Available online at www.sciencedirect.com
ScienceDirect
Journal of Interactive Marketing 29 (2015) 11 – 25
JOURNAL OF INTERACTIVE MARKETING
www.elsevier.com/locate/intmar

Information Accessed or Information Available? The Impact on Consumer Preferences Inferred at a Durable Product E-commerce Website☆

Imran S. Currim [a,1] & Ofer Mintz [b,*,1] & S. Siddarth [c,1]

[a] *Paul Merage School of Business, University of California, Irvine, CA 92697, USA*
[b] *E. J. Ourso College of Business, Louisiana State University, Baton Rouge, LA 70803, USA*
[c] *Marshall School of Business, University of Southern California, Los Angeles, CA 90089, USA*

Abstract

Most previous choice modeling research infers preferences by assuming that consumers consider all the information available at the point-of-purchase. Because e-commerce sites increasingly incorporate tracking technologies that can monitor consumer behavior on their site, our research studies how incorporating the information accessed by consumers into a choice model impacts model performance and inferred preferences. We use data from an electronic goods manufacturer that monitored the attribute information accessed by 582 shoppers while they made Customize and Buy decisions at the firm's website. We find that incorporating the information *accessed* by consumers into the choice model provides more valid estimates of attribute preferences and better fitting choice models than models based on information *available*. Because firms can easily obtain this type of information as a by-product of their online operations, we propose that managers who monitor information acquisition and apply the information accessed model will have a useful methodology to gain a better understanding of consumer preferences.

Keywords: Multi-attribute models; Consumer choice; Revealed preferences; Electronic commerce

Introduction

Imagine that the manager of the Kindle product line at Amazon wants to gain a better understanding of customer preferences with a view of making better product design, pricing, advertising, and targeting decisions. She is aware that there are hundreds of shoppers from all over the world at the Amazon website at any point of time considering various Kindle configurations presented in the form of a comparison chart with information on a variety of attributes such as price, connectivity, content, display, and battery life, as shown in Fig. 1. She has the ability to follow the clickstream of potential shoppers up to the point that they make a purchase decision, and wants to use the observed choices to obtain insights about consumer preferences for different attributes and alternatives.

An obvious way of achieving this objective is via a choice model that incorporates all the attribute information available in the comparison chart at the point-of-purchase as is common in the choice modeling literature. However, extensive laboratory research has shown that consumers typically do not access all information at the point-of-purchase due to search costs, information overload, prior knowledge, or heuristic-based shopping. Therefore, with 7 different Kindle configurations and 11 different attributes, she expects that shoppers may not pay attention to all of the available attribute data for all alternatives and wonders whether and how much the choice model's performance and diagnostics would improve if it incorporated the specific cells that a shopper actually looked at.

☆ The authors thank Internet Technology Group, Inc. (ITGi) for providing data; Randy Bucklin, Rico Bumbaca, Tim Gilbride, Jooseop Lim, and Ken Wilbur for their helpful comments; and the Dean's Office of the Paul Merage School of Business (SF10682) for the financial support.
* Corresponding author at: 2117 Business Education Complex, Louisiana State University, Baton Rouge, LA 70803.
E-mail addresses: iscurrim@uci.edu (I.S. Currim), omintz@lsu.edu (O. Mintz), siddarth@usc.edu (S. Siddarth).
[1] The authors contributed equally to this article and are listed in alphabetical order.

http://dx.doi.org/10.1016/j.intmar.2014.09.003
1094-9968/

	Kindle	Kindle Touch	Kindle Touch 3G	Kindle Keyboard	Kindle Keyboard 3G	Kindle DX	Kindle Fire
	$79.00	$99.00	$149.00	$139.00		$379.00	$199.00
Connectivity							
	Wi-Fi	Wi-Fi	Free 3G + Wi-Fi	Wi-Fi	Free 3G + Wi-Fi	Free 3G	Wi-Fi
Content							
	Millions of books, newspapers, magazines, games, and docs	Millions of books, newspapers, magazines, audiobooks, games, and docs		Millions of books, newspapers, magazines, audiobooks, games, and docs		Millions of books, newspapers, magazines, audiobooks, games, and docs	20 million movies, TV shows, apps, games, songs, books, newspapers, audiobooks, magazines, and docs
Web							
	Experimental browser	Experimental browser		Experimental browser		Experimental browser	Amazon Silk cloud-accelerated browser
Display							
	6" E Ink Pearl	6" E Ink Pearl		6" E Ink Pearl		9.7" E Ink Pearl	7" Vibrant Color IPS
Battery Life, Wireless Off							
	1 month	2 months		2 months		3 weeks	8 hours continuous reading or 7.5 hours video playback
Storage							
	2GB on device for 1,400 books	4GB on device for 3,000 books		4GB on device for 3,500 books		4GB on device for 3,500 books	8GB on device for 80 apps plus either 10 movies or 800 songs or 6,000 books
	Plus free cloud storage for all Amazon content so you never have to worry about running out of space	Plus free cloud storage for all Amazon content so you never have to worry about running out of space		Plus free cloud storage for all Amazon content so you never have to worry about running out of space		Plus free cloud storage for all Amazon content so you never have to worry about running out of space	Plus free cloud storage for all Amazon content so you never have to worry about running out of space
Dimensions							
	6.5" x 4.5" x 0.34"	6.8" x 4.7" x 0.40"		7.5" x 4.8" x 0.34"		10.4" x 7.2" x 0.38"	7.5" x 4.7" x 0.45"
Weight							
	5.98 ounces	7.5 ounces	7.8 ounces	8.5 ounces	8.7 ounces	18.9 ounces	14.6 ounces
Interface							
	5-way controller	multi-touch		keyboard		keyboard	multi-touch

Fig. 1. Comparison chart.

Because the choice model parameters are key to providing insights into how different features impact consumer choices and are the basis for managerial decisions, this central research question is especially important for online retailers of durable products who could easily collect this information as a by-product of their regular operations using approaches similar to ours.

To accomplish our objective, we employ data from an electronic goods manufacturer who monitored the attribute information accessed by shoppers while they made Customize and Buy decisions at the firm's website. We develop and estimate choice models that incorporate the observed choice set and attribute levels shoppers actually examine to investigate whether they provide more valid attribute preferences and improve model fits over those achieved based on a model calibrated using all the information available to consumers. The potential improvement from employing information accessed, i.e., the specific cells of the comparison matrix the consumer looked at, over employing information available, i.e., all the cells of the comparison matrix on the website, on choice model diagnostic and predictive abilities can be tested via two types of models. The first assumes that both sets of information, accessed and available, are exogenous to final choice, while the second assumes information accessed as endogenous. We begin by formulating and empirically testing models of the first type that model the final choice. Subsequently, in the additional analysis section, we discuss the results from estimating models of the second type, which also incorporate a customer's cell clicking behavior in addition to the final choice.

This work makes the following theoretical, methodological, substantive, and managerial contributions. Theoretically, we bridge two important literatures. The first is psychology-based consumer information processing theory (e.g., Bettman 1979; Dhar and Nowlis 2004; Payne 1976; Payne, Bettman, and Johnson 1993), which describes how product attribute information is accessed during the choice process, but does not use it to predict consumers' choice decisions or to infer consumer preferences. The second is utility-based choice theory (e.g., Lancaster 1966; Ratchford 1975) which infers consumer preferences for product attributes, or response to the marketing mix, using choice data. However, because researchers in this stream of research are not privy to the actual information accessed by shoppers, the resulting models are typically estimated based on all the product information that is available to consumers. The bridge between the two literatures is accomplished by employing the actual information accessed by shoppers and not all information available, to infer consumer preferences and predict choices. Methodologically, we propose a set of new models based on only the information accessed by shoppers. Substantively, we compare and demonstrate the value of employing only the information accessed by shoppers in the choice model, over employing all information available, which has not been accomplished heretofore. Taken together, we offer a practical modeling methodology for managers based on incorporating information actually accessed by shoppers into choice models.

Managerially, shoppers at manufacturer (e.g., Apple), retailer (e.g., Amazon), or intermediary (e.g., CNET) websites are often provided product comparison charts such as Fig. 1 or with tools that enable them to make attribute-by-attribute comparisons of products that they are interested in. It would be relatively easy for a firm that already provides its shoppers with the ability to do side-by-side comparisons to conduct a natural field experiment with a sample of shoppers over a limited time period to capture data on information accessed. This could be accomplished, as we demonstrate, by presenting the sample of shoppers with a product–attribute matrix in which the values in the cells are hidden and shoppers are required to click on cells to reveal the values. Alternatively, this could be accomplished by employing other technologies readily available like clickstream monitoring, eye-tracking, or mouse-tracking. Our results suggest that it would benefit firms to gather such data to provide a better basis for product design, pricing, advertising, and targeting decisions because it provides better diagnostics about consumer preferences and choices.

Background

Normative theory suggests that consumers access all information relevant to their decision and make trade-offs to optimize their choice. However, a fundamental problem with inferring preferences based on normative choice models is that limited cognitive resources drive consumers to not access all information available and instead employ heuristics to simplify decision making (Simon 1956). For example, many lab-based studies on information processing research have described how information load and prior knowledge can lead consumers to not access all the information available (Jacoby, Chestnut, and Fisher 1978; Lurie 2004; Malhotra, Jain, and Lagakos 1982). In contrast, constructive choice theory suggests that consumers make choices based on a trade-off between minimizing cognitive effort and maximizing accuracy (Bettman, Luce, and Payne 1998; Payne, Bettman, and Johnson 1993). Thus, consumers will access information until their perceived uncertainty for choice is sufficiently reduced to make a decision (Jacoby et al. 1994; Moorthy, Ratchford, and Talukdar 1997).

Based on a hybrid of normative and constructive choice theories, the current choice model literature, particularly the consider-then-choose models using scanner panel data (e.g., Andrews and Srinivasan 1995; Musalem et al. 2010; Siddarth, Bucklin, and Morrison 1995; Wu and Rangaswamy 2003) and models that estimate screening rules based on data gathered from choice-based conjoint (CBC) experiments (e.g., Gilbride and Allenby 2004), attempt to predict consumer choices and preferences. The former set of studies show that accounting for consumer consideration sets can improve choice model predictions, while the latter set of studies show that *imputing* cut-off levels for different attributes depending upon the choice rule being modeled (e.g., conjunctive, disjunctive) improves diagnostics about parameter estimates. However, both sets of studies rely on imputed choice sets and assume that consumers consider all information presented.

On the other hand, our methodology tracks the alternatives and attributes actually accessed by consumers and directly incorporates the alternative and attribute restrictions into the choice model. Not surprisingly, consistent with constructive choice theory, and consider-then-choose and screening rule models, we find that only 51% of shoppers in our retail setting access information on all alternatives, only 48% of shoppers access information on all attributes, and only 9% of shoppers access all alternative–attribute information.

Another important feature of previous lab-based screening rules and information processing research is that the choice alternatives presented to consumers are derived from an experimental design, which ensures that the chosen levels minimize attribute correlation across alternatives and respondents. In contrast, our sample of shoppers saw actual products in a real online retail setting in which multiple (sometimes even all) choice alternatives shared the *same* level of an attribute. This important characteristic of "real-world" comparison data at retail websites means that the resulting X-matrix may not vary much across shoppers or alternatives, with potential identification problems that can result in parameter estimates with incorrect signs and counter-intuitive diagnostics about attribute preferences. In these settings, we expect that using the actual information accessed by consumers in a choice model can potentially avoid identification problems because it generates variance in the X-matrix across shoppers. We therefore develop an approach to collect data on information accessed by consumers and incorporate this information in a model that reflects their choice-set and attribute restrictions.

Data

Our data consist of the product–attribute information accessed by 582 shoppers who visited a well-known electronic manufacturer/retailer's website to shop for a popular durable product category, similar to but different from a tablet. Access to the data for publication purposes was provided under the condition that the identity of the manufacturer/retailer and the product category remain confidential. The Decision Board (DB) (Mintz et al. 1997) was installed on the firm's website to collect the product–attribute information accessed by shoppers who visited the website on a particular weekend. The 582 shoppers that navigated to the DB product–attribute comparison chart web page were able to observe three actual products or choice alternatives and 11 attribute descriptors but not the specific attribute values in each cell, which were kept hidden. Shoppers were told that they could observe the features or attribute values in a cell by clicking on it, as is typical in behavioral information processing studies, before transitioning to their Customize and Buy decision, which was also located on the same webpage.[2]

This data collection procedure with alterations to a real website enabled us to obtain revealed preferences from actual shoppers who were unaware that they were part of a consumer behavior study.[3] The requirement for customers to click on individual cells to access attribute values is similar to many popular websites like Facebook, Expedia, and CNET that require users to click on cells or links to access (more) information about attributes or alternatives. It is also similar to clickstream, eye-tracking, and mouse-tracking market research techniques that many websites employ to study customer behavior.

A mockup of the matrix, as seen by a consumer during an intermediate stage of purchase, using the Amazon Kindle setting as an example, is shown in Fig. 2. All 582 shoppers were presented with the same matrix of alternatives, a decision made by the manufacturer/retailer. Each shopper in our dataset accessed more than one cell in the product–attribute matrix. A Customize and Buy command tab was prominently displayed at the bottom of each column or alternative similar to the Choose tab in Fig. 2. The DB kept track of the product–attribute cells accessed by each shopper and whether the shopper clicked on the Customize and Buy tab. As is common in many Internet-based retail settings, no other information, such as consumer demographics or prior shopper knowledge, was captured.[4]

Information Available

The matrix of information presented to shoppers appears in Panel A of Table 1. We provide ordinal rankings for each attribute based on its level for each alternative and show the attribute level for alternative. For attributes A1–A9 and A11, higher values indicate preferred attribute levels; while attributes 10a and 10b, which were observed together when a cell in row 10 was accessed, describe physical aspects of the product (e.g., height and weight) for which it is a priori difficult to judge whether larger or smaller values are more desirable. 42% of shoppers purchased alternative 1, 5% purchased alternative 2, 5% purchased alternative 3, and 48% did not purchase. Notice that alternative 1 is the lowest priced alternative while alternatives 2 and 3 have the same higher price. This may partially explain why alternative 1 has a greater share of purchases, but it is important to note that it does not dominate the other two alternatives completely since it has inferior levels of attribute A11 (relative to alternative 2) and attributes A5 and A6 (relative to alternative 3).

[2] Once a consumer clicks on an attribute to view the attribute he/she does not have to go back to view the original product-attribute matrix to view a different attribute, they are just viewing the same product-attribute matrix with opened (i.e., all the previous cells they have opened) and unopened cells. Consequently, similar to previous laboratory based studies using Information Display Boards, the consumer does not need to remember/recall the attribute for the particular alternative, making the task simple and straightforward.

[3] Previous experimental laboratory research using IDB computerized decision process tracers show that subjects display similar information processing, choices, and eye movements to subjects that are not using IDB computerized decision process tracers to make a decision (Johnson, Meyer, and Ghose 1989; Johnson, Payne, and Bettman 1988).

[4] To the extent that consumers knew some of the attribute levels based on prior visits that are not recorded in our data, it makes our record of the information accessed less accurate and works against our expectations that information accessed will improve the choice model performance. However, because we find that a model based on the recorded information accessed generated the best diagnostics on attribute preferences and estimation and validation sample fits, it implies that relatively little data about the information accessed was missing. Of course, keeping track of the information accessed on prior visits would only strengthen the empirical results.

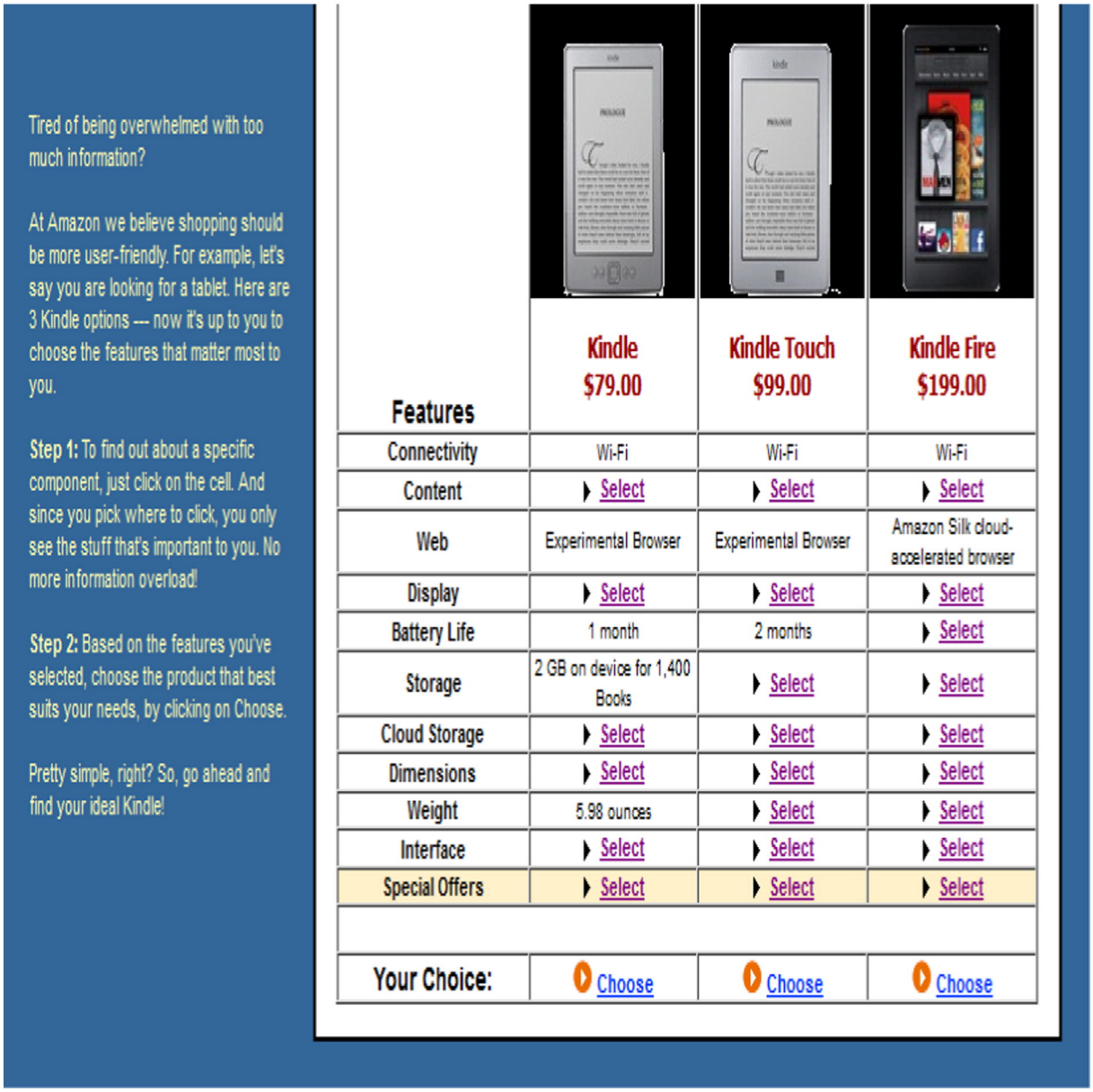

Fig. 2. Mock-up of the matrix presented to shoppers by the Decision Board Platform using Amazon Kindle as an example.

One interesting feature of the data is that the levels for five of the attributes (A2, A3, A4, A7, and A9) are exactly the same across all alternatives and, further, two of the three alternatives also have the same level for five other attributes (price, A1, A6, A8, and A11). Thus, our real world purchase setting not only has an X-matrix that is the same for *all* shoppers but also has multiple attributes with levels that are the same across alternatives. These features make tracking of the information accessed by a consumer critical to achieve variance in the X-matrix across alternatives, attributes, and consumers and for identifying model parameters correctly.

Attributes Accessed

The percentage of shoppers accessing attribute information is presented in Panel B of Table 1. Shoppers access an average of 8.4 attributes out of 11. Notice that only 52% of shoppers access information on all 11 attributes, the rest do not. Of those who choose to customize and buy, 59% access information on all attributes, while of those who choose not to customize and buy, 44% access information on all attributes. This indicates that consistent with the behavioral information processing and information overload literatures, a large percentage of shoppers (48%) do not access information on all attributes and that this observation is largely true for those who decide to customize and buy (41%) as well as those who decide to not customize and buy (56%).

Alternatives Accessed

An alternative was considered to be accessed by a consumer if at least one attribute level cell for that alternative was clicked on during the shopping trip. Panel C reports the percentage of

shoppers who accessed information on each choice alternative. Shoppers access an average of 2.1 alternatives. 51% of shoppers access information on all three alternatives, 8% on two alternatives, and 41% on only one alternative. Of those who choose to customize and buy, 47% access information on all three alternatives, while of those who choose to not

Table 1
Matrix of information presented to shoppers and descriptive statistics of information accessed by shoppers.

Panel A. Matrix of information presented to shoppers

Attribute	Row	Continuous or discrete	Range	Alternative 1 (Level)	Alternative 2 (Level)	Alternative 3 (Level)
Price	N/A	Continuous	$699.99–$849.99	1	2	2
A1	1	Continuous	1.3–1.5	2	1	1
A2	2	Discrete	1	1	1	1
A3	3	Discrete	1	1	1	1
A4	4	Continuous	40	1	1	1
A5	5	Continuous	14.1–15.4	2	1	3
A6	6	Discrete	1 or 2	1	1	2
A7	7	Discrete	1	1	1	1
A8	8	Discrete	1 or 2	2	1	2
A9	9	Discrete	1	1	1	1
A10a	10	Continuous	1–1.4	1	2	3
A10b		Continuous	5.29–6	2	1	3
A11	11	Discrete	1 or 2	1	2	1

Prices and alternative (model) names were available to shoppers without clicking a cell; Attributes 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. For the remaining attributes, lower levels indicate less value.

Panel B. Percent of consumers accessing attribute information

Number of attributes accessed	Number of shoppers	%	Customize and buy	%	Do not customize and buy	%
1	30	5%	8	3%	22	8%
2	27	5%	7	2%	20	7%
3	38	7%	21	7%	17	6%
4	23	4%	9	3%	14	5%
5	25	4%	10	3%	15	5%
6	27	5%	7	2%	20	7%
7	18	3%	9	3%	9	3%
8	23	4%	11	4%	12	4%
9	29	5%	14	5%	15	5%
10	39	7%	27	9%	12	4%
11	303	52%	178	59%	125	44%

Panel C. Percent of consumers accessing information for each alternative

Number of alternatives accessed	Number of shoppers	%	Customize and buy	%	Do not customize and buy	%
1	241	41%	139	46%	102	36%
2	47	8%	22	7%	25	9%
3	294	51%	140	47%	154	55%

Panel D. Percentage of shoppers accessing information in different cells

Attribute	Alternative 1	Alternative 2	Alternative 3	Overall
A1	86%	39%	37%	93%
A2	79%	29%	26%	86%
A3	78%	31%	29%	86%
A4	73%	28%	25%	79%
A5	70%	26%	24%	76%
A6	68%	28%	25%	73%
A7	66%	22%	20%	71%
A8	62%	22%	20%	67%
A9	63%	22%	20%	68%
A10	56%	19%	17%	60%
A11	70%	32%	30%	75%
Overall	95%	59%	55%	9%

Table 1 *(continued)*

Panel E. Percent of consumers accessing total amount of information

Number of cells accessed	Number of shoppers	%	Customize and buy	%	Do not customize and buy	%
<5	97	17%	36	12%	61	22%
6–10	103	18%	49	16%	54	19%
11–15	233	40%	137	46%	96	34%
16–20	36	6%	15	5%	21	7%
21–25	33	6%	21	7%	12	4%
26–30	21	4%	14	5%	7	2%
>30	59	10%	29	10%	30	11%
Total	582	–	301	52%	281	48%

Panel F. Attribute information accessed on an alternative and purchase of the alternative

Attribute/alternative	Accessed/not accessed	Purchased alternative 1	Purchased alternative 2	Purchased alternative 3
A1	Accessed	40%	4%	5%
	Not accessed	2%	1%	0%
A2	Accessed	37%	3%	4%
	Not accessed	4%	2%	1%
A3	Accessed	38%	4%	5%
	Not accessed	3%	1%	1%
A4	Accessed	35%	3%	4%
	Not accessed	7%	2%	1%
A5	Accessed	34%	3%	4%
	Not accessed	7%	2%	1%
A6	Accessed	34%	3%	4%
	Not accessed	8%	2%	1%
A7	Accessed	34%	3%	4%
	Not accessed	8%	2%	2%
A8	Accessed	31%	3%	4%
	Not accessed	10%	2%	2%
A9	Accessed	33%	3%	3%
	Not accessed	9%	2%	2%
A10	Accessed	30%	3%	3%
	Not accessed	12%	2%	3%
A11	Accessed	34%	4%	4%
	Not accessed	7%	1%	2%

customize and buy, 55% access information on all three alternatives. Overall, once again, consistent with the behavioral information processing and information overload literatures, a large percentage of shoppers (49%) did not access information on all alternatives, with minor differences among those who decided to customize and buy (53%) versus those who did not (45%).

Alternatives and Attribute Levels Accessed

The percentage of shoppers accessing information in different cells is presented in Panel D of Table 1. Of our sample of shoppers, only 9% access all information cells. Further, while 95% of shoppers access some information on alternative 1, the low-priced alternative, a large percentage of shoppers, between 55% and 59% access some information on alternatives 2 and 3, implying that the amount of information accessed varies substantially across alternatives.

Total Information Accessed

The percentage of consumers who access different amounts of information is presented in Panel E. The total amount of information that can be accessed is 33 cells (3 alternatives × 11 attributes). Shoppers access an average of 13.4 cells out of 33; with 75% of shoppers accessing 15 or fewer cells, or less than half of the available information. This level of access is about the same for those who chose to customize and buy (74% access 15 or fewer cells) and for those who did not (75% access 15 or fewer cells).

Relationship Between Information Accessed and Purchase

In Panel F, we show that the percentage of shoppers purchasing a particular alternative is found to be much higher if shoppers access any attribute of the alternative relative to if they do not access the attribute of the alternative. For example, of the shoppers who access attribute A1 of alternative 1, 40% purchase the alternative while 2% do not. This is true of 32 of the 33 paired comparisons in Panel F, indicating that accessed information is associated with purchasing the alternative and not merely with just getting noticed.

These statistics lead to three conclusions. First, contrary to the practice of using all data presented or available to the consumer to estimate current consumer preference and choice models, almost

half of the shoppers do not access information on all attributes and this observation is true for those who decide to customize and buy as well as those who decide to not customize and buy. While such a result is entirely expected from the constructive choice theory and information processing-based points of view, what is new here is that the subset of attribute information accessed by a shopper has not heretofore been observed and used to calibrate a choice model. Second, almost half of the shoppers do not access information on all alternatives and this observation is true for those who decide to customize and buy as well as those who do not, a pattern that is entirely expected based on the constructive choice theory and information overload literatures (e.g., Lurie 2004; Malhotra, Jain, and Lagakos 1982). What is new about our work is that we actually observe the subset of alternatives about which product–attribute information is accessed by a shopper and employ this information to calibrate a choice model.

Third, only 9% of shoppers are found to access all the feature information available. In addition, almost 75% access less than half of all the feature information available and this observation is found to be true for those who customize and buy as well as those who choose not to customize and buy. Again, while such a result is entirely expected from constructive choice theory, information processing, and information overload literatures, what is new here is that the feature information accessed by the shopper has not heretofore been observed and used to calibrate a choice model. Consequently, it is likely that models based on information accessed will perform better than models based on information available.

Models

We propose and estimate several multi-attribute choice models based on the literature progressively ranging from information available to information accessed in the comparison chart. For example, if a shopper clicked on and looked at all cells in the comparison chart there would be no difference between the information available in the comparison chart and the information accessed, since all available information was accessed. However, if a shopper only clicked on some cells of the comparison chart then the two information sets would be different. In this case, the subset of information accessed may be more relevant to the purchase decision and revealed preferences. Consequently, a choice model based on the information accessed may provide a better fit and generate better diagnostics about attribute preferences, similar to the way in which choice models that incorporate consideration sets improve on models that assume the consumer considers all alternatives.

We begin by specifying the utility function for each consumer and product alternative and specify a general equation to relate the utilities to the choice probability based on the information accessed by each consumer. Subsequently, we specify alternative benchmark models that progressively account for different levels of information available or acquired by the consumer and are representative of the most common choice models estimated in the literature.

Let i (i = 1, N) index consumers, j (j = 1, J + 1) index choice alternatives, k (k = 1, K) represent product attributes, and l (l = 1, L_k) index levels within each attribute. There are J products, and the J + 1th alternative represents the no-buy option. The deterministic part of consumer i's utility for product j can be written as:

$$U_{ij} = \sum_k \beta_k X_{jk} * \delta_{ijk}. \tag{1}$$

X_{jk} is the value of attribute k for product j, β_k is the preference parameters, and δ_{ijk} takes on different values depending upon the estimated model. The deterministic part of the utility of the no-buy option is set to zero. Assuming that the deterministic utilities are augmented by an additive random component, ϵ_{ij}, which is independent, identical, extreme value Type 1 distributed, the probability that consumer i chooses product j is given by:

$$\text{Prob}(ij) = \frac{\lambda_{ij} * \exp(U_{ij})}{\sum \lambda_{ij} * [\exp(U_{ij})]} \tag{2}$$

where λij = 1 if product j is in the choice set of consumer i, 0 otherwise, which, like δ_{ijk}, is defined based upon the estimated model. Note, the λij indicator term in the denominator is outside the $\exp(U_{ij})$ expression so if product j is not included in the choice set, its utility is zero nor does it contribute to the term in the denominator. Further, because λij for the no-buy option takes the value 1 by definition, the denominator will always include the value 1, which corresponds to the exponentiated value of its zero utility.

As is common in durable product categories, we only have one observation per shopper; consequently, we are unable to allow for heterogeneity in the parameter estimates at the individual level using a Bayesian model or at the segment level using a latent class or finite mixture model. However, our approach does address the heterogeneity problem in the following way. The estimated coefficients in our proposed model can be interpreted as effects on choice conditional on accessing information on that attribute/alternative. If the information is not accessed the coefficient becomes zero. Given that respondents are unlikely to access information that they are not interested in, access becomes predictive of choice. Access separates those interested in the attribute/alternative combination from those who are not, which is a way in which consumers are heterogeneous. We estimate two versions of the proposed model.

M1A: Information Acquisition Model

In this model, δ_{ijk} = 1 if consumer i accessed the kth attribute for any product j, and is 0 otherwise. Also, λ_{ij} = 1 only if at least one attribute of alternative j was accessed, and 0 otherwise.[5] Therefore this model incorporates *all* of the information actually accessed by consumer i. If a particular attribute is not accessed by the consumer, its value does not

[5] Note that in our empirical application, λ_{ij} = 1, only if the consumer accessed an attribute level *other* than price, because the price of each choice alternative was visible to all shoppers and did not require any searching on their part. For the same reason, δ_{ijk} = 1 for the price attribute of each choice alternative.

Table 2
Model fit results.

Model	M1A: information acquisition	M1B: nested information acquisition	BM1: all information available (MNL)	BM2: attribute acquisition	BM3: alternative acquisition
Estimation sample log likelihood	−375.385	−363.156	−452.536	−435.253	−413.225
BIC	−414.905	−402.676	−492.056	−474.773	−452.745
Validation sample log likelihood	−126.704	−123.752	−146.277	−144.953	−132.910

enter the utility function of that product because $\delta_{ijk} = 0$, consistent with the fact that the consumer is uninterested in its value. Furthermore, if a consumer does not access *any* attribute information about a particular alternative, then $\lambda_{ij} = 0$ and the alternative drops out of the denominator of the logit expression, consistent with the fact that the consumer is uninterested in the alternative. The rationale for setting attribute and alternative values that are not accessed to zero is based on previous theory that consumers focus on only a subset of information as the basis for decision making (see Bettman, Luce, and Payne 1998), research on information use in stated choice experiments (Hensher 2006) and findings about how consumers use cut-offs in discrete choice models from conjoint experiments (Gilbride, Allenby, and Brazell 2006).

M1B: Nested Information Acquisition Model

This model is a nested logit version of model M1A. Thus, instead of modeling the no-buy decision as an additional alternative in the choice model, the buy and no-buy options appear together at a higher level of the choice hierarchy and the three choice alternatives are nested together under the buy option at a lower level. Thus, this model explicitly allows the correlation between the no-buy alternative and the three choice options in the buy nest to be different from the correlation between these items within the nest. While the rest of the model specification is the same as in Model 1A, an additional intercept and an inclusive value term, captured by the log of the denominator of the choice probability expression in Equation 2, are included as additional explanatory variables in the buy/no-buy decision.

We now present a set of alternative benchmark models that are based on the information available or presented to the consumer rather than on the information accessed by the consumer. Nested versions of the proposed benchmark models are not presented because they did not improve the estimation and validation sample fits or parameter estimates relative to their non-nested counterparts.

Benchmark Models

BM1: All Information Available Model (MNL)

In this model, $\delta_{ijkl} = 1$ in Equation 1 for all attribute levels, all products, and all consumers and $\lambda_{ij} = 1$ in Equation 2 for all alternatives and consumers. Thus, this model completely ignores the information accessed by the consumer by assuming that all alternatives are in a consumer's choice set and that alternative utility depends on all attribute levels, resulting in a standard MNL model based on all information available. The ubiquity of this model in marketing leads us to present and estimate it, even though we recognize that the invariant X-matrix problem and the fact that multiple attributes have similar levels across alternatives will make it impossible to identify its parameters.

Table 3
Estimation sample results.

Variable	M1A: information acquisition	M1B: nested information acquisition	BM1: all information available (MNL)	BM2: attribute acquisition	BM3: alternative acquisition
Category value	—	.230 ***	—	—	—
Intercept	−4.148 **	1.033	−.015 ***	−6.420 ***	−.014 ***
Price	−.009 ***	−.012 ***	−.013 ***	−.011 ***	−.010 ***
A1	.625 **	.626	.003	.253	−.001
A2	−.095	−1.077	−.005 ***	.339	−.006 ***
A3	1.036 **	1.686 **	−.005 ***	1.012 **	−.006 ***
A4	−.020 **	−.014	.209 ***	−.022 **	.145 ***
A5	.009	.003	.058	.005	.066
A6	.398	.998 *	−.033	.189	−.005
A7	.459	.850	−.005 ***	.375	−.006 ***
A8	−.246	−.263	−.002	−.123	.017
A9	.439	1.117 *	−.005 ***	.211	−.006 ***
A10a	.613	−.568	−.015	−2.028	−.009
A10b	− .060	.251	.005	.470 *	.016
A11	.209	1.103 **	−.002	.046	−.026

Notes.
* $p < .10$.
** $p < .05$.
*** $p < .01$.

BM2: Attribute Acquisition Model (Attribute Set Model)

In this model, $\delta_{ijk} = 1$ if consumer i accessed the kth attribute for product j and is 0 otherwise. Thus, the utility function only includes price (which is observed for all alternatives) and those terms in which information about an attribute was actually accessed. However, $\lambda_{ij} = 1$ for all alternatives j, i.e., all alternatives are assumed to be in the choice set. Consequently, this model partially accounts for the attribute level information accessed by the consumer by including the correct terms in the utility function but ignoring the contents of the choice set. Since shoppers access different attributes, unlike BM1, the X-matrix will vary across shoppers.

BM3: Alternative Acquisition Model (Choice Set Model)

In this model, $\delta_{ijk} = 1$ for all alternatives, but $\lambda_{ij} = 1$ only if at least one attribute level of alternative j (other than price) was accessed and 0 otherwise. Thus, the utility function does not take into account the specific attribute information accessed by a consumer, but does force an alternative to drop out of the denominator in the choice probability expression if none of its attribute information is accessed. In other words, in this model an alternative is included in the choice set only when a consumer accessed at least one cell of non-price information. Since shoppers access different alternatives, unlike BM1, the X-matrix will vary across shoppers.

Table 4
Parameter estimates for models M1A, BM2 and BM3 based on a priori split of the original sample.

Panel A. Segmentation based on alternatives accessed

Model/variable	M1A: information acquisition		BM2: attribute acquisition		BM3: alternative acquisition	
Segment	1 alternative accessed	2 or 3 alternatives accessed	1 alternative accessed	2 or 3 alternatives accessed	1 alternative accessed	2 or 3 alternatives accessed
Intercept	.823	.124	−7.628 **	−5.221 ***	.073 ***	−.014 ***
Price	.029	−.010 ***	−.015 ***	−.010 ***	.082 ***	−.010 ***
A1	.024	.733 **	1.021	.197	−.212 ***	.000
A2	−.369	.025	−.883	.649	−.093 ***	−.006 ***
A3	1.141	.898 *	1.944 *	.953	−.093 ***	−.006 ***
A4	−.034	−.017	−.036	−.023 *	−3.324 ***	.146 ***
A5	.030	−.009	.119 *	−.025	3.723 ***	.056
A6	1.271	.455	−.336	.223	4.002 ***	−.013
A7	1.396 *	.196	1.836 **	.114	−.093 ***	−.006 ***
A8	−.687	−.101	−.810 *	.025	3.532 ***	.009
A9	.785	.291	.891	.073	−.093 ***	−.006 ***
A10a	−13.692	.961	−1.122	−.990	.804 ***	−.011
A10b	2.531	−.182	.341	.253	2.361 ***	.011
A11	.223	.351	−.748	.137	−3.801 ***	−.018
Estimation Sample Log Likelihood	−355.080		−419.990		−388.160	
BIC	−434.119		−499.029		−467.199	
Validation sample log likelihood	−140.760		−145.160		−210.600	

Panel B. Segmentation based on attributes accessed

Model/variable	M1A: information acquisition		BM2: attribute acquisition		BM3: alternative acquisition	
Segment	<9 attributes accessed	≥9 attributes accessed	<9 attributes accessed	≥9 attributes accessed	<9 attributes accessed	≥9 attributes accessed
Intercept	−4.083 **	−.452	−7.221 ***	23.164 ***	−.013 ***	−.014 ***
Price	−.008 ***	−.007	−.013 ***	.009 *	−.009 ***	−.010 ***
A1	.437	.569	−.021	12.632 ***	−.003	.000
A2	.300	−1.223 *	.726	−1.032	−.007 ***	−.006 ***
A3	.868 *	2.603 **	1.031 **	14.583 ***	−.007 ***	−.006 ***
A4	−.028 **	−.009	−.030 **	.038	.102	.156 ***
A5	.021	.018	.016	−1.082	.100	.055
A6	.355	.485	.152	−.096	.036	−.017
A7	1.064 **	−.185	.645	−1.428	−.007 ***	−.006 ***
A8	−.940 **	.266	−.719 **	.655	.052	.006
A9	−.020	.908	−.048	−.211	−.007 ***	−.006 ***
A10a	6.344	−.753	.785	−1.955	−.001	−.012
A10b	−1.181	.159	−.124	.420	.039	.009
A11	.202	.468	.124	−.929	−.064	−.014
Estimation sample log likelihood	−365.480		−423.640		−401.080	
BIC	−444.519		−502.679		−480.119	
Validation sample log likelihood	−126.350		−162.180		−134.720	

Table 4 *(continued)*

Panel C. Segmentation based on cells accessed

Model/variable	M1A: information acquisition		BM2: attribute acquisition		BM3: alternative acquisition	
Segment	<12 cells accessed	≥12 cells accessed	<12 cells accessed	≥12 cells accessed	<12 cells accessed	≥12 cells accessed
Intercept	−3.311	−1.344	−7.749***	10.453***	−.013***	−.014***
Price	−.007***	−.010**	−.013***	−.009*	−.009***	−.010***
A1	.407	.801	.269	−.629	−.002	.000
A2	−.039	−.421	.398	.080	−.007***	−.006***
A3	1.218***	1.580	.943*	15.307***	−.007***	−.006***
A4	−.026**	.019	−.028**	.017	.125**	.147***
A5	.042	.001	.031	.041	.091	.043
A6	−.108	1.240***	−.048	.326	.022	−.023
A7	1.193**	.065	.932*	−.578	−.007***	−.006***
A8	−.587**	−.071	−.450	.473	.041	−.001
A9	.864	−.308	.813	−.937	−.007***	−.006***
A10a	5.880*	−2.084	−2.493	−2.927	−.004	−.013
A10b	−.990*	.414	.608	.563	.032	.003
A11	−.193	1.145***	−.420	.856*	−.051	−.008
Estimation sample log likelihood	−353.530	−418.530	−410.550			
BIC	−432.569	−497.569	−489.589			
Validation sample log likelihood	−132.840	−161.250	−134.300			

Notes.
* $p < .10$.
** $p < .05$.
*** $p < .01$.

Results

We randomly assigned the 582 consumers into estimation and validation samples containing 437 (75%) and 145 (25%) consumers. We coded the various attribute levels in our dataset as follows. First, for the continuous attributes, i.e., price, A1, A4 (which has the same value for all alternatives), A5, A10a, and A10b, we simply use the corresponding numerical values for each product and 0 for the no-buy alternative. Second, for each of the two-level categorical attributes, A6, A8, and A11, we employ a specification in which its lower level is coded as 1, its higher level is coded as 2, and an uninspected attribute is coded as 0. Thus, this coding scheme permits us to distinguish between the absence of an attribute (when not accessed) and its two values when present, which the typical dummy variable coding with two levels could not accomplish. For example, suppose A6 referred to the web attribute in Fig. 2. Then in our specification, the values would be 0 when the web attribute was not accessed, 1 for the experimental browser, and 2 for the Amazon Silk cloud-accelerated browser. We did not use −1/1 effect coding because it implicitly assumes that the lower level of an attribute (the −1) is below the value when the attribute is absent or was not seen. Third, for the remaining categorical attributes, A2, A3, A7, and A9, for which all alternatives have identical levels, we used a dummy variable that is set to 1 for each choice alternative and 0 when the attribute level was not accessed. As an example, if A2 referred to the Connectivity attribute in Fig. 2, then A2 = 1 when this attribute was accessed for an alternative, 0 otherwise.

We assess model performance based on (i) statistical fit and (ii) the validity of the parameter estimates. In Table 2 we present the fitted log-likelihood from each model for the estimation sample. The estimated parameters for each model were used to calculate the log-likelihood in the validation samples, which are also reported. Models that only use the information accessed by the shopper (M1A and M1B) are found to provide better estimation and validation sample fits than models that rely on all information available to the shopper (BM2 and BM3). The nested specification results in a better fit for M1B over M1A for the estimation (Chi-square = 24.46, $p < .01$) and validation samples (Chi-square = 5.90, $p < .05$).

In Table 3 we present the parameter estimates of the estimated models. Notice that many of the parameter estimates from the information available model (BM1 or MNL) have incorrect signs reflecting the expected identification problems. Because higher levels of these attributes are expected to yield higher levels of utility, the negative parameter estimates for attributes A2, A3, A7, and A9 are counter-intuitive. Thus, as expected, the all information available model (BM1 or MNL) is unable to correctly identify the impact of these attribute levels. Second, the coefficients for attributes A2, A3, A7, and A9 have counter-intuitive negative signs in the alternative acquisition model (BM3), and the coefficient of attribute A4 has a counter-intuitive negative sign in the attribute acquisition model (BM2), while coefficients for all the attributes in the nested information accessed model (M1B) yield correct signs. Note that the values of attributes A2, A3, A4, A7 and A9 in our real-world empirical data do not vary across alternatives; however, A3 is statistically significant in M1B because data on A3 can be accessed for one option but not others and this model explicitly accounts for the information accessed.

We also investigated whether incorporating parameter heterogeneity in the benchmark models BM2 and BM3 would improve model fit and resolve the problem of counter-intuitive signs observed in the corresponding aggregate version reported

in Table 3. Because we only have one observation per consumer, typical in infrequently purchased product categories, and no demographic information, we allow for heterogeneity using an a priori segmentation approach (e.g., Currim 1981) based on information access patterns. Specifically, we used three segmentation schemes based on the number of (a) alternatives accessed, (b) attributes accessed, and (c) cells accessed. For each of the three segmentation schemes we defined two segments and chose cut-off points to ensure a reasonable sample size in the estimation and validation sub-samples. Specifically, the estimation (validation) samples are (a) 175 (67) shoppers accessed information on only 1 alternative and 262 (78) accessed information on 2 or 3 alternatives; (b) 284 (87) shoppers accessed 8 or fewer attributes and 153 (58) accessed 9 or more attributes; and (c) 261 (95) shoppers accessed 11 or fewer cells and 176 (50) accessed 12 or more cells.

The fit statistics and parameter estimates for these models are presented in Table 4. However, incorporating segment level heterogeneity in the information access model (M1A) and the benchmark models (BM2 and BM3) did not improve the estimation and validation sample fits nor resolve the problem of the counter-intuitive sign observed in the corresponding homogenous version reported in Table 3. Therefore, we conclude that incorporating heterogeneity in the parameter estimates using these a-priori approaches does not improve the validation sample fits or resolve the issue of incorrectly signed parameter estimates observed in the aggregate versions of the baseline models. Only M1B which incorporates all alternative and attribute information accessed by shoppers has all correctly signed parameter estimates and the best validation sample fits.[6]

In summary, because the information accessed models incorporate the cells of the comparison matrix actually opened by a consumer, they can overcome estimation issues such as an X-matrix that does not vary across shoppers and attribute values that are invariant across all alternatives. In contrast, the benchmark models that rely on all information available or incorporate only certain aspects of information accessed in their estimation, have identification problems and/or counterintuitive signs. Thus, our empirical analysis shows that employing the information accessed, with and without a priori segmentation, improves validation sample based forecasts of consumers' choices and provides more valid estimates of consumers' attribute preferences over employing information available.

Additional Models

To establish robustness of our findings and to account for specific aspects of our data setting, we estimate two other models that we term (i) inferred attributes and (ii) endogenous information accessed.

Inferred Attributes

In this model we set the value of an attribute that is not accessed to be the maximum (minimum) value of the attribute in the data matrix. The theoretical rationale is that consumers may have inferred the value of an attribute even if they did not access information on it, either because the consumer believes that the values of all attributes in the row are similar or because once they click on an attribute they infer the remaining attribute values in the row. Consequently, $\delta_{ijk} = 1$ for all i, j, and k, but the value of X_{jk} is set to the minimum (maximum) (X_{jk}) if the attribute level was actually not accessed and $\lambda_{ij} = 1$. For attributes that only have a single value, that specific value is used. In other words, this model assumes that all alternatives and attributes are accessed, but the value of each particular attribute level is either its actual value if accessed or the maximum (minimum) value if not accessed. Thus, this model controls for the amount of information accessed per alternative and reduces the penalty for attributes that are not accessed.

The fit and parameter estimates for the Inferred Attributes model are also found to be worse than M1B.[7] A potential explanation for why there is no impact of modeling the attribute as zero versus a positive value in the Inferred Attributes model is as follows. Attributes A2, A3, A4, A7, and A9 do not vary across the products and, consequently, whether we assume the value is 0 or the maximum (or minimum) for information not accessed (incidentally the amount of information not accessed is high, see Table 1 Panels B–E), the values for all alternatives on these five remain the same. In addition, many shoppers purchase the lower priced alternative (alternative 1), because they are price sensitive, i.e., they are less concerned about higher priced alternatives (alternatives 2 and 3) and their attribute levels other than price. If such attribute levels are not accessed for higher priced alternatives (see Table 1 Panels D and F) whether one assumes a value of 0 or the maximum (or minimum) value for the higher priced alternatives it does not matter because the higher priced alternative is not chosen. Additionally, these results further support the idea that tracking the information accessed is critically important to getting useful parameter estimates.

Endogenous Information Access

It is possible that consumers access more cells of alternatives for which they have a greater preference making cell clicks depend upon the previously uncovered information about an alternative. In other words, the attribute values that enter the choice model also depend upon preference and not modeling consumer information access jointly with the final choice decision can create an endogeneity problem that can bias the parameter estimates.

To explore this possibility, we extended model M1B to take into account a person's entire clicking history in addition to the final choice decision. Let n represent the total number of cells opened by the consumer at a particular point in the process, j

[6] We performed a simulation to test whether M1B is able to recover the underlying choice parameters. Specifically, we used a known vector of utility weight parameters (β_k) to simulate a single choice for every consumer in the sample based on the actual search pattern used by that consumer. We then estimated the corresponding model on this set of simulated choice data, and repeated the data generation and estimation process 200 times to confirm that we were able to recover the β_k that had generated the choices.

[7] In addition, we tested a model wherein the value of attributes not accessed was set to the average value of the attribute. The resulting log-likelihood of the model, while better than the benchmark models, was inferior to M1A and M1B.

and k represent the column and row for an unopened cell C_{jk}, and let O^n_{jk} be the propensity to open this cell. The utility in Equation 1, U^n_{jk}, is re-specified to be updated upon every click, resulting in a dynamic choice probability $Prob(ij)^n$.

Cell opening endogeneity is captured by making the opening propensity of an unopened cell C_{jk} a function of the dynamic choice probability $Prob(ij)^n$ for that alternative, following the previous behavioral literature on alternative-based processing, and the cumulative number of cells opened in that row until that point, num^n_k, following the previous behavioral literature on attribute-based processing (e.g., see Bettman, Luce, and Payne 1998). In this way, the explicit link between cell opening and product preference or alternative-based processing is captured by the first term while the analogous effect in the horizontal direction or attribute-based processing is captured by the second term. Adding an iid extreme value Type 1 Gumbel error term to each propensity yields a multinomial logit expression for the probability of opening a cell that has not yet been opened.

By way of example, consider a person who has opened ten cells (n = 10) at a particular point. The uncovered attribute levels yield updated utilities and probabilities for each of the alternative $Prob(ij)^{10}$ while the number of cells opened in each row k is available in num^{10}_k. These two values define the cell opening propensities for each one of the 23 remaining cells, O^{10}_{jk} and, in turn, the corresponding cell-opening probabilities. The parameters of the cell opening model are jointly estimated with the preference parameters of the choice model yielding the endogenous version of M1B.[8]

The parameter values for the coefficients of $Prob(ij)^n$ and num^n_k were positive and statistically significant, indicating that consumers are more likely to open cells for products that they like and in rows that they have already opened. Interestingly, the coefficients of the endogenous model, which also incorporates the information accessed by the consumer, also had the right signs underscoring our basic finding that modeling information accessed is better than simply modeling information available. We used the preference coefficients from M1B (endogenous) to calculate the choice likelihood in the calibration and holdout samples but found them to be inferior to those from M1B (exogenous). In summary, M1B (exogenous) based on information accessed by consumers is found to outperform three benchmark models based on the current literature and two additional models based on inferred attributes and endogenous information accessed.

A potential explanation for why the original model M1B (exogenous information accessed) does better than M1B (endogenous information accessed) is as follows. The endogenous model needs to fulfill two tasks, (i) predict product choices and (ii) fit the cell opening model. In contrast, the parameters of the exogenous model need only accomplish the sole goal of producing a product choice model with the best estimation sample fit. Because two tasks are more difficult to accomplish than one, consequently, M1B (exogenous information accessed) results in better estimation and validation sample fits than M1B (endogenous information accessed).

Discussion

Consumers typically do not access all information at the point-of-purchase due to search costs, information overload, prior knowledge, or heuristic shopping (Bettman, Luce, and Payne 1998; Diehl and Zauberman 2005; Jacoby, Chestnut, and Fisher 1978). Even though previous information processing research has for over forty years recorded the information accessed by consumers in lab settings, this data has not been employed in choice models to infer consumer preferences or predict consumer choices. In contrast, the choice-based conjoint and consider-then choose models in marketing, also for over forty years, have employed the information available on all attributes and alternatives to estimate the parameters of a multi-attribute choice model, while ignoring the information accessed by consumers.

We combine important elements of these established research streams by (a) collecting the information consumers' access in an online durable product retail setting and (b) investigating whether incorporating information accessed in such a setting provides more valid diagnostics about attribute preferences and improves model estimation and validation sample fits over those achieved based on information available. We accomplish this by focusing on consumers' choices and what information is accessed, and by making small changes to a manufacturer/retailer's website to obtain information on product-attributes accessed and revealed preferences that are often only obtained through laboratory experiments.

Our database recorded what attribute information was accessed by 582 shoppers at the point-of-purchase of an electronic manufacturer/retailer's website and found that 48% of shoppers did not access information on all attributes, 49% did not access information on all alternatives, and only 9% of consumers accessed all feature information available. Most importantly we show that monitoring, recording, and employing the information accessed by the shopper can improve validation sample based forecasts of consumers' choices and provide more valid estimates of consumers' attribute preferences.

Our research provides a practical model based methodology for managers. It demonstrates the potential for managers to incorporate these relatively inexpensive and straightforward efforts that utilize the authentic shopping behavior of a sample of customers on their retail websites to improve the diagnostics about attribute preferences and predictive accuracy of their marketing models. We show in our data analysis that employing choice models with information available instead of information accessed would lead to harmful recommendations for product design and advertising decisions for attributes A2, A3, A7, and A9, which potentially would decrease product sales. Hence, incorporating information accessed is a first step in focusing on what people are actually doing at the point-of-purchase on the web, subsequent to which diagnostics about attribute preference

[8] We also tested several different specifications for the cell opening model using, different combinations of num_k, $(num_k)^2$, $Prob(ij)^n$, $(Prob(ij)^n)^2$, $\log[Prob(ij)^n]$, and $(\log[Prob(ij)^n])^2$.

from choice models utilizing this information should be employed to improve product design, pricing, advertising, and targeting decisions. For example, Mozilla recently tracked which buttons consumers clicked most often on their Firefox toolbar browser to help redesign their Firefox web browser, Microsoft Bing used eye-tracking to analyze the percentage of consumers who look at information beyond the 5th search result in an effort to overhaul their search results display, and *The New York Times* employed mouse-tracking to investigate consumers viewing patterns on their webpage to assist in their redesign of their webpage.

In addition, managers can use information accessed by shoppers to enhance their follow-up communications with shoppers who did not purchase. For example, if managers collect information on which products and attributes were accessed by non-buyers, this information could be employed in follow-up communications by focusing non-buyers on certain alternatives and attributes they accessed and hence demonstrated an interest in. A request to get the non-buyer to just reconsider their decision, or combining the request with an incentive to buy, e.g., a temporary price promotion, could result in a decision to buy, especially for those non-buyers who considered the lowest price alternative on at least one visit to the manufacturer/retailer website.

The managerial and methodological implications of our findings can also be extended to offline brick and mortar stores. In traditional stores such as Best Buy, often not all information on the product is available on the shelf talker and not all brands and models in a particular category will contain information on the same attributes so that the kind of assessments implied in current choice-based conjoint and consider-then-choose models could be difficult for consumers to accomplish. The consumers may also find all the information too burdensome to process and may sample information selectively and not access all the information on attributes and alternatives as assumed in current state-of-the art choice models. In addition, if the same X-matrix of alternatives and attributes is available to all consumers and if attribute levels are the same across brands and models, standard multinomial logit choice models may produce non-identifiable or counter-intuitive parameter estimates. Thus, our work offers a method for managers to obtain diagnostics and parameter estimates based on the information consumers actually accessed at the point-of-purchase to help overcome such difficulties.

A limitation of our research is that in our shopping setting the electronic manufacturer/retailer required that we present three specific existing products to shoppers only for a short 50-hour time period. In other words, we could not employ hypothetical products or show shoppers different sets of existing products to make choices from as is commonly done in CBC studies. Consequently, we are unable to study preferences for new attribute levels or study heterogeneity across consumers on attribute preferences as is possible in CBC studies.

However, we are able to observe the information on attributes and products actually accessed by the shopper at the point-of-purchase before making a Customize and Buy decision, and as a result, are able to investigate the relevance of information accessed, relative to information available at the point-of-purchase, in improving inferences from choice models. One direction for future research is to investigate whether data collected using actual products at the point-of-purchase, as we do in this paper, can be combined with data collected using hypothetical product profiles in CBC studies to improve the scope and performance of consumer models. This could allow us to harness the advantage of capturing information accessed in point-of-purchase settings with the advantages of traditional CBC studies.

In addition, while randomizing the sequence in which attributes are presented is impossible in a field study (relative to a controlled lab study), any non-random presentation of attributes is common across models and will not benefit one model over another. Another direction for future research is to further enhance the endogeneity model by making the cell opening probability an explicit function of the dynamic search process and incorporating the decision to stop searching into the model. We hope such future research will build on our efforts.

References

Andrews, Rick L. and T.C. Srinivasan (1995), "Studying Consideration Effects in Empirical Choice Models Using Scanner Panel Data," *Journal of Marketing Research*, 32, 1, 30–41.

Bettman, James R. (1979), *An Information Processing Theory of Consumer Choice*. Reading, Massachusetts: Addison-Wesley Publishing Company.

———, Mary Frances Luce, and John W. Payne (1998), "Constructive Consumer Choice Processes," *Journal of Consumer Research*, 25, 3, 187–217.

Currim, Imran S. (1981), "Using Segmentation Approaches for Better Prediction and Understanding from Consumer Mode Choice Models," *Journal of Marketing Research*, 18, 3, 301–9.

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.

Diehl, Kristin and Gal Zauberman (2005), "Searching Ordered Sets: Evaluations from Sequences Under Search," *Journal of Consumer Research*, 31, 4, 824–32.

Gilbride, Timothy J. and Greg M. Allenby (2004), "A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules," *Marketing Science*, 23, 3, 391–406.

———, ———, and Jeff D. Brazell (2006), "Models for Heterogeneous Variable Selection," *Journal of Marketing Research*, 43, 3, 420–30.

Hensher, David A. (2006), "How Do Respondents Process Stated Choice Experiments? Attribute Consideration Under Varying Information Load," *Journal of Applied Econometrics*, 21, 6, 861–78.

Jacoby, Jacob, Robert W. Chestnut, and William A. Fisher (1978), "A Behavioral Process Approach to Information Acquisition in Nondurable Purchasing," *Journal of Marketing Research*, 15, 4, 532–44.

———, ———, ———, James J. Jaccard, Imran Currim, Alfred Kuss, Asim Ansari, and Tracy Troutman (1994), "Tracing the Impact of Item-by-item Information Accessing on Uncertainty Reduction," *Journal of Consumer Research*, 21, 2, 291–303.

Johnson, Eric J., John W. Payne, and James R. Bettman (1988), "Information Displays and Preference Reversals," *Organizational Behavior & Human Decision Processes*, 42, 1, 1.

———, 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.

Lancaster, Kelvin J. (1966), "A New Approach to Consumer Theory," *Journal of Political Economy*, 74, 2, 132–57.

Lurie, Nicholas H. (2004), "Decision Making in Information-rich Environments: The Role of Information Structure," *Journal of Consumer Research*, 30, 4, 473–86.

Malhotra, Naresh K., Arun K. Jain, and Stephen W. Lagakos (1982), "The Information Overload Controversy: An Alternative Viewpoint," *Journal of Marketing*, 46, 2, 27–37.

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.

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.

Musalem, Andrés, Marcelo Olivares, Eric T. Bradlow, Christian Terwiesch, and Daniel Corsten (2010), "Structural Estimation of the Effect of Out-of-stocks," *Management Science*, 56, 7, 1180–97.

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.

Ratchford, Brian T. (1975), "The New Economic Theory of Consumer Behavior: An Interpretive Essay," *Journal of Consumer Research*, 2, 2, 65–75.

Siddarth, S., Randolph E. Bucklin, and Donald G. Morrison (1995), "Making the Cut: Modeling and Analyzing Choice Set Restriction in Scanner Panel Data," *Journal of Marketing Research*, 32, 3, 255–66.

Simon, Herbert A. (1956), "Rational Choice and the Structure of the Environment," *Psychological Review*, 63, 2, 129–38.

Wu, Jianan and Arvind Rangaswamy (2003), "A Fuzzy Set Model of Search and Consideration with an Application to an Online Market," *Marketing Science*, 22, 3, 411–34.