---	---
Product 3	-2.860	(0.58)	-7.054	(0.52)	-27.621	(1.81)	---	---	---
Information Processing Parameters									
τ	---	---	-2.749	(0.01)	-2.405	(0.13)	---	---	---
ϕ (Row)	---	---	0.847	(0.00)	0.850	(0.03)	---	---	---
ϕ (Column)	---	---	0.564	(0.00)	0.402	(0.05)	---	---	---
Continuous Variables Endpoints									
			\overline{Min}_k	\overline{Max}_k	\overline{Min}_k	\overline{Max}_k			
A1	---	---	2.01	3.01	0.26	2.88	---	---	---
A4	---	---	25.57	94.42	43.94	76.06	---	---	---
A5	---	---	1.51	28.06	11.08	18.42	---	---	---
A10a	---	---	0.52	1.82	0.47	2.04	---	---	---
A10b	---	---	2.20	9.25	2.36	9.07	---	---	---

Parameter estimates in **bold** indicate that the 95% highest posterior density did not include 0, i.e., the parameter is significant at the 95% level.

Table 6. Hold-out Sample Predicted Probabilities

Shopper ID	Predicted Probability			
	Prod 1	Prod 2	Prod 3	None
1	0.219	0.106	0.514	0.161
2	0.006	0.008	0.750	0.236
3	0.193	0.093	0.451	0.262
4	0.186	0.074	0.267	0.473
5	0.381	0.019	0.014	0.587
6	0.241	0.074	0.046	0.640
7	0.080	0.191	0.072	0.657
8	0.017	0.303	0.015	0.665
9	0.085	0.126	0.066	0.723
10	0.124	0.091	0.056	0.729
11	0.157	0.063	0.012	0.768
12	0.020	0.188	0.018	0.773
13	0.063	0.092	0.057	0.788
14	0.063	0.092	0.057	0.788
15	0.063	0.092	0.057	0.788
16	0.063	0.092	0.057	0.788

Online Technical Appendix for Estimation Algorithms

The estimation algorithms will be given in detail for model 4 in Table 4, a model using the Max-Min method of determining the benefit of processing additional information, and assuming unopened product attributes are not used in the final product evaluation. Modifications for the Expected Value will also be detailed. Modifications for the Hybrid and Inferred Mean models are straightforward.

Max-Min, Not Used

We begin with the Max-Min, Not Used model (model 4 in Table 4) to setup key concepts and to provide an understanding of the basic structure of the algorithm. Three discrete pieces of information are available in the data set for each consumer: a sequence of choices of which product-attribute to inspect, a sequence of decisions on whether to continue processing new information or stop, and finally which of the three products to select for customization and purchase or the “none” option. Equations (6), (7), and (8) give the probabilities and equation (9) summarizes the likelihood function for individual i . Let K index the total number of attributes and K' index the number of continuous attributes. The parameters and their priors are specified as:

$$\beta_k \sim N(0, 100) \text{ for } k=1, \dots, K$$

$$\tau \sim N(0, 100)$$

$$x_k^{(\text{Max})} \sim N(0, 100) \text{ for } k'=1, \dots, K'$$

$$x_k^{(\text{Min})} \sim TN(0, 100, 0 < x_k^{(\text{Min})}) \text{ for } k'=1, \dots, K'$$

Where $N(\mu, \sigma^2)$ signifies a normal distribution with mean μ and variance σ^2 , and $TN(\mu, \sigma^2, c < x)$ is a corresponding truncated normal distribution with a lower truncation point c . Let the symbol “-“ represent the remaining elements of a set such that $\{\beta_{k-}\}$ indicates all members of the set excluding β_k .

Random-walk Metropolis-Hastings steps are used to sequentially draw parameters from their posterior distributions.

1. Generate $\beta_k | \{\beta_{k-}\}, \{x_k^{(\text{Max})}, x_k^{(\text{Min})}\}, \tau$ for $k=1, \dots, K$

Let $\beta_k^{(p)}$ represent the proposed new draw and $\beta_k^{(o)}$ represent the current or old draw. Form $\beta_k^{(p)} = \beta_k^{(o)} + \eta$ where η is a draw from $N(0, \sigma_k^2)$, where σ_k^2 is adjusted during the burn-in period to assure proper mixing with a target rejection rate of 50%.

For $i=1, \dots, I$ and $n_i = 1, \dots, N_i$, where n_i indexes the sequence of revealed attribute levels for person i and N_i is the total number of revealed product-attributes, initialize $\tilde{X}_i = 0$ and then:

- a) Compute the values of $\gamma_{ijk}^{n_i+}$ per equation (10) for all product-attributes which have not been revealed so far.
- b) Compute the choice probability for the revealed product-attribute given by equation (6).
- c) If $n_i < N_i$, compute the probability of continuing to process information given by equation (7).
 - i) Update the matrix \tilde{X}_i^n with the revealed product-attribute level x_{jk} .
 - ii) Return to step a)
- d) If $n_i = N_i$, compute the probability of not continuing to process information as one minus the quantity calculated in equation (7).
- e) If $n_i = N_i$, compute the final product choice probability given by equation (8).

Repeating steps a) – e) for $\beta_k^{(p)}$ and $\beta_k^{(o)}$ for each individual gives rise to the individual level likelihood functions $l_i^{(p)}$ and $l_i^{(o)}$ specified by equation (10). After repeating this over the sample of consumers, one can accept or reject the scalar $\beta_k^{(p)}$ with probability:

$$\text{Min: } \left(\frac{\prod_{i=1}^I l_i^{(p)} \times \exp\left(-\frac{1}{2 \times 100} (\beta_k^{(p)} - 0)^2\right)}{\prod_{i=1}^I l_i^{(o)} \times \exp\left(-\frac{1}{2 \times 100} (\beta_k^{(o)} - 0)^2\right)}, 1 \right)$$

2. Generate $\tau | \{\beta_k\}, \{x_k^{(\text{Max})}, x_k^{(\text{Min})}\}$

Let $\tau^{(p)}$ represent the proposed new draw and $\tau^{(o)}$ represent the current or old draw. Form $\tau^{(p)} = \tau^{(o)} + \eta$ where η is a draw from $N(0, \sigma_\tau^2)$, where σ_τ^2 is adjusted during the burn-in period to assure proper mixing with a target rejection rate of 50%.

Repeat steps a) – e) from step 1. Accept the new draw $\tau^{(p)}$ with probability:

$$Min: \left(\frac{\prod_{i=1}^I l_i^{(p)} \times \exp\left(-\frac{1}{2 \times 100} (\tau^{(p)} - 0)^2\right)}{\prod_{i=1}^I l_i^{(o)} \times \exp\left(-\frac{1}{2 \times 100} (\tau^{(o)} - 0)^2\right)}, 1 \right)$$

3. Generate $x_{k'}^{(Max)}, x_{k'}^{(Min)} \mid \{\beta_k\}, \{x_{k'-}^{(Max)}, x_{k'-}^{(Min)}\}, \tau$ for $k'=1, \dots, K$,

The values of $x_{k'}^{(Max)}$ and $x_{k'}^{(Min)}$ are drawn jointly and are centered on the actual mean value of the attribute levels across the products. Let $\bar{x}_{k'}$ represent the actual mean value of the continuous attribute k' as calculated from the product attribute matrix. Let $r_{k'}^{(o)}$ equal the range, $x_{k'}^{(Max)[o]} - x_{k'}^{(Min)[o]}$. Form $x_{k'}^{(Max)[pl]} = \bar{x}_{k'} + 0.5 * (r_{k'}^{(o)} + \eta)$ and $x_{k'}^{(Min)[pl]} = \bar{x}_{k'} - 0.5 * (r_{k'}^{(o)} + \eta)$, where η is a draw from $N(0, \sigma_{k'}^2)$ and where $\sigma_{k'}^2$ is adjusted during the burn-in period to assure proper mixing with a target rejection rate of 50%. If $x_{k'}^{(Min)[pl]} < 0$, then set $x_{k'}^{(Min)[pl]} = 0$.

Repeat steps a) – e) from step 1. Accept the new draws of $x_{k'}^{(Max)[pl]}$ and $x_{k'}^{(Min)[pl]}$ with probability:

$$Min: \left(\frac{\prod_{i=1}^I l_i^{(p)} \times \exp\left(-\frac{1}{2 \times 100} (x_{k'}^{(Max)[pl]} - 0)^2\right) \times \exp\left(-\frac{1}{2 \times 100} (x_{k'}^{(Min)[pl]} - 0)^2\right)}{\prod_{i=1}^I l_i^{(o)} \times \exp\left(-\frac{1}{2 \times 100} (x_{k'}^{(Max)[o]} - 0)^2\right) \times \exp\left(-\frac{1}{2 \times 100} (x_{k'}^{(Min)[o]} - 0)^2\right)}, 1 \right).$$

Note that the normalizing constant for the truncated normal prior on $x_{k'}^{(Min)}$ cancels out in the calculation of the acceptance probability.

All parameters are drawn individually except for A10a and A10b which are drawn together since they are revealed at the same time. Attribute A10a has its own values of β_k and $x_{k'}^{(Max)}, x_{k'}^{(Min)}$ as does Attribute A10b. However, when calculating step 1a there is only one value of $\gamma_{ijk}^{n_i+}$ which determines the expected value of jointly revealing A10a and A10b.

For models using the Actual Mean, initial values in the matrix \tilde{X}_i are set equal to mean value for each continuous product attribute, calculated from the actual product attribute matrix. For

discrete attributes, $\bar{x}_k = 0.5$ or $\bar{x}_k = -0.5$ each with probability 0.5. Let K^* index the discrete attributes. On each iteration of the MCMC chain:

4. Generate $\bar{x}_{k^*} | \{\bar{x}_{k^*-}\} \{\beta_k\}, \{x_{k^*}^{(Max)}, x_{k^*}^{(Min)}\}, \tau$ for $k^*=1, \dots, K^*$

Let $l_i^{(\bar{x}_{k^*}=0.5)}$ represent the likelihood for individual i evaluated at $\bar{x}_{k^*} = 0.5$ and $l_i^{(\bar{x}_{k^*}=-0.5)}$ represent the likelihood for individual i evaluated at $\bar{x}_{k^*} = -0.5$. Then choose $\bar{x}_{k^*} = 0.5$ with probability:

$$\left(\frac{\prod_{i=1}^I l_i^{(\bar{x}_{k^*}=0.5)}}{\prod_{i=1}^I l_i^{(\bar{x}_{k^*}=0.5)} + \prod_{i=1}^I l_i^{(\bar{x}_{k^*}=-0.5)}} \right)$$

Otherwise $\bar{x}_{k^*} = -0.5$. This is an example of a Griddy-Gibbs with two possible values and an equal prior probability on each.

For models using the Inferred Mean method, as detailed in the text, \tilde{X}_i is comprised of vectors \tilde{x}_{jk} of length $2K$. The first K rows of \tilde{X}_i are initialized to 0 and the next $K+1$ to $2K$ rows are set equal to 1. In the algorithm above, expand the vector β to be length $2K$ and updates to all parameters follow as detailed above. Change step 1.c.i. to the following:

- i) When a consumer inspects the cell corresponding to product j and attribute k , the value x_{jk} replaces the corresponding 0 for \tilde{x}_{ijk} and in the second half of \tilde{x}_{ij} , the value of \tilde{x}_{ijk+K} is switched from 1 to 0.

Expected Value, Not Used

The major changes to the Expected Value, Not Used model (model 2 in Table 4) are drawing the parameters for the prior distributions of x_k and evaluating the value of γ_{ijk}^{n+} in equation (4). The parameters and their priors are specified as:

$$\ln(\beta_k) \sim N(0, 100) \text{ for } k=1, \dots, 9 \text{ and } 11$$

$$\beta_k \sim N(0, 100) \text{ for } k=10 \text{ and product intercepts}$$

$$\tau \sim N(0, 100)$$

$$x_k \sim \log \text{ normal}(\mu_k, \sigma_k^2) \text{ for continuous attributes}$$

$$CV(x_k) \sim U(0, 5)$$

$x_k \sim \text{Bernoulli}(0.5)$ for discrete attributes

Although we assume continuous $x_k \sim \text{log normal}(\mu_k, \sigma_k^2)$, we found it easier to specify and estimate the model in terms of $E(x_k)$ and $\text{CV}(x_k)$, the coefficient of variation. In particular, the model mixed much more efficiently by drawing $\text{CV}(x_k)$. We assume $E(x_k) = \bar{x}_k$, the mean of the actual product attributes.

Random-walk Metropolis-Hastings steps are again used to sequentially draw parameters from their posterior distributions.

1. *Generate* $\text{CV}(x_k) | \{\text{CV}(x_k)\}, \{E(x_k)\} \{\beta_k\}, \tau$ for $k=1, \dots, K$

Let $\text{CV}(x_k^{(p)})$ represent the proposed new draw and $\text{CV}(x_k^{(o)})$ represent the current or old draw. Draw $\text{CV}(x_k^{(p)})$ from a uniform $U(0,5)$ distribution.

a. Given $E(x_k)$ and $\text{CV}(x_k^{(p)})$ compute $\text{Var}(x_k^{(p)})$ as:

$$\text{Var}(x_k^{(p)}) = \left[E(x_k) \times \text{CV}(x_k^{(p)}) \right]^2$$

b. Given $E(x_k)$ and $\text{Var}(x_k^{(p)})$ compute $\sigma_k^{2(p)}$ for the lognormal distribution as:

$$\sigma_k^{2(p)} = \ln \left[\frac{e^{2 \ln E(x_k)} + \text{Var}(x_k^{(p)})}{e^{2 \ln E(x_k)}} \right]$$

c. Given $E(x_k)$ and $\sigma_k^{2(p)}$ compute $\mu_k^{(p)}$ for the lognormal distribution as:

$$\mu_k^{(p)} = \ln \left[E(x_k) \right] - \frac{\sigma_k^{2(p)}}{2}$$

Step a. is from the definition of the coefficient of variation; steps b. and c. are from the definition of the lognormal distribution.

For $i=1, \dots, I$ and $n_i = 1, \dots, N_i$, where n_i indexes the sequence of revealed attribute levels for person i and N_i is the total number of revealed product-attributes:

a) The value of γ_{ijk}^{n+} from equation (4) is estimated via simulation for all product-attributes which have not been revealed so far. For $m=1, \dots, M$:

- i) Draw $x_k^{(m)}$ from $N(\mu_k^{(p)}, \sigma_k^{2(p)})$, and compute $x_k^{(m)} = e^{x_k^{(m)}}$ for continuous attributes. Draw $x_k^{(m)}$ from Bernoulli(0.5) for discrete attributes, where $x_k^{(m)} = 0.5$ with probability 0.5 and $x_k^{(m)} = -0.5$ with probability 0.5.
- ii) Compute $Z^{(m)} = \ln(\iota' \exp[\beta' X_i^{n+}])$ where the draw of $x_k^{(m)}$ is inserted into the appropriate row and column of X_i^{n+} .
- iii) Compute $\hat{Z} = \frac{1}{M} \sum_{m=1}^M Z^{(m)}$
- iv) Compute $\hat{\gamma}_{ijk}^{n+} = \hat{Z} - \ln(\iota' \exp[\beta' X_i^n])$ following equation (4).

The value $\hat{\gamma}_{ijk}^{n+}$ is used to compute individual level likelihoods following the same procedures b) – e) from the Max-Min, Not Used model algorithm, step 1. The draws of $x_k^{(m)}$ are done using a modified latin hypercube sampling scheme with $M=20$; other values of M were evaluated using simulated data and values greater than 20 offered no substantive improvement in parameter estimates. Accept the new draw $CV(x_k^{(p)})$ with probability:

$$Min: \left(\frac{\prod_{i=1}^I L_i^{(p)}}{\prod_{i=1}^I L_i^{(o)}}, 1 \right)$$

Note that the likelihood is not weighted by the prior because $CV(x_k^{(p)})$ is drawn directly from its prior distribution.

Draws of other parameters and modifications for Actual Mean follow those outlined earlier.

Online Technical Appendix for Inferred Mean Method for Unseen Product-Attributes

The “Actual Mean” method of dealing with unseen product-attributes relies quite heavily on the specified value of \bar{x}_k since it appears in the indirect utility function. The “Not Used” method requires the potentially problematic assumption that consumers use different information sets for deciding the benefit of inspecting product-attributes compared to the final product choice; we would like to avoid these assumptions. We build on Branco, Sun, and Villas-Boas (2012) and propose an expectation deviation method of handling unseen product-attributes. In our model, there is an intercept term which represents the product labeled for each column in the IDB. Consider a product class with only two attributes, then the indirect utility is given by:

$$V_{ij} = \beta_0^{\bar{x}} + \beta_1 x_{j1}^* + \beta_2 x_{j2}^* + \varepsilon_{ij}. \quad (\text{TA3})$$

Here the product-attribute values x_{jk}^* represent deviations from the expected values \bar{x}_k and β_0 is the indirect utility when the product-attribute levels all equal their expected values. If we assume that for unseen product-attribute levels that $x_{jk} = \bar{x}_k$, then before any product-attribute levels are inspected, $V_{ij} = \beta_0^{\bar{x}} + \varepsilon_{ij}$. Rewriting (TA3) in expectation deviation format and expanding $\beta_0^{\bar{x}}$ we get:

$$V_{ij} = \beta_0^* + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2 + \beta_1 (x_{j1} - \bar{x}_1) + \beta_2 (x_{j2} - \bar{x}_2) + \varepsilon_{ij}. \quad (\text{TA4})$$

Here $\beta_0^{\bar{x}} = \beta_0^* + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2$. When product-attribute 1 is inspected and product attribute 2 is not, we get:

$$V_{ij} = \beta_0^* + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2 + \beta_1 (x_{j1} - \bar{x}_1) + \beta_2 (\bar{x}_2 - \bar{x}_2) + \varepsilon_{ij} \quad (\text{TA5})$$

$$V_{ij} = \beta_0^* + \beta_2 \bar{x}_2 + \beta_1 x_{j1} + \varepsilon_{ij}.$$

This result suggests a new parameterization of the indirect utility function. Let $d_{ijk} = 1$ when product-attribute j, k has not been inspected by consumer i and $d_{ijk} = 0$ if it has. Then:

$$V_{ij} = \beta_0^* + \beta_1 x_{j1} (1 - d_{ij1}) + \beta_2 x_{j2} (1 - d_{ij2}) + \delta_1 d_{ij1} + \delta_2 d_{ij2} + \varepsilon_{ij}. \quad (\text{TA6})$$

In this model $\delta_k = \beta_k \bar{x}_k$. Here we can again assume that if a product-attribute is not inspected that $x_{jk} = \bar{x}_k$ but we do not need to specify \bar{x}_k , it is estimated in the parameter δ_k . We will refer to this method of handling unseen product-attributes as “Inferred Mean.”

The Inferred Mean method requires expanding the parameter vector β by K , the number of attributes in the product choice set, in order to estimate δ . In terms of our model set-up, the vector \tilde{x}_{ij} is now of length $2 \times K$. Before any cells are inspected in the IDB, the first $1, \dots, K$ elements of \tilde{x}_{ij} are set equal to 0 while the last $K+1, \dots, 2 \times K$ elements are set equal to 1, with these last elements representing the binary variables in equation (13). When a consumer inspects the cell corresponding to product j and attribute k , the value x_{jk} replaces the corresponding 0 for \tilde{x}_{ijk} and in the second half of \tilde{x}_{ij} , the value of \tilde{x}_{ijk+K} is switched from 1 to 0. The sequence of product-attributes revealed by a consumer and the differences between the final set of product-attributes across consumers identifies the binary variables in equation (13). The remainder of the model set-up does not change.