View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by The University of Utah: J. Willard Marriott Digital Library

UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

Working Paper No. RHS-06-141

November 15, 2010

Clicks to Conversion: The Impact of Product Information and Price Incentives

Vandana Ramachandran

University of Utah - Department of Operations and Information Systems

Siva Viswanathan

University of Maryland - Robert H. Smith School of Business

Henry C. Lucas

University of Maryland - Robert H. Smith School of Business

This Paper can be downloaded without charge from the
Social Science Research Network Electronic Paper Collection
http://ssrn.com/abstract=1717047

Clicks to Conversion: The Value of Product Information and Price Incentives

Vandana Ramachandran
David Eccles School of Business
University of Utah, Salt Lake City, UT 84112
vandana@business.utah.edu

Siva Viswanathan
Robert H Smith School of Business
University of Maryland, College park, MD 20742
sviswana@rhsmith.umd.edu

Hank Lucas
Robert H Smith School of Business
University of Maryland, College park, MD 20742
hlucas@rhsmith.umd.edu

Abstract

This study uses clickstream data obtained from a large online durable goods retailer to examine how different types of information – product-related and price-related information provided by retailers – impact purchase-related outcomes for consumers. Using mixture-modeling techniques to analyze latent differences among customers, we find that consumers fall under three distinct categories - *directed shoppers, deliberating researchers* and *browsers*. In examining the impacts of information on purchase outcomes, we find that product and price-related information impacts consumers in these three shopping states differently. While product information highlighting features of product alternatives in a category has the strongest impact on *deliberating researchers*, specific price incentives related to category-level discounts increases the likelihood of purchase for both *directed shoppers* as well as *browsers*. Price incentives relating to site-wide free shipping have a positive impact on purchase for all consumers. Surprisingly, category-level discounts have a negative impact on *deliberating researchers*, while rich product information hampers the purchase process of *directed shoppers*. We discuss the managerial implications of our findings and the role of clickstream analytics in designing dynamic targeting and information provisioning strategies for online retailers.

Keywords: online shopping, product information, price incentives, states of shopping

Electronic copy available at: http://ssrn.com/abstract=1717047

INTRODUCTION

Providing shoppers with the right information at the right time can have a significant impact on individual purchase behaviors and consequently a retailer's bottom line. However, given that shoppers differ not only in their shopping goals but also their purchase propensities (Moe 2003; Wolfinbarger and Gilly 2001), understanding what constitutes the "right information" and the "right time" for each type of consumer becomes paramount. This concern is especially important for online retailers of infrequently purchased durable goods. Our study uses detailed micro-level data to examine the impacts of two specific categories of information –related to *products* and *price incentives* - on the purchase outcomes of online customers. While past research has found that product and price information have significant impacts on consumers' choices and outcomes (e.g., Diehl et al. 2003; Klein and Ford 2003; Lynch and Ariely 2000), they treat consumers as homogeneous. We specifically seek to understand whether the influence of product information and price incentives differ systematically across consumers with varying shopping-related needs.

Traditionally marketers have sought to differentiate consumers on the basis of geography, demographics, psychographics, and purchase history including recency, frequency, and value of purchase (e.g., Rossi et al. 1996). However, similar information is scarce for online retailers that are faced with visits from relatively "unidentifiable" visitors who form a significantly higher proportion of traffic than "loyal" or "registered" customers. This difficulty is more pronounced for online retailers of infrequently purchased durables, who, faced with a slim dossier on each customer, must seek alternate ways to learn about the customer's needs and preferences.

The interactive web channel combined with recent technological advances has placed large amounts of rich micro-level clickstream data in the hands of retailers – data that can provide valuable insights into consumers' online behaviors. Prior literature has found that

Electronic copy available at: http://ssrn.com/abstract=1717047

consumers' online navigation behaviors can be used to classify sessions into groups that differ in their shopping-related goals (e.g., Moe 2003; Hoffman and Novak 1996). Consumers shopping for durable goods often have differing goals – some are close to finalizing the purchase, while others are browsing through a product category, and still others may be researching the available product features and forming their consideration sets. We describe the latent differences that drive these variations in consumer behavior as *states of shopping.* In this paper, we theorize that consumers in different shopping states not only have varying shopping goals, but also correspondingly varying information needs – and will therefore be influenced differently by product information and price incentives. We seek to determine *when* product-related information and price incentives increase a consumer's propensity to buy, and when they lead them to abandon their decision to purchase.

We use finite mixture modeling techniques to discover an optimal set of states that are distinguished in their navigation patterns and responses to category level product information, category discounts and shipping offers. Using clickstream data that covers visits from over 36,000 customers to four best-selling durable products carried by a leading click-and-mortar retailer in the U.S. market in late 2006, we find that a three-state model consisting of *directed shoppers, deliberating researchers* and *browsers* best describes the latent shopping-relevant differences across sessions in our data.

Our main findings pertain to purchase conversion *within a session.* We find that product information that highlighted features of product alternatives in a category has the strongest impact on *deliberating researchers*, while price discounts related to a specific product category (e.g., 10% off refrigerators) has a significant impact on both *directed shoppers* and *browsers*. Price incentives related to site-wide free shipping interestingly have a positive impact across all

Electronic copy available at: http://ssrn.com/abstract=1717047

three states, highlighting the value placed on free shipping by online shoppers. More surprising are the two negative effects. Category-level discounts have a negative impact on *deliberating researchers*, while product information hampers the purchase process of *directed shoppers*. One explanation that our theorizing provides is that match (mismatch) between information and the goals of the consumer in a particular state of shopping can aid them in narrowing down (expanding) the choices and by increasing the attractiveness of the focal (non-focal) alternatives.

Our secondary results that compare the effects of information/incentives on within-session and across-session outcomes highlight an important tradeoff. Whereas price-related incentives positively influence within-session conversion for a greater number of sessions, they were less effective in attracting consumers to repeat visits when they left without purchasing. In contrast, product-related information positively influences a smaller set of deliberating customers to convert within a session, but has a stronger impact on across-session outcomes, influencing consumers in all three states to purchase in later sessions. Our findings are robust to a number of different considerations – using a subset of cookies/sessions that display higher engagement with or interest in the focal products, controlling for potential endogeneity in product information using a matching estimator, controlling for price/product sensitivity of users, alternate specifications of the purchase, and using the cart as an outcome.

Our study makes a number of important contributions. It is among the first to use a rich clickstream dataset to show the varying impacts of product information and price incentives on consumers who differ in their latent shopping needs. Second, we theorize and demonstrate that the results observed in our data are tied to the differences in shopping-relevant information needs across the states. Third, the study takes into account both session-level and cookie-level heterogeneity in examining consumers' purchase decisions. Third, by studying the online purchase incidence of durable goods, our work adds to the existing literature that primarily focuses on targeting messages to consumers purchasing frequently purchased products. Findings from our study can help firms develop a targeted information provision strategy that operates contemporaneously as consumers' shopping intentions are ascertained in real-time. Our micro-level approach has the potential to be highly interactive and complementary to existing targeting strategies used to attract consumers to e-tailer stores. Our work, thus, contributes to the burgeoning interest in business intelligence by developing a clickstream analytic technique to help learn important relationships between purchase, product information, price incentives, and unobserved latent states of consumers using a large scale micro dataset of visits to a retailer.

The rest of the paper is structured as follows. We begin by surveying existing research and discussing our conceptual framework. We then describe the clickstream data and develop models for uncovering latent states of shopping, followed by cookie-panel models to study the effects of product information and price incentives. We then discuss the robustness of our findings, and conclude with a discussion of the implications for retailers.

RELATED RESEARCH

Two main streams of literature inform our work. In the first stream, a large body of existing work spanning computer science, information systems and marketing has studied consumer behaviors in online channels, and broadly suggests that search (paths, patterns, and volume) can predict outcomes. Early work studied users' traversals on the web and classified their navigation strategies to determine interesting patterns (e.g., Catledge and Pitkow 1995; Yang and Padmanabhan 2007). This research has since been extended to use these patterns to predict outcomes (e.g., Srivastava et al. 2000). More recent works by IS researchers have attempted to infer latent user intentions and contextual factors that give rise to observed search behaviors to better understand their drivers (Adomavicius et al. 2005; Jin et al. 2004). The resulting models have been used to implement better document pre-fetching systems, and recommendation and adaptive personalization systems in online environments. Also, fueled by the explosion of online search, several recent studies have found aggregate measures such as the volume of web searches to accurately predict future consumption outcomes for films, games and songs (Goel et al. 2010).

As clickstream technologies have evolved, researchers have turned their focus to the modeling of finer-grained e-commerce data gleaned from consumers' online trails. Some prior work has used session-level metrics (e.g., time spent and pages visited) to predict the likelihood of purchase conversion. A few studies have examined paths taken by consumers across websites (e.g., Johnson et al. 2004; Park and Fader 2004); while others - like ours – have focused on search within a website (e.g., Bucklin and Sismeiro 2003; Montgomery et al. 2004; Moe and Fader 2004; Sismeiro and Bucklin 2004). Among the latter set of studies, some have examined search within a session (e.g., Moe 2003; Sismeiro and Bucklin 2004) and others have modeled sessions over time (e.g., Moe and Fader 2004). However, majority of the studies have either lacked access to, or have not modeled the effects of the different types of content (or information) viewed by consumers - which is likely to influence their purchase behaviors.

A second stream has examined the impact of price and product information found online on consumers' purchase outcomes (e.g., Ratchford et al. 2003; Viswanathan et al. 2007; Zettelmeyer et al. 2005). However, prior work typically does not distinguish their impacts across consumers with different shopping goals, treating them as homogeneous.

Our research intersects and extends these streams to understand how search patterns in clickstream data can be used to discern underlying differences in consumers' shopping goals that are driven by theories of consumer behavior and are not merely interesting patterns, and how the states thus identified respond to rich product information versus price incentives. Our work is closest in spirit to research that studies consumer responses to marketing communications and prescribe strategies to optimally target messages to customers. Zhang and Krishnamurthi (2004) study the related questions of when-how much-and to whom to promote to in an online market for frequently purchased products on the basis of purchase history. Manchanda et al. (2006) study the effect of banner advertising on purchasing behavior for consumers who bought at least once. While their study uses clickstream data, they do not observe the content of ads - and therefore cannot distinguish the impacts of price and product information.

CONCEPTUAL FRAMEWORK

States of Shopping

Not all shoppers shop with the same intent or goals and these underlying differences are often reflected in consumers' search and decision-related behaviors in offline channels (Cox 1967; Putsis and Srinivasan 1994). When consumers adopt the Internet to conduct research and make purchases, such variances are likely to translate to the online market as well (e.g., Moe 2003). Clickstream data, in particular, are composed of navigation trails from diverse customers who may have varying purchase goals (Bucklin and Sismeiro 2003; Moe and Fader 2004). To the extent that consumers with differentiated shopping needs may value product information and price differently, treating consumers as homogenous would be naive. The limited ability to track consumers' offline behaviors meant that earlier models relied on summary behaviors, whereas the online channel offers nuanced insight into detailed individual search behaviors.

A commonly drawn distinction differentiates two extremes of consumer navigation behaviors in the online channel – exploratory vs. goal-oriented (Hoffman and Novak 1996). Exploratory browsing is often undirected, less-deliberate and stimulus-driven (Janiszewski

1998). This type of search, as found in prior literature, may not necessarily be motivated by a specific goal. Here, consumers derive utility not from the outcomes of search, rather, but from the process of gathering knowledge that may be useful in the future. Such exploratory behaviors are often part of a consumer's ongoing search process (Hoffman and Novak 1996; Wolfinbarger and Gilly 2001). By contrast, goal-oriented searchers are purposeful and obtain utility by clicking and traversing through paths that allow them to gather information related to a specific goal – a product of interest or an impending purchase (Childers et al. 2001; Hoffman and Novak 1996). Moe (2003) showed that both exploratory and goal-oriented consumers may purchase.

More importantly, consumers in different clusters displayed significantly different navigation patterns during online sessions at retailer stores (Moe 2003). We surveyed past work on categorizing online sessions based on consumer navigation and found that a combination of measures related to the *breadth*, *depth* and *intensity* of search within a session is helpful in differentiating sessions (e.g., Hoffman and Novak 1996; Moe 2003; Wolfinbarger and Gilly 2001). These studies suggest that variations in consumers' observable navigation behaviors are likely driven by the differences in their shopping-related needs. We therefore use a model-based approach to derive groupings of sessions characterized by their varying navigation patterns, while simultaneously seeking to differentiate the impacts of information/incentives across the states. The exact number of states however is context-dependent and is empirically determined.

Online Information and Incentives

Our interest lies in examining the impacts of information and incentives that is provided by and controlled by online retailers (as opposed to those offered by manufacturers such as brand-specific discounts), and which apply to a subset of products available at the online retailer (such as products in a category). *Product information* provides consumers with greater knowledge related to the capabilities, features, uses and applications of the products in a product category, thereby allowing consumers to better "experience" products (Lucas 2001). In online markets, such non-price information may include the use of multimedia and microsites to provide product configurators, buying and comparison guides, and video/audio demonstrations of features that enable consumers to compare across and evaluate alternatives. This knowledge may help consumers increase their utility for products in that category by allowing them to find the best suited alternatives. *Category-specific price* incentives offer consumers discounts to purchase products from select categories (such as "savings on home furnishing-10% off", "End of season special values on all kitchen appliances"). *Generic price* incentives include a price reduction or discount that may be applied to purchases at the firm's website and not specific to any one particular category, such as shipping and delivery offers (e.g.,"free shipping on orders over $X").

DATA

Given our interest in purchase incidence rather than brand choice, we fix the products of interest. We chose four best-selling products (henceforth referred to as *focal products*) at a leading click-and-mortar retailer of durable goods that vary in category level information and incentives available. This ensures that our sample contains reasonably healthy conversion rates and purchase is not an extreme event, which may complicate the identification of effects in our models[1]. Further, this design helps control for the effects of product level attributes and price (that may drive purchase) since they do not vary in our setting within each focal product. We obtained an extensive dataset that includes all searches conducted by consumers who visited the retailer's website and clicked on one of the focal products during a contiguous 30-day period in late 2006. The retailer did not make any significant changes to the layout or organization of the website during this time period that could contaminate our results.

[1] Due to restrictions on the data that the retailer was able to provide, we limited our data to four best-sellers.

Each recorded consumer session consists of an ordered and time-stamped sequence of clicks to the online store pages. The clickstream data is a rich source of information about consumers' activities at a website, including the pages visited and the use of search tools and decision-aids to refine/screen products. The clicks were matched with corresponding site pages that we downloaded during the time of data collection to ascertain the type of content viewed including category, product, and information pages, promotions, customer service, catalogs etc.

The data was collected by the retailer with the aid of a third party tool that uses Javascript technology to track user movements across website pages. It also reads and records cookies and session level identification. This type of clickstream data offers some benefits over webserver log data. Unlike the latter where each page request by a user generates several server hits (from graphics, multimedia, and content on the page) that have to be aggregated to correspond to a meaningful user page request, each page view in our clickstream corresponds to a single URL making it much cleaner and more complete. However, clickstream data requires extensive pre-processing before it can be analyzed. We filtered the text using custom-built parsers written in PERL, which encode the text into numeric form amenable to quantitative analysis.

Sample Construction

The retailer provided us with 86,321 sessions, identified by a unique combination of cookie ID and session ID. This dataset however had to be extensively cleaned. We eliminated sessions that included only one page view, and also removed sessions where no focal product pages were viewed. An important limitation of clickstream data is that we cannot determine with certainty what product was purchased if we do not observe the product that the consumers clicked on to add to the shopping cart. We therefore limit our examination to sessions where the consumer clicked on a focal product to view it. We also ensured that when consumers did not purchase one of the focal products in a session, they also did not purchase any other product; and that a focal product was visited in all included sessions. This cleaning resulted in 43,041 sessions[2].

Finally, to reduce the chance of capturing only the repeat sessions of visitors who might have made their first visit in the days preceding our data collection, we dropped single sessions in the first two days of our time period. This choice is supported by a study of over 150 million online transactions across 800 retailers that found that when shoppers left an online store due to concerns about security, brand trust, and the need to price-compare, nearly 80% of those who return did so within 1-2 days (McAfee 2010). Our final sample consists of a total of 40,740 sessions from 36,636 unique users (cookies). The total number of sessions that are repeat visits is 4,102 resulting in 7,104 total sessions (17.44%) from 3,002 repeat visitors. In the session level sample, the visit to cart ratio is 9.31%, the visit to buy ratio is 2.06%, whereas the conditional cart to buy ratio is 22.12%. At the cookie level 2.30% of consumers make a purchase.

Measures

Outcome: $Purchase$ is a binary variable that indicates whether a consumer completed the purchase. As a first step, we track whether consumers added a product to the shopping cart (measured using a binary variable $Cart$). Following this, consumers complete a number of steps in order to purchase. We use this to construct a count measure ($Purchase\_cnt$) that is closely related and highly correlated to $Purchase$ – the number of purchase related steps completed by a user during a session, with a higher count indicating greater likelihood of completing the process.

Information and Incentives: $ProdInfo$ measures whether consumers, during the online session, clicked on multimedia/rich media content that offer information on product features,

[2] These errors are beyond our control, and occur in the clickstream data generation process at the retailer end. We however found that they have significantly fewer clicks (mean = 3.44, s.d.= 2.18) and no conversion. What remains is a clean sample of customers that are randomly selected from the population of all users that visit the focal product pages. Visits occurring greater than 30 days apart have a high probability of being for different products – hence our choice of a contiguous time period, where customers are likely to repeat visit to view the same focal products.

ideas for their usage or applications, and guides for buying and product comparisons. As an example, in the case of appliances, the online retailer offered how-to documents and configurators to help users compare features such as capacity, volume, material, energy efficiency, style and finish among others, and a listing of "points to consider" when selecting among alternatives. All four product categories offered multiple kinds of product information. Specific price incentive ($SPriceInc$) refers to promotions available in specific product categories. There were category-specific sales for three product categories. Generic price incentive ($GPriceInc$) is a free shipping offer available store-wide for 16 days during our data collection period. These three types of information had differing patterns of availability at the store, thereby allowing us to separately identify the effects of each. Further, the retailer did not target any price and non-price information to customers. A binary variable indicates whether consumers obtained each type of information - $ProdInfo, GPriceInc, SPriceInc$.

States of Shopping: Individual sessions form the basis for categorizing consumers' state of shopping, which as theorized, can change across sessions (over time) for a given consumer[3]. However since the state is actually latent, we infer it from observed navigation patterns of consumers. We borrow from past work in using the breadth, depth, and intensity of search to differentiate between directed vs. browsing behaviors (Moe 2003; Wolfinbarger and Gilly 2001).

Breadth of search is defined using two measures - the number of unique product departments viewed($DeptBreadth$), and the number of unique product categories viewed ($CatBreadth$)[4]. The first refers to search across departments or unrelated product categories, e.g., garden, appliances, and flooring. The second refers to search within a department and across related product categories- e.g., hardwood, tile and laminate within flooring. Depth of search

[3] For example, a customer may begin by browsing a product category, then in another session transition to a state where she may more carefully form her consideration set, and decide either to not buy or buy a considered product.
[4] The online store is broadly organized as departments (Appliances)→ categories (refrigerators)→ products.

($Depth$) is the extent of hierarchical drilldown (or narrowing of results) within the product category of the focal product conducted by the customer. This measure is normalized since the four product categories allow for a different maximum depth by design of the category.

Intensity of search is measured using a set of variables that describe the level of involvement of the shopper in a given session. We calculate the total time spent in minutes ($TotalTime$)[5], the number of pages visited ($TotalPages$) and the number of unique product pages viewed ($TotalProducts$) by the customer. A product page is counted as viewed only when the customer clicks on a particular product to view its details. Additionally, we create two related ratios –the number of product pages accessed per minute ($ProdPagesPerMin$) with a lower number indicating that the consumer is more engaged with (reading and processing rather than skimming) the content, and the ratio of product pages to the number of categories visited during the session ($RatioProdtoCatPages$) where a larger number would indicate either that the customer was focused on a few categories and/or that she viewed many product pages. All variables are described in Table 1. Sample means and correlations are provided in Table 2.

MODELING SESSION BEHAVIORS

We are interested in determining latent states of shopping where consumers vary in their shopping related needs, and consequently exhibit different observable navigation behaviors. More importantly, we expect that differences in consumers' needs in different states lead to varied influences from product information and price incentives. Our goal is to obtain groupings of sessions that are both responsive and identifiable.

[5] The time spent on the last page is not known (Bucklin and Sismeiro 2003; Montgomery et al. 2001). So we use the second-last click to determine time based measures. While this is a limitation, it is likely that the time spent on the last clicked page is the longest for directed shoppers. If this information were available it would only strengthen the existing differences across the three states (directed shoppers have the highest mean for *TotalTime*) – see Table 4a.

The finite mixture technique offers a model-based approach to simultaneously estimate both session membership in latent states and their response parameters to improve both the identification of states and model fit. Segmenting consumer sessions on only one of the bases – descriptive characteristics (here navigation) or response to marketing interventions – will not allow the firm to target and reach the desired segments effectively. Clustering based on dual criteria is advantageous when there is some concordance between segments derived from different bases. Theory suggests that consumers in different sessions have different underlying shopping goals, and exhibit varying navigation patterns ranging from directed to exploratory. In turn, these varying shopping-related needs are expected to consequently drive different valuations (and responses) for rich product information, category discounts and shipping offers.

Past studies have found that while there may be qualitative agreement in the characterizations of segments obtained using model based and non model-based approaches, the assignments (and sizes) may be nearly independent (e.g., Andrews et al. 2010; Vriens et al. 1996). Thus, it is important to choose the best approach on the basis of the goals of the study. Model-based approaches have been observed to provide better fit than traditional clustering, and are preferred when the goal is to understand the true segmentation structure in the data along with the nature of the regression relationship within sessions (e.g., Andrews et al. 2010; McLachlan and Peel 2000; Wedel and Kamakura 2000). Further, it gives us a formal way to select the optimal number of clusters, and can handle data belonging to different measurement scales (McLachlan and Peel 2000). In terms of modeling heterogeneity, finite mixture models provide us an attractive middle ground between random coefficient models, that apply a continuous mixing distribution to efficiently estimate average effects but remain uninformative about responses at specific disaggregated levels, and hierarchical Bayesian models that estimate individual-level parameters. Given our interest in identifying groups of sessions that are relatively homogeneous, we determine that a mixture model is especially useful and managerially appealing in our context.

Our data display a high proportion of zero outcomes, as is expected for a purchase dataset. One approach to handle this is to use hurdle or zero inflated models that separate the probability of obtaining a zero outcome from the probability of nonzero outcomes (Winkelmann 2008). However, consumers belonging to any state can experience a non-zero probability of purchase, and conversely, consumers who have a zero outcome do not necessarily belong to the same subpopulation (e.g., Wang et al. 1996; Deb and Trivedi 1997). For example, browsers who do not buy have different underlying reasons than goal-oriented buyers who do not buy. The finite mixture model is appropriate for our needs as it allows zero and non-zero values to be realized from the same underlying stochastic process, and provides better fit than alternatives.

While our primary dependent variable is $Purchase$ – due to technical complications in the identification of binary mixtures, we use the count outcome (there is a high correlation between $Purchase$ and $Purchase\_cnt$ with ρ = 0.904, p = 0.000) where we model components as derived from the Poisson distributional family, for which generic identifiability has been shown (Teicher 1963; Titterington et al. 1985) [6]. In later models, we show robustness of the responses to information and incentives to alternate specifications.

Model of States of Shopping

Let Y be the non-zero integer valued random variable that measures the count of purchase completion pages visited by the user in a session. In a Poisson mixture model, the probability mass function of Y is given by $\boldsymbol{P(Y = y) = \int \frac{\exp(-\lambda)\lambda^y}{y!} f(\lambda) d\lambda}$, where λ, the mean or E[Y], is

[6] Identification in a binary outcome model generally requires that we observe consumers repetitively in T> 2K-1 sessions (K is the number of component distributions), which would lead us to drop a large mass of our sample.

treated as a stochastic variable with a discrete mixing density function f(λ). In a finite mixture model, f(λ) arises from a fixed number of components G.

$$\boldsymbol{P(Y=y)} = \sum_{g=1}^{G} \frac{\mathbf{exp}(-\boldsymbol{\lambda_g})\boldsymbol{\lambda^{y_g}}}{\boldsymbol{y}!}\boldsymbol{\pi_g} \tag{1}$$

Where π_g is the prior probability that an observation belongs to g = G. $\mathbf{0} < \boldsymbol{\pi_g} < 1$ and $\sum_{g=1}^{G} \boldsymbol{\pi_g} = \mathbf{1}$. For identification, we follow the labeling restriction that $\boldsymbol{\pi_1} \geq \boldsymbol{\pi_2} \geq \cdots \geq \boldsymbol{\pi_G}$ (Titterington et al. 1985). λ_g is the component specific mean or rate. The log of the component-specific rate is modeled as a linear function of covariates thought to exhibit differences across latent groups of consumers. Given that mixture models get easily complicated to estimate when the parameters grow, we specify a simple yet parsimonious model to determine the usefulness of a mixture setup for our data. The covariates include the breadth, depth and intensity (BDI) measurers (*RatioProdtoCatPages* is excluded due to collinearity), and whether consumers obtained each of the three types of online information.

$$\begin{aligned} \boldsymbol{\lambda_{ig}} &= \boldsymbol{\lambda_g}(\boldsymbol{x_i}, \boldsymbol{\beta_g}) = \boldsymbol{exp}(\boldsymbol{\beta_g'}' \boldsymbol{x_i}) \\ &= \boldsymbol{exp}\ (\boldsymbol{ProdInfo_i'\beta_{prod}} + \boldsymbol{SPriceInc_i'\beta_{sprice}} + \boldsymbol{GPriceInc_i'\beta_{gprice}} + \boldsymbol{BDI_i'\beta_{bdi}}) \end{aligned} \tag{2}$$

Where $i = 1, \dots \dots, N$ sessions and y_i^* is the unobserved continuous random outcome variable; $\boldsymbol{\beta_{prod}}, \boldsymbol{\beta_{sprice}}, \boldsymbol{\beta_{gprice}}$ are the coefficients of online information; and $\boldsymbol{\beta_{bdi}}$ are the coefficients of BDI measures used to characterize latent states. The mixture distribution is given by the weighted sum across the **g** components, and can be expressed as

$$\begin{aligned} &\boldsymbol{f}(\boldsymbol{y_i} |\ \boldsymbol{ProdInfo_i}, \boldsymbol{SPriceInc_i}, \boldsymbol{GPriceInc_i}, \boldsymbol{BDI_i};\ \boldsymbol{\Omega_1}, \dots \boldsymbol{\Omega_G};\ \boldsymbol{\pi_1}, \dots, \boldsymbol{\pi_G}) = \\ &\sum_{g=1}^{G} \boldsymbol{\pi_g f_g}(\boldsymbol{y_i} |\ \boldsymbol{ProdInfo_i}, \boldsymbol{SPriceInc_i}, \boldsymbol{GPriceInc_i}, \boldsymbol{BDI_i}\ ;\ \boldsymbol{\Omega_g}) \end{aligned} \tag{3}$$

$\boldsymbol{\Omega_g}$ are component parameters that are estimated by maximizing the following log likelihood

$$\mathbf{max}_{\pi,\Omega}\ \boldsymbol{LL} = \sum_{i=1}^{N} \left(\mathbf{ln}\left(\sum_{g=1}^{G} \boldsymbol{\pi_g f_g}(\boldsymbol{y_i} |\ \boldsymbol{ProdInfo_i}, \boldsymbol{SPriceInc_i}, \boldsymbol{GPriceInc_i}, \boldsymbol{BDI_i};\ \boldsymbol{\Omega_g})\right)\right) \tag{4}$$

The posterior probability that observation $\boldsymbol{y_i}$ belongs to component **g is** given by Bayes theorem conditional on observed covariates and the outcome.

$$\mathbf{Pr}\ (\boldsymbol{y_i} \in \textbf{population g} \mid \boldsymbol{ProdInfo_i}, \boldsymbol{SPriceInc_i}, \boldsymbol{GPriceInc_i}, \boldsymbol{BDI_i}, \boldsymbol{y_i};\ \boldsymbol{\Omega}) =$$

$$\frac{\pi_g f_g(y_i \mid \boldsymbol{ProdInfo_i}, \boldsymbol{SPriceInc_i}, \boldsymbol{GPriceInc_i}, \boldsymbol{BDI_i}, \boldsymbol{\Omega_g})}{\sum_{g=1}^{G} \pi_g f_g(y_i \mid \boldsymbol{ProdInfo_i}, \boldsymbol{SPriceInc_i}, \boldsymbol{GPriceInc_i}, \boldsymbol{BDI_i}, \boldsymbol{\Omega_g})} \tag{5}$$

Each session is then assigned membership into a group representing a different latent state of shopping for which it has the largest (posterior) probability.

Estimation and Results

The model is estimated using the EM algorithm within the maximum likelihood framework (Dempster et al.1977). For each fixed value of the number of components, the unobserved component memberships of the observations are treated as missing values and the data are augmented by estimates of the component memberships, i.e. the estimated a-posteriori probabilities, iteratively. Estimation requires the provision of initial membership values, and we re-ran the models with multiple random starting points in order to avoid local optima. A necessary condition for identification – that the matrix of covariates are of full rank – is verified.

The model is estimated by increasing the components from 1 to 4, and the best model is chosen using information criteria: AIC and AIC3 (Bozdogan 1987), and BIC (Schwarz 1978). Following the principle of parsimony, we prefer a model with fewer parameters for the same log likelihood, all else equal. The best model minimizes $-2 * LL + p * d$; where p is the number of free parameters in the model and $d = 2$ for AIC, $d = log\ N$ for BIC, and $d = 3$ for AIC3. These criteria suggest that the 3-component solution provides the best fit as shown in Table 3. We calculate the entropy measure to assess the degree of separation in the estimated posterior probabilities (Ramaswamy et al. 1993; Wedel and Kamakura 2000, p.92). The measure $E_G =$

$1 - \frac{\sum_{n=1}^{N}\sum_{g=1}^{G} -p_{ng} ln p_{ng}}{N\ lnG}$ where p_{ng} is the posterior probability that session n belongs to state g, is the highest for the the 3-state solution (0.87), suggesting reasonably good assignments.

The Latent States: Identification and Characterization

Our goal in using mixture models is to uncover underlying differences across consumer sessions. We present the differences in BDI characteristics across the obtained states in Table 4a. Sessions in State 1 had the lowest value for department and category breadth – that is they visited fewer categories compared to sessions in State 3, which had the highest numbers on both breadth measures. Customers in State 2 performed the highest number of hierarchical drill-downs or depth of search while customer sessions in State 3 contained the fewest. Customer sessions categorized as State 1 viewed the highest number of pages and spent the longest time on the website. Sessions in State 2 and 3 differed little along these two attributes. However, customers in State 2 viewed a significantly higher number of product pages (nearly double that of customers in state 1). Thus while customers in State 1 viewed more pages overall, only a small share were product pages; a majority were store, promotions, specials, and policy related pages.

Another distinguishing variable is the ratio of product to category level pages – or intensity of product search within (focal) product categories - which is the highest for sessions in State 2, followed by sessions in State1 and then 3. The low number for State 3 indicates either that customers viewed fewer product pages or conducted a dispersed search across many categories. Customers in States 1 and 2 viewed significantly fewer product pages per minute than those in State 3, also indicating that the former were more engaged with products.

These differences are also supported by the multinomial logit analysis used to predict the posterior probability of belonging to each state presented in Table 4b. In distinguishing state 1 from state 3, the most important factors were that: state 1 viewed significantly fewer detailed product pages; spent more *TotalTime*, had lower *CatBreadth*, but higher *DeptBreadth*, suggesting they viewed few product categories but many (purchase relevant) departments, and had significantly lower *Depth*. In distinguishing state 3 from state 2, the most significant factors were that: state 3 spent significantly lower *TotalTime*, had a higher *CatBreadth* and *DeptBreadth*, and viewed significantly fewer detailed product pages *TotalProducts*.

On the basis of the above results, we classify State 1 as *directed shoppers* (DS) who are the closest to a purchase decision, State 2 as *deliberating researchers* (DR) who are researching and learning about products in the focal category, and State 3 as *browsers* (BR) whose interests were relatively less directed. It is relevant here to compare our categorization of sessions to past works. Our mapping jointly considers the varying impacts of information and price incentives and navigation behaviors in the mixture models; whereas previous studies did not aim to do that. The closest work is Moe (2003) that used cluster analysis and found two additional types of sessions - knowledge builders and shallow visitors. Knowledge builders viewed mostly informational pages to learn about the store and saw few product pages. Shallow visitors viewed only two or fewer shopping pages. These two types of sessions are excluded from our dataset by construction. Recall that we only include sessions where visitors viewed products, and we dropped cookies with a single session with very few clicks. As described by Moe (2003), customers in both these sessions do not have shopping-related goals. Our set of states appears to adequately cover the space of visits with shopping-relevant goals found in prior work.

We next examine purchase outcomes – which should differ across states if our labeling provides face validity (see Table 5a). Sessions from directed shoppers had the highest propensity to both add products to the cart (14.49%) and complete the purchase (5.13%), whereas sessions from browsers had the lowest proportions for both. Interestingly, conditional on adding to cart, browsers had a higher likelihood of completing the purchase (20.2%) than deliberating researchers (17.2%), but lower than directed shoppers (35.4%). Further, customers conducting research were the least likely to use decision aids to screen alternatives, indicating their greater reliance on compensatory choice processes in building their consideration sets. Directed shoppers and browsers displayed greater non-compensatory search through the use of decision aids to quickly narrow down the assortment. Directed shoppers also displayed high usage of text search to directly find products they wanted. Finally, directed shoppers had the highest likelihood of being a repeat visit for a cookie.

In Table 5b, we present the response parameters for information and incentives obtained from the finite mixture model in column (1). The results show that product information had a small negative impact on directed shoppers, whereas it had a positive effect on browsers and the strongest impact on deliberating researchers. Category specific price incentives had a positive effect for directed shoppers and browsers, whereas it had a negative, albeit insignificant, impact on deliberating researchers. The effect of the free shipping incentive was positive in all three states. The results from a mixture model run using only the responses to information / incentives as a segmentation basis are robust (column 2). Table 5b provides preliminary evidence of the different influences of product information and price incentives across the states. Next, we develop a panel model to that also accounts for unobservables at the cookie level, and controls for other relevant factors that impact purchase..

INFORMATION, INCENTIVES, AND STATES OF SHOPPING

Durable goods can involve significant costs and are a high-involvement purchase category. Search for information is therefore an important step in consumer decision-making and an integral element of major consumer behavior frameworks (Bettman 1979; Howard and Sheth

1969). The framework by Putsis and Srinivasan (1994, hereafter P&S) – which addresses the issues of information acquisition behavior as well as the duration of the purchase deliberation process - is useful in describing consumers' decision-making for durable goods . It proposes that consumers are differentiated in the distance between their conditional indirect utilities with and without purchase, and will buy only when this utility difference exceeds zero (see Figure 1 for details). Of particular interest to our context is the fact that this distance can be altered by purchase-relevant information. While P&S describe consumers as differing in the distance between utilities, they do not tie the differences to varying shopping needs, as we do in this study. We theorize about the different impacts that product information and price incentives will have on the utilities of consumers with different shopping goals.

Customers interested in durable goods must perform a number of tasks prior to facing the final purchase decision. Marketers have long referred to a buying funnel where consumers pass through stages of decision-making which often include awareness/interest, research, consideration and finally, action. As a group, online consumers at a retailer's website express a relatively greater level of awareness and interest in shopping for durable goods than consumers who view advertisements and marketing interventions in non-shopping contexts. However, unlike browsers who display general interest, deliberating researchers are actively viewing and visiting detailed product pages, while shoppers display directed purchase-oriented behaviors; suggesting that consumers in the three states of shopping have different levels of progression through the funnel. They, therefore, also have different purchase relevant goals.

Of the three states, deliberating researchers have the greatest need to compare and evaluate the alternatives available at the retailer. As seen in Table 5a, they were more likely to use compensatory strategies to evaluate (a subset of) alternatives. They attempt to learn about the products, and ultimately narrow down the options to a smaller set for more serious consideration. Browsers and directed shoppers, on the other hand, are less driven by the need to perform detailed comparisons of alternatives during their online session. The main difference between the latter two is that while browsers typically lack a refined set of products that they are seriously considering buying from, directed shoppers usually have already formed such a set and buy when they find attractive deals or prices. Thus, a browser jointly seeks a desirable product and a suitable price, whereas a shopper searches for a suitable price for a product that she values. Consumers in both these states are more likely to evaluate a product by itself (benefits vs. price), rather than perform relative evaluations and comparisons across alternatives as deliberating researchers do. These unique differences in goals, we argue, will lead consumers to be influenced differently by product information, category discounts and shipping offers.

Information that matches consumers' shopping needs increases the likelihood that their utility difference with and without the good turns positive, and they cease search and purchase. Alternatively, some types of information may increase this utility difference by introducing uncertainties that require consumers to negotiate difficult trade-offs between alternatives thereby making choice more difficult. When this occurs, consumers have been observed to defer their decisions in order to address the newly formed uncertainties (e.g.,Chernev 2003; Dhar 1997; Gourville and Soman 2005; Iyengar and Lepper 2000).

While directed shoppers and deliberators exhibit goal-oriented shopping behaviors, and are more likely to buy, browsers may also purchase (see Table 5a). Researchers have put forth several inconclusive explanations for why browsers buy. Past work has found that impulse buying is not confined to any particular product type – consumers have been observed to buy several durable goods on impulse (e.g., Kollat and Willett 1967). However, given that durable goods involve significant costs, we expect that browsers who purchase are less likely to have been driven purely by impulse; rather it is likely that they have been in the market for some time, conducting ongoing search and accumulating relevant knowledge in the product category of interest. Stern (1962) showed that impulse buys could range from pure impulse to planned buys.

We now consider the influence of the three types of information and incentives. $ProdInfo$ informs consumers about features and highlights the pros (and cons) of alternatives, thereby making it easier for consumers to compare alternatives. It educates and enhances consumers' product experience, while also helping them build a consideration set that best matches their needs. Thus, product information is most useful for goal-oriented deliberating researchers, and can move them closer to completing the purchase. Browsers, on the other hand lack a immediate purchase need, and may not search purposefully but may still purchase when the stimuli they encounter causes the difference in their indirect utilities with and without purchase to turn positive (Bloch et al.1986). Research has shown that, "webmospherics" that engages browsers, and aids them to better experience the product may be successful in driving impulse buys (Childers et al. 2001). Since browsers are likely to evaluate products individually rather than conduct relative and thorough comparisons across alternatives, $ProdInfo$ may induce browsers to purchase by causing emotional reactions to and greater involvement with the selected products (Rook, 1987; Rook and Gardner, 1993).

The other group of goal-oriented sessions represents directed shoppers; but they have different shopping goals than deliberating researchers. The information provided by rich multimedia, buying guides and comparators induces directed shoppers to focus again on features and attributes of alternatives, and re-evaluate the available assortment of products. This can create ambivalence and cause negative reactance (Fitzsimons and Lehmann 2004) especially when it contradicts consumers' original choices by highlighting product-relevant aspects that the consumer may have overlooked before (e.g., new criteria to use for comparisons), For a consumer who has or is near finalizing the product to purchase, $ProdInfo$ may therefore complicate decision-making and stimulate her to search more to reduce the newly arisen ambiguity (Xia and Monroe 2004).

P1. Product information has positive effects on conversion for deliberating researchers followed by browsers, but has a negative effect for directed shoppers.

Next, we consider $SPriceInc$. Directed shoppers are the closest to the final purchase decision, having already determined the most desirable product(s). Learning about a sale in the focal product category can thus be extremely successful in incentivizing her to purchase, and preventing her from delaying, or worse, abandoning the site in search of better prices elsewhere. Browsers display general interest in the category but do not evaluate alternatives the way deliberators do. Since they are usually not seeking specific products, they tend to be influenced by the overall value of a product including its price. $SPriceInc$ can thus be extremely valuable in converting browsers. Nearly 75% of respondents in an industry study (Yankee Group and Ernst and Young 2002) cited a special sale price as the top factor that contributed to a spontaneous impulse purchase. This is likely to be the case when browsers aware of a need in a product category are mindful (albeit passively) about associated deals, and obtain positive utility in purchasing when discounts increases the overall attractiveness of a product in that category.

However, a sale in a specific category raises the attractiveness of all products in that category, and may increase the number of alternatives that satisfy the feasibility (budget) constraints of a consumer. For deliberating researchers attempting to narrow their alternatives, such information can have negative purchase consequences. Encountering larger selections has been found to cause choice overload and increase choice conflict, resulting in choice deferral (e.g., Iyengar and Lepper 2000). The delay was observed to be greater when the assortment considered by the consumer was increased to include alternatives that were non-alignable (Chernev 2003; Gourville and Soman 2005). This may happen when, for example, more appliances fall into a consumer's feasible set, but they include alternatives that vary in features, thereby making comparison among them more difficult for the deliberating consumer. We therefore expect that the conversion of deliberating researchers will be negatively influenced by $SPriceInc$ which causes them to delay purchase in order to resolve the newly formed uncertainties. . A similar effect is not expected to occur for browsers and shoppers because of their lower tendency to compare and evaluate available alternatives relative to each other.

P2. Category specific price incentive has a positive effect on conversion for directed shoppers and browsers, but has a negative effect for deliberating researchers.

Finally, when buying durable goods in an online market, consumers are likely to be influenced by relevant channel-related concerns - the most prominent being the presence of shipping/delivery fees for products purchased online but not offline. Indeed, recent industry studies conducted by PayPal and comScore found that the leading cause of online shopping cart abandonment cited by 46% of respondents was high shipping charges[7]. Lewis et al. (2006) find that the increased salience of online shipping fees causes consumers to often overweight shipping charges, and this is likely to be so for durable goods with substantial shipping charges. $GPriceInc$ (like $SPriceInc$) is a limited time offer that generates a sense of urgency among shoppers, but is not directly tied to a particular product. Inman et al. (1997) found that consumers in conditions of high need for cognition were the least affected by such generic offers. We expect that deliberating researchers who are actively seeking information about products, and engage in

[7] Eighth Annual Merchant Survey (April 2009) sponsored by PayPal and comScore.

effortful and systematic thinking would be least affected by $GPriceInc$. Shipping offers are however most likely be effective for customers closest to making a purchase and actively searching for the best purchase price such as directed shoppers. Also, browsers who display passive interest in the category may respond impulsively to attractive limited time offers.

P3. Generic price incentives have positive effects on conversion for directed shoppers and browsers, and to a lesser extent for deliberating researchers.

In addition to the within-session impacts, we also examine the impacts of information and incentives obtained in one session on purchase behaviors in the following visit for the sample of cookies with repeat visits. Findings from past works suggest that when consumers are in a state of flow – described as intrinsic shopping enjoyment with concentration and attention – they are more likely to experience stickiness to the website, and be more likely to return (Koufaris 2002). As compared to price incentives, rich product information engages the customer to spend time at the website and learn about its offerings, thereby increasing consumers' level of flow and attention. Such information can increase customer loyalty, and likelihood to return to visit the store even when they do not purchase during a current session. Price discounts however attract consumers who are induced by the lower price,and because they are less engaged with the retailer itself, may not return when they leave without purchasing in a session and find better prices elsewhere. Next, we develop a model to assess the relative effects of product information and price incentives on outcomes *within* session versus *across* sessions.

P4. Product information is more successful than price incentives in influencing conversion in future sessions when consumers return.

Model of Purchase Outcomes

We use a random-effects model to account for two forms of heterogeneity. The first is cookie-level unobserved heterogeneity which captures the underlying dependence across sessions from a single cookie, and is time-invariant across its sessions, modeled using random effects. Additionally sessions from a cookie may belong to different latent states of shopping across repeat visits. This time-variant heterogeneity is modeled using session-level dummies.

We are primarily interested in the effects of online information/incentives, and examine how their coefficients vary across sessions using interactions between product information, category discounts and generic shipping offers, and the three states- DS, DR, and BR. Since information about products and prices can be accumulated across visits to a retailer, we create lagged cumulative measures -$\boldsymbol{PastProdInfo}, \boldsymbol{PastSPriceInc}, \boldsymbol{PastGPriceInc}$. We also include dummies to represent the state of the preceding session ($\boldsymbol{LATSTATE}$) to help control for the effects of state transition on purchase outcomes[8]. Next, we describe the model setup for our primary outcome – the binary measure Purchase .

Model

$$\begin{aligned} \boldsymbol{y}_{it}^{*} = {} & \boldsymbol{PastProdInfo}'_{it}\,\boldsymbol{\gamma}_{pProd} + \boldsymbol{PastSPriceInc}'_{it}\,\boldsymbol{\gamma}_{pSprice} + \boldsymbol{PastGPriceInc}'_{it}\boldsymbol{\gamma}_{pGprice} \\ & + \boldsymbol{PastCart}'_{it}\boldsymbol{\gamma}_{pCart} + \boldsymbol{ProdInfo}'_{it}\,\boldsymbol{\gamma}_{prod} + \boldsymbol{SPriceInc}'_{it}\,\boldsymbol{\gamma}_{sprice} \\ & + \boldsymbol{GPriceInc}'_{it}\,\boldsymbol{\gamma}_{gprice} + \boldsymbol{LATSTATE}'_{it-1}\,\boldsymbol{\gamma}_{pLatstate} + \boldsymbol{LATSTATE}'_{it}\,\boldsymbol{\gamma}_{latstate} \\ & + [\boldsymbol{ProdInfo}'_{it} + \boldsymbol{SPriceInc}'_{it} + \boldsymbol{GPriceInc}'_{it}] * \boldsymbol{LATSTATE}'_{it}\,\boldsymbol{\gamma}_{info*latstate} \\ & + \boldsymbol{BDI}'_{it}\,\boldsymbol{\gamma}_{bdi} + \boldsymbol{x}'_{it}\,\boldsymbol{\gamma}_{x} + \boldsymbol{\alpha}_i + \boldsymbol{\eta}_{it} \end{aligned}$$

$$where\ i = 1, \dots\dots, N\ \ cookies\ \ t = 1, \dots. T\ sessions \tag{6}$$

$\boldsymbol{y}_{it} = \mathbf{1}(\boldsymbol{y}_{it}^{*} > 0)$ and $\boldsymbol{\nu}_{it} = \boldsymbol{\alpha}_i + \boldsymbol{\eta}_{it}$

$\boldsymbol{\gamma}_{pProd}, \boldsymbol{\gamma}_{pSprice}, \boldsymbol{\gamma}_{pGprice}$ are the coefficients of information accumulated in past sessions

[8] We cannot model the impact of information/incentives on state transition patterns because of the possibility that customers may visit the offline store in between their online visits. We therefore treat online state as exogenous.

$\boldsymbol{\gamma_{pCart}}$ is the coefficient of the number of cart adds in past sessions

$\boldsymbol{\gamma_{prod}}, \boldsymbol{\gamma_{sprice}}, \boldsymbol{\gamma_{gprice}}$ are the coefficients of information obtained in the current session

$\boldsymbol{\gamma_{pLatstate}}$ are the dummies that represent the state of the immediately previous session

$\boldsymbol{\gamma_{latstate}}$ are the dummies to represent the (latent) state of the current session

$\boldsymbol{\gamma_{info*latstate}}$ are the coefficients for the interactions between information and latent state

$\boldsymbol{\gamma_{bdi}}$ are the coefficients of the breadth, depth and intensity variables

$\boldsymbol{\gamma_{x}}$ are the coefficients for observed session-level control variables.

α_i is the unobserved cookie-level random effect which is assumed to be uncorrelated with the covariates. These individual effects are distributed $\boldsymbol{\alpha_i \sim N(0, \sigma_\alpha^2)}$. η_{it} is the i.i.d. random error term; $\boldsymbol{\eta_{it} \sim N\ (0, \sigma_\eta^2)}$ and represents unobservables that are uncorrelated across sessions and cookies. The variance of ν_{it} is given by $\boldsymbol{Var(\nu_{it}) = \sigma_\nu^2 = \sigma_\alpha^2 + \sigma_\eta^2}$ and $\boldsymbol{cov(\nu_{it},\ \nu_{is}) = \sigma_\alpha^2}$ if the sessions belong to the same cookie, and 0 otherwise. The variance of the pure shocks is normalized to one. The fraction of the total error variance due to the individual consumer level component of the error term is given by the intra-group correlation coefficient

$$\boldsymbol{\rho = \frac{\sigma_\alpha^2}{\sigma_\nu^2} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\eta^2} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + 1}} \tag{7}$$

Let z contain the covariates in equation (7), then the probability of observing the given outcome conditional on the cookie random effect is given by

$$\mathbf{Pr}(\boldsymbol{y_{it} \mid z_{it}, \alpha_i}) = \boldsymbol{\Phi[(z_{it}'\gamma + \alpha_i)(2y_{it} - 1)]} \tag{8}$$

The likelihood for each unit is given below

$$\boldsymbol{L_i} = \mathbf{Pr}\big(\boldsymbol{y_{i1}, y_{i2,\dots,} y_{iT}}\big) = \int_{-\infty}^{\infty} \boldsymbol{\phi\ (v_{i1}, v_{i2}, \dots\ v_{iT} | \ \alpha_i)\ \phi(\ \alpha_i)\ \partial\alpha_i} \tag{9}$$

The dependence between v_{it}'s is attributable to the shared variation in α_i (due to the assumed independence between z_{it} and α_i). By conditioning on the α_i we integrate them out of the likelihood and evaluate the one-dimensional integral in (10) by using Gauss-Hermite quadrature (Greene 1997, p.190). The Log-likelihood of the model described in (7)-(10) is given by:

$$LL = \sum_i \sum_t y_{it}\, ln\ F(z'_{it}\gamma) + (1 - y_{it}) \ln (1 - F(z'_{it}\gamma)) \quad (10)$$

Results

The results from the cookie-panel model are presented in Table 6. In columns (1)-(3), we present the coefficient estimates using step-wise additions of the interaction terms between the three states of shopping and product information/price incentives. The full model is presented in column (4), and is the one we use to examine the effects below.

States of shopping: To assess the effects of latent states, we convert the coefficients to obtain probability estimates for a change in the value of the state dummy from 0 to 1, when all other covariates are held at their mean (or median for binary variables). Directed shoppers had the highest rate of conversion followed by deliberators and browsers, controlling for covariates. The coefficient of $PastLatState\_DS$ is negative, while the coefficients of $PastLatState\_DR$ and $PastLatState\_BR$ are positive and significant. A directed shopper that does not purchase but returns (in any state) has a significantly lowered tendency to purchase. However, returning to shop after having been in either the deliberating or browsing states significantly increased the likelihood of purchase. Thus, sessions abandoned by directed shoppers are a costly loss. This highlights the importance of converting directed shoppers in the current session itself. In Appendix A, we provide further details on state transitions.

Information and Incentives: To examine the within-session purchase impacts of information and incentives obtained by a consumer during the session, we need to assess both the main-effect coefficients and their interactions with states of shopping. For ease of comparison, these coefficients are presented in Table 7 in column (1).

Product Information had the strongest impact on within-session purchase for deliberating researchers who were gathering information about, and assessing or comparing the alternatives in a category. Buying guides and multimedia tools aid the deliberating consumer in lowering choice uncertainty and choosing a product that best matches her needs. This result supports our expectation. In contrast, product information lowered the likelihood of purchase in a given session for directed shoppers. This effect, though surprising, is in line with our theorizing that information that increases a consumer's purchase uncertainty, relative to her state of shopping, may cause her to delay (or abandon) purchase until the uncertainty is resolved. A directed buyer who has chosen a focal product, when presented with rich product information, guides and tools, may learn about or be reminded of additional features and attributes that she had not considered or had overlooked, thereby causing her to re-evaluate the set of alternatives. The impact on browsers suggests a significant albeit small effect. Browsers were more likely to purchase in the presence of product information than in its absence, implying that rich media positively engages and attracts browsers towards conversion.

Category specific price incentive had significant positive impacts on both directed shoppers and browsers, as expected, leading them to convert more often than in its absence. Directed shoppers who are likely to be seeking the best price for the product(s) that they have selected may price-shop across retailers in search of deals; obtaining a promotion can incentivize them to purchase from the said retailer. Browsers typically lack an immediate purchase need, but rather possess an ongoing or passive interest in a focal product. Learning about a discount in the product category increases the attractiveness of the focal product for the browser, resulting in a conversion. Specific price incentive however did not induce similar effects in deliberating researchers; rather, it had a negative effect on their purchase behavior. Such consumers are in the process of comparing alternatives and narrowing down their consideration set; a category level price discount increases the set of alternatives that is now newly feasible, and is therefore likely to increase their choice uncertainty. This result also supports our theorizing that information that increases purchase uncertainties relevant to a state of shopping hinders purchase.

Generic price incentive or shipping related offers were found to have across-the-board positive impacts on sessions belonging to all three states, suggesting that online consumers highly value shipping offers, and treat them differently than other price incentives. In other words, the absence of shipping promotions lowered the purchase likelihood for all consumers. The strongest positive impact of shipping offers is interestingly observed for browsers, followed by directed shoppers and then deliberators. This result suggests that directed shoppers, while positively affected by shipping offers, are less likely than browsers to abandon their purchase when such an offer is not available. One explanation is that directed shoppers are close to finalizing their purchase and have a deeper commitment to the purchase than browsers. Deliberators who have not yet completed their evaluations and formed their preferences experienced weaker (nevertheless significant) effects on buying.

Overall, these findings are consistent with our theorizing about within-session impacts. Directed shoppers were strongly influenced by $SPriceInc$ and negatively affected by $ProdInfo$. Deliberating researchers were strongly influenced by $ProdInfo$, while $SPriceInc$ hampered their purchase, and finally, browsers experienced the strongest purchase reaction to $SPriceInc$ and to a much smaller extent to $ProdInfo$. $GPriceInc$ had a universal positive effect with the effect on deliberating researchers comparatively smaller than on browsers and directed shoppers.

We additionally observe important effects on across-session outcomes. In Table 6, the coefficient of $PastProdInfo$ is positive and significant, while the coefficients of $PastSPrice$ and $PastGPrice$ are negative, with only the effect of $PastSPrice$ significant. The cumulative effects of product information obtained in earlier sessions had a positive impact on conversion in the current session. By contrast, the accumulated effects of category specific price promotions obtained in the past sessions had a negative effect on purchase in a given session. This finding highlights the value of $ProdInfo$ in helping engage the customer and in building a relationship with them that extends beyond a given session, while encountering too many price incentives in the past sessions had a negative effect on purchase in a given session, when such discounts were no longer available.. Given concerns echoed by retailers about consumers who are price-sensitive and respond only to price promotions but are typically not loyal and hunt for deals (e.g., McWilliams 2004), our results show that retailers can benefit by investing in creating rich product experiences for their customers. Additionally, this finding highlights the negative future effects of promotions when consumers expect them but they may no longer be available.

These findings provide support for propositions P1, P2 and P3, and suggest that differences in consumers' search behaviors across the states plausibly produce different impacts of product information and price incentives on online purchase outcomes.

Further, comparing the effects of online information on within-session and across-session outcomes highlights important tradeoffs. $ProdInfo$ has significant within-session influence on purchase behaviors for deliberating researchers (and to a smaller extent browsers), but it negatively influences the within-session conversion of directed shoppers. Exposure to $ProdInfo$, however, had the strongest positive influence on purchase decisions for returning consumers irrespective of their state of shopping in the previous session. On the contrary, both types of price incentives displayed strong positive effects on within-session purchase for customers in two latent states of shopping- directed shoppers and browsers, but had weak to strong negative impacts on purchase for returning customers. This result supports proposition P4.

ASSESSING ROBUSTNESS

In this section, we assess the robustness of our findings to a number of potential concerns.

Restricted sample of interested customers: One limitation of using clickstream to study consumer's purchase behaviors is that we cannot ascertain the true intent of consumers. While consumers may have visited a product page during the session, it may not translate into true interest in the product and need not suggest that the product was considered for purchase. We therefore place a stronger restriction on the sessions that we include in the second sample. We require the customer to have displayed "substantial interest" in the focal product during at least one of her visits to the store. We consider adding a focal product to the shopping cart as a sufficient condition, and therefore include all sessions from this customer. For customers that did not add a product to the cart, we include sessions from only those who viewed the focal product multiple times during at least one of their visits. This *interested sample* contains 11,076 sessions. The main results from the *full sample* (column (4) in Table 6) are repeated in column (1) in Table 7 (for ease of comparison) and the results for the interested sample are displayed in column (2).

We also run an additional model restricted to only users that add the focal product to the shopping cart in at least one of their sessions – the *add to cart sample*. Since consumers may purchase the product offline, we restrict analysis to only consumers who indicate online purchase intent by adding the product to the shopping cart. This analysis is run using 4,885 sessions, and the results are presented in column (3) in Table 7. The robustness of the findings in columns (2) and (3) engender confidence in our main findings in column (1).

Endogeneity in product information: Unlike $SPriceInc$ and $GPriceInc$ that are offered to all customers on the days that the incentives were available, $ProdInfo$ is selected by the customer. $ProdInfo$'s impact on outcomes may suffer from endogeneity bias if customers who were more likely to purchase were also the ones more likely to obtain it. As a first step, we consider the proportion of customers in each of the three states that obtained information (see Table 5). We observe that fewer deliberating researchers, who appear to have the strongest positive impact on purchase from $ProdInfo$, obtained product information than directed shoppers, suggesting that endogeneity may not be a concern. To more rigorously address self-selection, we use the matching method to estimate the effects of $ProdInfo$.

We use propensity score matching to estimate average treatment effects by comparing the outcomes of treated and control groups that have been matched on the BDI covariates – which measure underlying differences that may potentially drive customers' decision to obtain $ProdInfo$. The propensity score is the conditional probability of being in the treated rather than the control group given the relevant observed covariates (Rosenbaum & Rubin, 1983). The groups thus obtained are, on average, observationally identical. More details are in Appendix B. The results from the matching analyses limited to consumer sessions with a common support are presented in Table 8. Matches are found using caliper matching ($r = 0.1$) in column (1) and block-stratified matching in column (2). The standard errors are bootstrapped. As observed there, our primary results remain robust. The coefficient of $ProdInfo$ is positive and significant for deliberators and browsers, whereas for directed shoppers, it continues to be negative.

Price vs. Brand Sensitivity of Consumers: Our premise is that the purchase behaviors and varying impacts of product information and price incentives that we observe across consumers are driven by membership in different (latent) states. An alternate explanation is that consumers who completed the purchase when provided with price (product) information were merely more price-sensitive (brand/feature-sensitive). We assess this possibility.

An important feature of shopping online is the availability of tools that allow consumers to refine/screen products, and thereby affect the consideration sets and the final products that they buy (e.g., Alba et al. 1997; Haubl and Trifts 2000; Lynch and Ariely 2000). We measure consumers' use of price ($PriceFacetedSearch$) vs. product ($ProdFacetedSearch$) attributes to screen alternatives as proxy measures of their price vis-à-vis product sensitivity. If price-product sensitivity drove our main results, we should find that deliberators are more product-sensitive than shoppers; and that shoppers and browsers are more price-sensitive than deliberators.

In Table 5, we found that on average, deliberating researchers had the lowest counts of product/brand refining (μ= 0.172, s.d. = 1.148), followed by browsers (μ = 0.371, s.d. =1.756) and directed buyers (μ = 0.570, s.d. =2.879). Deliberating researchers also had the fewest number of price refining counts (μ = 0.250, s.d. =1.413), while browsers (μ = 0.555, s.d. =2.135) and directed buyers (μ = 0.579, s.d. = 2.843) were similar. Browsers conducted more price-based than feature/brand-based refining operations. Directed shoppers were equally likely to refine using both types of attributes. Deliberators were least likely to use either refining criterion (but relied slightly more on price-based screening). These patterns do not indicate that a systematic correlation between latent states and price vs. brand sensitivities is what drives the results. .

In Table 9, we estimate a model where we include interaction terms between the three types of information/incentives and both types of refining to separately control for their effects. The relevant coefficients are displayed in Table 9 in column (1). ,Our main results for the effects of information and incentives obtained within a session on conversion remain consistent with our earlier findings from column (1) in Table 7. Even after controlling for price/brand sensitivity, the state of shopping remained a significant moderator of the impacts of information and incentives.

Correlated random effects: The random effects model used in Table 6 assumes that there is no correlation between the unobservable α_i and the $\boldsymbol{z_{it}}$ (esp the BDI covariates)**.** To relax this assumption we consider a correlated random effects model (Mundlak 1978; Wooldridge

2009). Essentially, a parametric relationship is specified $\alpha_i = \overline{z_i}'\theta + \upsilon_i$. Assume that $\boldsymbol{\upsilon_i}$ is i.i.d. normal and independent of $\mathbf{z_{it}}$ and the errors in the random effect specification, and has variance $\sigma_\upsilon{}^2$. Then, $E[y_{it}|z_{it}, \bar{z}_{it}, \upsilon_i] = h(z'_{it}\beta + \alpha + \overline{z}'_i\theta\ + \upsilon_i\ )$. These results for the main variables of interest are presented in column (2) in Table 9. Two variables –the mean of total time and the total pages viewed – were significant, suggesting some correlation with $\boldsymbol{\alpha_i}$. These results are consistent, suggesting that the assumption did not critically affect our results.

Alternate specifications of the outcome: We use $Purchase\_cnt$ as the dependent variable, and re-estimate the model using both a random effects panel Poisson model (column 3 in Table 9) and an ordered probit model (column 4 in Table 9). While some coefficients are only marginally significant, the results generally support our main findings here as well.

Adding to Shopping Cart as an outcome: Finally, we extend our analysis to examine an important intermediate outcome. We re-estimated our main models using adding to the shopping cart ($Cart$) as the outcome variable as shown in column (5) in Table 9. The results are largely consistent with our main purchase model. Deliberating researchers were more likely to add a product to the shopping cart upon retrieving relevant product information; whereas both directed shoppers and browsers were more likely to do so when they received either type of price-related (sales and shipping) information. This result is interesting because it suggests that the same type of information influences customers in a given shopping state to both add the product to the shopping cart and complete the purchase. This is counter to the belief that once customers have added products to the cart, only promotion information will influence them to consummate. We also test a joint model as described in Appendix C.

CONCLUSION

As noted by Montgomery et al (2004), clickstream data offers the ability to analyze not just the purchase occasion alone, but also the sequence of events that lead to various outcomes within a website. Thus a systematic analysis of clickstream data can provide valuable insights into customer needs and behaviors.

The main results of our study are the following. When focusing on conversion within a session, both browsers and directed shoppers are best influenced by price incentives (discounts, free shipping etc.). However, customers who are deliberating about alternatives responded best to product information. In our sample, we observed that the sessions from deliberators formed the largest group, slightly greater than the sessions from browsers and nearly three times larger than the sessions from directly buyers. This finding suggests that online retailers have the ability to induce customers to convert using rich product information if they are able to identify and target them when they are in the deliberation state. By persuading deliberating researchers to complete the purchase within a session, the retailer reduces the need to attract them using price levers when they return later as directed shoppers or browsers. This approach allows the retailer to then offer sales to the customers in states that obtain the greatest value from price incentives, and more importantly might have abandoned the session in their absence. Thus, by uncovering the unobserved state of the shopper, the retailer can appropriately target price vs. product information to the customer, thereby avoiding the need to offer margin-eroding promotions in order to incentivize customers to purchase. In fact, our results highlight the surprising negative effect of category price promotions on deliberating researchers. We also observed that product information distracted directed shoppers and led them to delay their purchase. Our within-session results shed light on the varied impacts of information across customers and also draw attention to the possible undesired consequences of mis-targeted information.

When examining conversion and purchase-related behaviors across sessions*,* our study suggests that there may be important tradeoffs in the impacts of information on purchasing within a session as compared to influencing customers to return to purchase in future sessions. Irrespective of the shopping state of the customer, product information had a significant positive impact on influencing non-purchasers in a given session to both return to the store and buy in that category in a future session. This highlights the important ability of product information to create stickiness in the site and loyalty among its customers. However, both types of price incentives – that had a positive impact on within-session conversion for browsers and directed buyers- appeared to have unfavorable impacts on the likelihood of purchase in future sessions when consumers did not buy in the current session. Finally, free shipping – that had broad positive impacts on within-session conversion for all three states of shopping - failed to have any impact on across-session purchase for all three states, highlighting its short-term effects. Further, it appears that online customers perceive category sales differently than shipping related offers

This study adds to a growing stream of work that suggests ways in which firms can improve their customer's experience by making websites more usable and navigable (e.g., Agarwal and Venkatesh 2002; Palmer 2002; Venkatesh and Agarwal 2006). Along these lines, our study sheds light on the impacts of providing product and price related marketing interventions for consumers in different shopping states.

Implications

The availability of micro-level consumer behavior data promises to bring online retailers closer to achieving truly customized interactions with their customers (Alba et al. 1997; Ansari and Mela 2003; Hoffman and Novak 1996). By analyzing the exploding volumes of clickstream data that are generated today, firms can quickly transform data into intelligence, and better understand the performance of alternate marketing interventions and make better business decisions. The clickstream tools typically available in the industry provide page-based views requiring more work to translate a series of clicks to pages into underlying shopping related needs of consumers. We show a simple approach which online firms can use to harness real time behaviors to infer differences across consumers and then provide appropriate information and incentives to aid in their conversion. Our study and its findings provide firms with knowledge that can be a useful starting point for generating business intelligence from clickstream data.

A particularly interesting aspect of our study is the use of observed and easily available real-time navigation activity on the website itself to generate the covariates required to determine the latent states of anonymous customers. This approach provides a number of benefits. First, it avoids the pitfalls surrounding the use of sensitive customer information that needs to be tracked over long periods of time. The use of real-time customer behaviors allows retailers to partially overcome the problem of the "gift-shopper" who is offered irrelevant promotions for children's toys when she later tries to search for business apparel. Second, since historical actions and pre-determined profiles are not always needed, these techniques allow retailers to actively target and interact with *new* visitors to their web store. Third, our model is consistent with shifting emphasis from the "static" user model to the "dynamic" behavior model which allows for the same consumer to be targeted in different ways on different occasions based on changing needs/preferences. By using such within-consumer targeting (different strategies for the different states of shopping a consumer may progress through), retailers can reduce the use of more controversial across-consumer targeting methods.

Further, our study questions the common practice of offering promotions such as free shipping and category discounts to all customers that are visiting a store. We argue that this strategy is suboptimal and results in retailers providing unnecessary promotions to customers who would have purchased anyway. We show that by learning about customers' latent states of shopping, retailers can instead optimally target information to customers who are less likely to purchase in its absence, thereby increasing the lift created by information.

Limitations and future extensions

Our current work is based on a sample observed over a short period that precludes us from studying purchases that may have occurred from customers returning after our observation period. We also group together different kinds of rich product information in this work, but it would be useful to tease apart the effects of constituents. This study should also be extended to examine the effects of user generated content (e.g., reviews).

In future studies, it will be useful to also examine the pathways of influence – how product information vs. price incentives differently affects customers' underlying purchase oriented structural parameters. For example, what is the impact of information on the buying threshold? Relatedly, when information does not incentivize customers to buy, does it help them to progress through the shopping funnel? In this study, we do not model the impact of price and product information on state transitions due to the limitation of unobserved offline store visits by customers. However, such a model with appropriate data can further help shed light on customer behaviors. Along these lines, it may also be useful to consider state transitions within a session; whereas we assume that consumers belong to a single state in a session, given by the most dominant behaviors within a session. Finally, while our current work is focused on the impacts of information obtained any time during the session, knowledge about timing or when in the session to provide different types of information would be complementary, and help firms make more specific decisions related to optimal provision of online information.

REFERENCES

Adomavicius, G., Sankaranarayanan, R., Sen, S., and Tuzhilin A. 2005. "Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach," *ACM Transactions on Information Systems* (23), pp.103–145

Agarwal, R., and Venkatesh, V. 2002. "Assessing a firm's Web presence: A Heuristic evaluation procedure for the measurement of usability," *Information Systems Research* (13:2), pp.168–186

Alba, J., Lynch, J., Weitz, B., Janiszewski, C., Lutz, R., Sawyer, A. and Wood, S. 1995. "Interactive Home Shopping: Consumer, Retailer, and Manufacturer Incentives to Participate in Electronic Marketplaces," *Journal of Marketing* (61: July), pp.38-53.

Allenby, G.M. and Rossi, P.E. 1999. "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics* (89: March/April), pp.57–78.

Andrews, R.L., Ansari, A. and Currim, I.S. 2002. "Hierarchical Bayes Versus Finite Mixture Conjoint Analysis Models: A Comparison of Fit, Prediction, and Partworth Recovery," *Journal of Marketing Research* (39: February), pp. 87-98.

Andrews, R. L., Brusco, M.J., Currim, I.S., and Davis, B. 2010. "An Empirical Comparison of Methods for Clustering Problems: Are There Benefits from Having a Statistical Model?." *Review of Marketing Science*, 8 (Article 3).

Ansari, A., and Mela, C. 2003. "e-Customization," *Journal of Marketing Research* (40: May), pp.131-145.

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

Bloch, P.H., Sherrell, D.L., and Ridgeway, N.M. 1986. "Consumer Search: An Extended Framework," *Journal of Consumer Research* (13), pp. 119-126

Bozdogan, H. 1987. "Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions," *Psychometrika* (52: 3), pp. 345–370.

Bucklin, R.E., and Sismeiro, C. 2003. "A Model of Web Site Browsing Behavior Estimated on Clickstream Data," *Journal of Marketing Research* (August), pp.249-267

Catledge, L.D and Pitkow, J.E. 1995. "Characterizing Browsing Strategies in the World-Wide Web," *Computer Networks and ISDN Systems* (27:6), pp. 1065–1073.

Chernev, A. 2003. "When More Is Less and Less Is More: The Role of Ideal Point Availability and Assortment in Consumer Choice," *Journal of Consumer Research* (30: September), pp.170–83.

Childers, T.L., Carr, C.L., Peck, J., and Carson, S. 2001. "Hedonic and Utilitarian Motivations for Online Retail Shopping Behavior," *Journal of Retailing* (77), pp.511-533

Chintagunta, P. K. 1993. "Investigating the Sensitivity of Equilibrium Profits to Advertising Dynamics and Competitive Effects," *Management Science* (39:9), pp.1146-1163.

Cox, D. F. 1967. *Risk Taking and Information Handling in Consumer Behavior*, Cambridge, MA: Harvard University Press.

Deb, P. and Trivedi, P. 1997, Demand for medical care by the elderly: a finite mixture approach, *Journal of Applied Econometrics*, 12.

Dempster, A., Laird, N., and Rubin, D. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B* (39:1), pp.1–38.

Dhar, R. 1997. "Consumer Preference for a No-Choice Option," *Journal of Consumer Research* (24: (September), pp.215–31.

Diehl, K., Kornish, L.J, and Lynch, J.G. 2003. "Smart Agents: When Lower Search Costs for Quality Information Increase Price Sensitivity," *Journal of Consumer Research* (30: 1), pp.56-71

Fitzsimons, G.J., and Lehmann, D.R. 2004. “Reactance to Recommendations: When Unsolicited Advice Yields Contrary Responses,” *Marketing Science* (23:1), pp. 82–94

Greene, W. 2003. *Econometric Analysis*, 5th Edition, Englewood Cliffs, NJ: Prentice-Hall.

Goel, S., Hofman, J.M., Lahaie, S., Pennock, D.M., and Watts, D.J. 2010. “Predicting Consumer Behaaior with web Search,” *Proceedings of the National Academy of Sciences* (107:41), pp. 17486-17490

Gourville, J.T., and Soman, D. 2005. “Overchoice and Assortment Type: When and Why Variety BackFires" *Marketing Science* (24:3), pp.382-395.

Häubl, G., and Trifts, V. 2000. “Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids,” *Marketing Science* (19:1), pp. 4-21.

Hoffman, D. and Novak, T. 1996. “Marketing in Hypermedia Computer-mediated Environments: Conceptual Foundations,” *Journal of Marketing* (60: July), pp.50–68.

Howard, J.A., and Sheth, J.N. 1969. *The Theory of Buyer Behavior*, NY: John Wiley & Sons/

Inman, J.J., Peter, A.C., and Raghubir, P. 1997. “Framing the Deal: The Role of Restrictions in Accentuating Deal Value,” *Journal of Consumer Research* (21:1), pp.68-79

Iyengar, S.S. and Lepper, M.R. 2000. “When Choice Is Demotivating: Can One Desire Too Much of a Good Thing,” *Journal of Personality and Social Psychology* (79:6), pp. 995–1006.

Jain, D.C., and Vilcassim, N.J. 1991. Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach, *Marketing Science* (10: Winter), pp.1-23

Janiszewski, C. 1998. “The Influence of Display Characteristics on Visual Exploratory Search Behavior,” *Journal of Consumer Research* (25:3), pp.290-301.

Jin, X., Zhou, Y., and Mobasher, B. 2004. "Web Usage Mining Based on Probabilistic Latent Semantic Analysis," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (KDD'04), Seattle

Johnson, E. J., Moe, W. W., Fader, P. S., Bellman, S., and Lohse, G. L. 2004. “On the Depth and Dynamics of Online Search Behavior,” *Management Science* (50:3), pp. 299-308.

Klein, L.R. and Ford, G.T. 2003. “Consumer Search for Information in the Digital Age: An Empirical Study of Prepurchase Search for Automobiles,” *Journal of Interactive Marketing* (17:3), pp.29-49.

Kollat, D.T. and Willett, R.P. 1967. “Customer Impulse Purchasing Behavior ,” *Journal of Marketing Research* (4:1), pp.21-31

Koufaris, M. 2002. “Applying the Technology Acceptance Model and Flow Theory to Online Consumer Behavior,” *Information Systems Research* (13: 2), pp. 205-223.

Lewis, M., Singh, V. and Fay, S. 2006. “An Empirical Study of the Impact of Nonlinear Shipping and Handling Fees on Purchase Incidence and Expenditure Decisions,” *Marketing Science* (25:1), pp.51-64

Lucas, P. 2001. “How Product Demos are Helping Shoppers Kick the Tires on the Web,” *Internet Retailer* Featured Article, Feb 2001, http://www.internetretailer.com/internet/marketing-conference/91973-how-product-demos-are-helping-shoppers-kick-tires-web.html

Lynch, J. G., Jr., and Ariely, D. “Wine Online: Search Cost and Competition on Price, Quality, and Distribution,” *Marketing Science*, 19 (Winter), 2000, pp. 83–103.

Manchanda, P., Dube, J-P., Goh, K Y and Chintangunta, P.K. 2006. “The Effects of Banner Advertising on Internet Purchasing,” *Journal of Marketing Research* (43**:** February), pp.98-108.

McAfee Report. 2010 “Digital Window Shopping: The Long Journey to Buy,” http://www.mcafeesecure.com/pdf/Digital-Window-Shopping_FinaL-web.pdf

McWilliams, G. 2004. Analyzing Customers, Best Buy Decides Not All Are Welcome, *Wall Street Journal*. Nov 8. http://online.wsj.com/article/0,,SB109986994931767086,00.html

McLachlan, G. J., and Peel, D. 2000. *Finite mixture models*. New York: J. Wiley
Moe, W. 2003. “Buying, Searching, or Browsing: Differentiating between Online Shoppers using In-Store Navigational Clickstream,” *Journal of Consumer Psychology* (13: 1&2), pp.29-40.
Moe, W., and Fader, P. 2004. “Dynamic Conversion Behavior at E-Commerce Sites,” *Management Science* (50:3), pp. 326-335.
Montgomery, A.L., Li, S., Srinivasan, K., and Liechty, J. C. 2004. “Modeling Online Browsing and Path Analysis Using Clickstream Data,” *Marketing Science* (23:4), pp.579-595.
Mundlak, Y. 1978. “On the Pooling of Time Series and Cross Section Data,” *Economet85rica* (46), pp. 69-
Murthi, B. P. S. and Sarkar, S. 2003. “The Role of the Management Sciences in Research on Personalization,” *Management Science* (49:10), pp. 1344-1362.
Palmer, J. W. 2002. "Web site usability, design, and performance metrics," *Information Systems Research* (13:2), pp.151–167.
Park, Y-H. and Fader, P.S. 2004. “Modeling Browsing Behavior at Multiple Websites,” *Marketing Science* (23: 3), pp. 280-303.
Putsis, W.P., and Srinivasan, N. 1994. "Buying or Just Browsing: The Duration of Purchase Deliberation," *Journal of Marketing Research* (31:3), pp. 393–402.
Ratchford, B.T., Lee,M-S. and Talukdar, D. 2003. “The Impact of the Internet on Information Search for Automobiles,” *Journal of Marketing Research* (40:2), pp.193-209.
Rook, D.W. 1987. “The buying impulse,” *Journal of Consumer Research* (14), pp.189–99
Rook, D.W., and Gardner, M.P. 1993. “In the Mood: Impulse Buyings’ Affective Antecedents”, in *Research in consumer behavior*, vol. 6. Janeen A-C, Russell WB (eds.), Greenwich (CT): JAI Press, pp. 1-28.
Rossi, P.E., McCulloch, R.E. and Allenby, G.M. 1996. “The Value of Purchase History Data in Target Marketing,” *Marketing Science* (15), pp. 321-340.
Schwarz, G. 1978. “Estimating the dimension of a model,” *Annals of Statistics* (6), pp.461-464.
Sismeiro, C., and Bucklin, R.E. 2004. "Modeling Purchase Behavior at an E-Commerce Web Site: A Task Completion Approach," *Journal of Marketing Research* (41:3), pp.306-323
Srivastava, J., Cooley R., Deshpande, M., and Tan, P.N. 2000. "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," *SIGKDD Explorations* (1:2), pp.12-23.
Stern, H. 1962. “The Significance of Impulse Buying Today,” *Journal of Marketing* (26: April), pp. 59-62.
Tauscher, L., and Greenberg, S. 1997. "How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems," *International Journal of Human Computer Studies, Special issue on World Wide Web Usability* (47:1), pp. 97-138.
Teicher, H. 1963. “Identifiability of Finite Mixtures,” *The Annals of Mathematical Statistics* (34), pp.1265-1269
Titterington, D., Smith, A., and Makov, U. 1985. "Statistical Analysis of Finite Mixture Distributions," New York: John Wiley & Sons.
Venkatesh, V., and Agarwal, R. 2006. “From Visitors to Customers: A Usability-Centric Perspective on Purchase Behavior in Electronic Channel s,” *Management Science* (52:3), pp.367-382.
Viswanathan, S., Kuruzovich, J.N. Gosain, S., and Agarwal, R. 2007. “Online Infomediaries and Price Discrimination: Evidence from the Auto-Retailing Sector,” *Journal of Marketing* (71:3), pp. 98-107.
Vriens, M., Wedel,M. and Wilms, T. 1996. “Metric Conjoint Segmentation Methods: A Monte Carlo Comparison,” *Journal of Marketing Research* (33: February), pp. 73–85.
Wang, P., Puterman, M.L., Cockburn, I. and Le, N.D., 1996, Mixed Poisson regression models with covariates dependent rates, *Biometrics*, 52.

Wedel, M., and Kamakura, W. A. 2000. *Market Segmentation: Conceptual and Methodological Foundations*. Dordrecht: Kluwer.

Winkelmann, R. 2008. *Econometric Analysis of Count Data* (5th ed.), Berlin: Springer-Verlag

Wolfinbarger, M. and Gilly, M. 2001. "Shopping Online for Freedom, Control, and Fun", *California Management Review* (43:2), pp.35–55.

Wooldridge, J. 2009. *"Correlated Random Effects Models with Unbalanced Panels,"* manuscript, Michigan State University

Xia, L., and Monroe, K.B. 2004. "Consumer Information Acquisition: A Review and an Extension," *Review of Marketing Research,* pp. 101-152.

Zettelmeyer, F., Scott Morton, F., and Silva-Risso, J. 2005. "Cowboys or Cowards: Why Are Internet Car Prices Lower?" Discussion paper, Haas School of Business, University of California, Berkeley.

Zhang, J. and Krishnamurthi, L. 2004. "Customizing Promotions in Online Stores". *Marketing Science* (23:4), pp.561-578.

Yang, Y. and Padmanabhan, B. 2007. "Behavioral Signatures in Web Browsing Data," In *Proceedings of the First China Summer Workshop on Information Management,* July 22-23

Yankee Group and Ernst and Young. 2002. "What Causes Customers to Buy on Impulse?," *User Interface Engineering* White Paper. http://www.uie.com/publications/whitepapers/ImpulseBuying.pdf

TABLES AND FIGURES

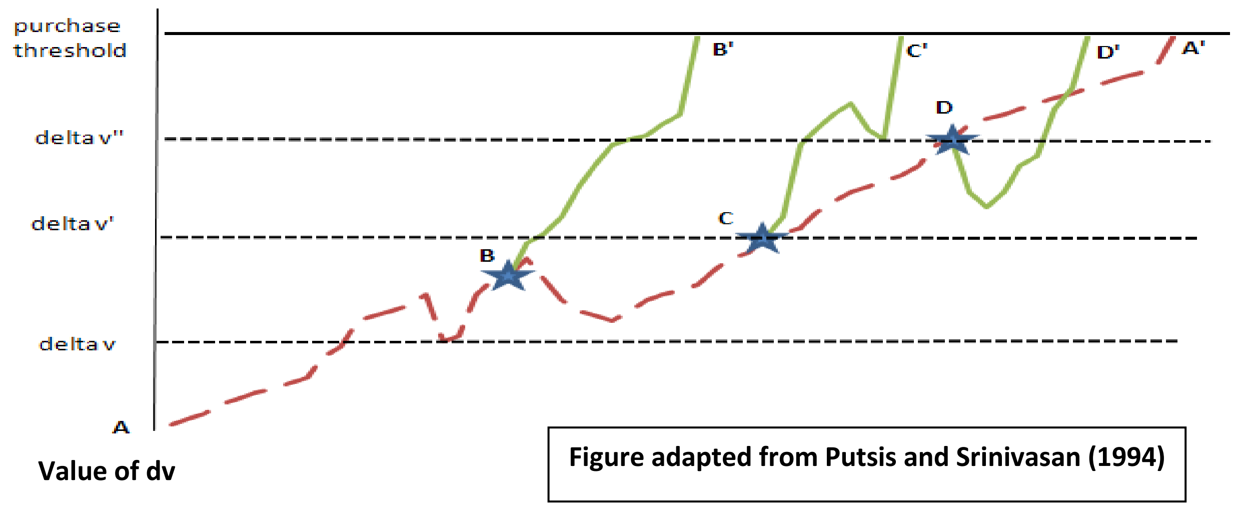

Figure 1.
AN EXAMPLE OF THE DELIBERATION MODEL FOR CUSTOMERS WITH VARYING LEVELS OF DIFFERENCE UTILITY

Notes: In the diagram above, the vertical axis represents the distance between the two indirect utilities for a consumer with and without purchase (called dv or difference utility), and the horizontal axis represents progress of time or the shopping process. The difference utility can change due to both informational and non-informational factors. Consumers only participate in formal pre-purchase search when the difference utility exceeds a certain threshold. However, since consumers in different states of shopping possess different needs, there exist multiple search thresholds that correspond to various states of shopping. These are represented by an arbitrarily chosen number of three thresholds - delta v, v' and v''. The difference between the search and purchase threshold (the solid black line) represents the extent of purchase uncertainties.

The dashed red line is an example of the progression of dv for a baseline consumer who is originally out of the market (dv < delta v), then progresses through three search thresholds, while increasingly lowering her purchase uncertainties until dv turns positive (at the purchase threshold) and she buys at A'. Compare her to three other customers (blue stars) who enter the market in different states of shopping – at points A, B and C. The green solid line depicts how their shopping process progresses in comparison to the baseline consumer. The path BB' is closest to a browser who has the largest difference utility dv at start, and upon obtaining relevant information or stimuli buys at B'. CC' and DD' depict consumers who are closer to the goal-oriented end of the shopping goals spectrum. Any information that increases (decreases) their purchase uncertainty lowers (raises) dv. They continue to search until their difference utility reaches the purchase threshold, and they buy at C' and D' respectively. Notice how appropriate information interventions help consumers B, C ad D to lower their purchase uncertainties faster than the baseline consumer.

Table 1. Measures

Measure	Definition
Purchase	Binary to indicate whether a user completed all three steps for purchase
Purchase_cnt	Count of the number of purchase related steps completed by the user
Cart	Binary to indicate whether a user added a focal product to the shopping cart
ProdInfo	Whether consumer viewed rich product information in the focal product category
GPriceInc	Whether there were any shipping related offers on the day of the session
SPriceInc	Whether there were any category-level discounts on the day of the session
DeptBreadth	number of unique product departments viewed
CatBreadth	number of unique product categories viewed
Depth	extent of hierarchical drilldown within the product category of the focal product
TotalTime	Session time in minutes
TotalPages	Number of pages viewed in session
TotalProducts	Number of product pages viewed in session
ProdPagesPerMin	Ratio of product pages to total time spent in session
RatioProdtoCatPages	Ratio of product pages to category level pages
PriceFacetedSearch	Count of times consumers screened alternatives by price -"under X dollars", and "between X and Y dollars"
ProdFacetedSearch	Count of times consumers screened alternatives using product related attributes such as brands and features
TextSearch	Count of times consumers used a textbox to search/ locate items of interest
CompMatrix	Count of the use of a side-by-side comparison matrix to compare selected products
Date	Calendar day
TimeofDay	Dummies to indicate whether session started in the morning, afternoon, or evening/night
Weekend	Dummy to indicate whether session was conducted on a weekend
MonthofVisit	Calendar month
RepeatVisit	Dummy to indicate whether the session is from a repeat customer
OrdSession	A count of the order of the session for a given cookie.
ProductType	Dummy to indicate that the session user conducted a search for one of the four focal products
Account	Dummy to indicate whether the consumer logged into a user account at the website
UGCReviews	Number of views of user generated content such as reviews and ratings
HomePage	Count of visits to the home page
StorePages	Count of visits to the local store pages
ExternalPages	Count of visits to the pages external to the store, but affiliated to or linked from it
ErrorPages	Count of visits to error pages

Table 2. Means and Correlations

Purchase	0.02 (.14)	1.00																
TotalPages*	2.55 (.85)	0.09*	1.00															
TotalTime*	1.68 (1.10)	0.12*	0.77*	1.00														
TotalProducts*	1.06 (.56)	0.11*	0.43*	0.35*	1.00													
DepthBreadth*	1.19 (.95)	0.04*	0.51*	0.40*	0.49*	1.00												
CatBreadth*	-1.04 (1.26)	-0.01	0.13*	0.07*	-0.11*	0.14*	1.00											
Depth*	0.33 (0.92)	0.00	0.03*	-0.01	-0.33*	0.26*	0.30*	1.00										
RatioProdTocatPages	0.68 (0.73)	0.02*	-0.14*	-0.06*	0.19*	-0.09*	-0.37*	-0.25*	1.00									
ProdpagesPerMin*	-0.72 (1.30)	0.03*	0.28*	0.23*	0.09*	0.18*	0.17*	0.19*	-0.10*	1.00								
PriceFacetedSearch*	-1.96 (1.05)	0.03*	0.27*	0.18*	0.10*	0.17*	0.17*	0.02*	-0.14*	0.05*	1.00							
ProdFacetedSearch*	-2.04 (0.92)	0.04*	0.27*	0.20*	0.13*	0.20*	0.14*	0.02*	-0.12*	0.06*	0.19*	1.00						
ProdInfo	.10 (.29)	0.07*	0.25*	0.27*	0.14*	0.15*	0.01	0.05*	0.00	0.09*	0.06*	0.06*	1.00					
GPriceInc	.50 (.49)	0.05*	0.01	0.00	0.00	0.02	0.00	0.01	0.00	0.01	0.01	0.00	0.02	1.00				
SPriceInc	.28 (.45)	-0.01*	0.13*	0.06*	0.12*	0.03*	-0.17*	-0.04*	0.04*	-0.01*	0.01	0.00	-0.01	-0.63*	1.00			
Past_ProdInfo	.09 (.51)	0.08*	0.05*	0.04*	0.37*	0.02*	-0.05*	-0.07*	0.10*	0.00	-0.01	0.02*	0.03*	0.02*	0.06*	1.00		
Past_GPriceInc	.02 (.17)	0.02*	0.06*	0.07*	0.22*	0.02*	-0.02*	-0.04*	0.05*	0.02*	0.02*	0.03*	0.17*	0.08*	-0.03*	0.40*	1.00	
Past_SPriceInc	.05 (.32)	0.05*	-0.03*	-0.03*	0.29*	-0.02*	-0.03*	-0.04*	0.06*	-0.01*	-0.01*	0.01	0.00	0.16*	-0.10*	0.43*	0.28*	1.00

Note * - indicates variables that are logged

Table 3. Examining fit across multi-component models

# components	LL	AIC	AIC3	BIC
1	-8221.305	16466.61	16478.61	16569.99
2	-8135.561	16321.12	16346.12	16536.49
3	-8044.895	16165.79	16203.79	16493.15
4	-8032.992	16167.98	16218.98	16607.34

Table 4a. BDI measures used to characterize latent states of shopping

	State 1 N=5833 (14.32%)		**State 2 N=17580 (43.18%)**		**State 3 N=17323(42.53%)**	
Variable	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
DeptBreadth	0.337	0.600	0.415	0.602	0.869	0.807
CatBreadth	1.526	2.949	2.448	4.327	4.369	5.688
Depth	2.198	1.205	3.330	1.142	2.094	0.871
TotalPages	30.222	34.326	15.050	14.756	18.692	18.069
TotalTime	18.435	18.019	9.258	11.884	7.093	8.964
TotalProducts	4.614	6.283	8.933	3.973	3.180	3.714
RatioProdtoCatPages	0.794	0.627	0.915	0.962	0.405	0.201
ProdPagesPerMin	2.118	1.730	2.100	1.621	3.179	1.996

Table 4b. Predicting the posterior probability of state using BDI measures

BDI	**(1) State 1 (vs. State 2)**		**(2) State 3 (vs. State 2)**	
	β	**s.e.**	***β***	**s.e.**
DeptBreadth	0.531***	.024	0.341***	0.018
CatBreadth	-0.329***	0.015	0.587***	0.010
Depth	-0.273***	0.022	-0.032+	0.017
TotalPages	0.017***	0.001	0.028***	0.001
TotalTime	0.849***	0.022	-0.656***	0.016
TotalProducts	-1.094***	0.048	-0.107***	0.030
ProdPagesPerMin	0.176***	0.013	0.086***	0.010
Intercept	-2.987***	0.061	0.888***	0.035

Note: RatioProdtoCatPages is dropped due to multicollinearity

Table 5a. Validating the latent states of shopping

	State 1 N=5833 (14.32%)		State 2 N=17580 (43.18%)		State 3 N=17323(42.53%)	
Variable	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
Buy	0.051	0.221	0.018	0.133	0.013	0.113
Cart	0.145	0.352	0.105	0.306	0.064	0.244
Conditional Buy	0.354	0.478	0.172	0.378	0.202	0.402
Repeat session	0.153	0.360	0.110	0.312	0.074	0.262
PriceFacetedSearch	0.579	2.843	0.250	1.413	0.555	2.135
ProdFacetedSearch	0.570	2.879	0.172	1.148	0.371	1.756
TextSearch	1.524	4.683	0.853	2.789	0.333	1.751
CompMatrix	0.375	1.604	0.207	1.374	0.131	0.888
ProdInfo	0.160	0.366	0.098	0.297	0.071	0.257
SPriceInc	0.310	0.462	0.284	0.451	0.263	0.440
GPriceInc	0.507	0.500	0.504	0.500	0.503	0.500

Table 5b. Effects of information from Poisson mixture modeling (3 components)

Information/incentives obtained	(1) β	(1) s.e.	(2) β	(2) s.e.
Directed shopper				
ProdInfo	-0.524***	0.059	-0.115+	0.064
SPriceInc	1.157***	0.102	0.721***	0.108
GPriceInc	1.063***	0.106	0.923***	0.103
Deliberating researcher				
ProdInfo	2.205***	0.042	1.624***	0.046
SPriceInc	-0.107	0.082	-0.096	0.085
GPriceInc	1.035***	0.067	0.946***	0.069
Browsers				
ProdInfo	1.664***	0.062	0.878***	0.066
SPriceInc	2.407***	0.178	1.732***	0.181
GPriceInc	2.414***	0.175	2.058***	0.176

Note: The dependent variable is Purchase_cnt. We estimate finite mixture models with response to information/incentives and BDI as the classification bases in model (1) and response to information as classification basis and BDI variables to characterizing the probability of belonging to a state in model (2)
*** p<0.001, **p<0.01, *p<0.05, + p<0.1

Table 6. Estimating the within-session and across-session impacts of online information and incentives on completing a purchase

	(1)		(2)		(3)		(4)	
	β	s.e.	β	s.e.	β	s.e.	β	s.e.
LatState_DR	-0.566***	0.055	-0.280***	0.057	-0.540***	0.081	-0.126	0.147
LatState_BR	-0.612***	0.062	-0.491***	0.066	-0.501+	0.087	-1.025***	0.225
PastLatState_DS	-0.349***	0.107	-0.333**	0.108	-0.340**	0.109	-0.323**	0.109
PastLatState_DR	0.196*	0.084	0.207*	0.084	0.199*	0.084	0.212*	0.085
PastLatState_BR	0.203*	0.089	0.205*	0.090	0.195*	0.089	0.225*	0.090
PastProdInfo	0.177***	0.040	0.184***	0.039	0.180***	0.039	0.180***	0.040
PastSPriceInfo	-0.392***	0.095	-0.408***	0.095	-0.402***	0.095	-0.401***	0.095
PastGPriceInfo	-0.085+	0.048	-0.115*	0.048	-0.094*	0.047	-0.115*	0.048
PastCart	-0.217***	0.059	-0.235***	0.059	-0.220***	0.059	-0.235***	0.060
ProdInfo	-0.344***	0.102	0.204***	0.057	0.201***	0.057	-0.352***	0.103
SPriceInc	0.271**	0.099	0.440***	0.114	0.250*	0.098	0.753***	0.148
GPriceInc	0.672***	0.156	0.716***	0.158	0.604***	0.169	0.802***	0.197
ProdInfo*DR	0.792***	0.119					0.821***	0.120
ProdInfo*BR	0.664***	0.134					0.672***	0.135
SPriceInc*DR			-0.650***	0.118			-1.012***	0.179
SPriceInc*BR			0.053	0.108			0.433+	0.238
GPriceInc*DR					0.202*	0.098	-0.365*	0.155
GPriceInc*BR					0.033	0.102	0.440+	0.229
CatBreadth	0.003	0.018	0.002	0.018	0.001	0.018	0.005	0.018
DepthBreadth	-0.161***	0.025	-0.169***	0.025	-0.166***	0.025	-0.166***	0.025
Depth	0.150***	0.026	0.158***	0.026	0.157***	0.026	0.150***	0.026
TotalPages	0.104*	0.042	0.105*	0.042	0.101*	0.042	0.110**	0.042
TotalPages^2	0.000	0.000	0.000+	0.000	0.000	0.000	0.000	0.000
TotalTime	0.287***	0.031	0.296***	0.031	0.300***	0.030	0.279***	0.031
TotalTime^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TotalProducts	0.307***	0.049	0.324***	0.049	0.313***	0.048	0.324***	0.049
ProdPagesPerMin	-0.054	0.040	-0.055	0.040	-0.056	0.040	-0.052	0.040
OrdSession	0.517***	0.071	0.517***	0.070	0.509***	0.070	0.527***	0.071
OrdSession^2	-0.048***	0.009	-0.046***	0.009	-0.047***	0.009	-0.047***	0.009
PriceFacetedSearch	-0.006	0.016	-0.008	0.016	-0.006	0.016	-0.008	0.016
ProdFacetedSearch	0.019	0.017	0.019	0.017	0.021	0.017	0.016	0.017
TextSearch	0.016	0.013	0.013	0.013	0.013	0.013	0.016	0.014
CompMatrix	0.069***	0.021	0.066**	0.021	0.066**	0.021	0.069***	0.021
HomePage	0.035	0.039	0.033	0.039	0.035	0.039	0.031	0.039
StorePages	0.071***	0.020	0.064***	0.020	0.063***	0.020	0.073***	0.020
ExternalPages	0.104	0.065	0.105	0.064	0.100	0.065	0.113+	0.065
UGCReviews	0.304**	0.107	0.311**	0.107	0.315**	0.106	0.299**	0.107
AccountPages	0.174***	0.032	0.176***	0.032	0.174***	0.032	0.177***	0.032
ErrorPages	-0.195***	0.053	-0.185***	0.053	-0.186***	0.053	-0.197***	0.053
Intercept	-2.542***	0.388	-2.748***	0.389	-2.605***	0.388	-2.725***	0.404
σ_α	0.518	0.031	0.517	0.031	0.518	0.031	0.517	0.031
ρ	0.212	0.020	0.211	0.020	0.212	0.020	0.211	0.020

Note: The dependent variable across all models is Purchase. We estimate cookie-panel models with cluster robust standard errors. Col (1)-(4) use the full sample, while col (5) uses the interested sample.
*** p<0.001, **p<0.01, *p<0.05, + p<0.1

Table 7. Impacts of information and incentives obtained within a session on purchase

Information/ Incentives obtained	(1) Full sample		(2) Interested sample		(3) Add to cart Sample	
	β	s.e.	β	s.e.	β	s.e.
PAST SESSIONS						
Past_ProdInfo	0.177***	0.040	0.175***	0.043	0.108*	0.043
Past_SPriceInc	-0.392***	0.095	-0.371***	0.099	-0.091	0.109
Past_GPriceInc	-0.085+	0.048	-0.128*	0.051	-0.233***	0.064
Directed shopper						
ProdInfo	-0.352***	0.103	-0.465***	0.122	-0.486**	0.186
SPriceInc	0.753***	0.148	0.997***	0.179	0.756+	0.408
GPriceInc	0.802***	0.197	0.932***	0.243	1.372*	0.591
Deliberating researcher						
ProdInfo	0.469***	0.077	0.402***	0.091	0.839***	0.157
SPriceInc	-0.259*	0.127	-0.285+	0.149	-0.177+	0.100
GPriceInc	0.437*	0.169	0.398+	0.214	0.800	0.552
Browsers						
ProdInfo	0.319***	0.098	0.274*	0.115	0.489*	0.226
SPriceInc	1.187***	0.203	1.495***	0.236	1.204**	0.454
GPriceInc	1.242***	0.238	1.471***	0.286	1.645*	0.655

Note: The dependent variable is Purchase. We estimate cookie-panel models with cluster robust standard errors. (1) is our main model. (1) uses the full sample, (2) uses the interested sample, and (3) uses a restricted sample of user who add a focal product to the shopping cart during some session.
*** p<0.001, **p<0.01, *p<0.05, + p<0.1

Table 8. The impact of product information using matching techniques

	N treated	N control	Average treatment effect	s.e.	T-stat
Radius matching					
All	3883	2389	0.026	0.004	7.094
Directed shopper	932	502	-0.014	0.009	-1.687
Deliberating researcher	1724	1085	0.041	0.006	6.950
Browser	1227	782	0.021	0.006	3.816
Stratified matching					
All	3883	36853	0.032	0.004	8.559
Directed shopper	932	4506	-0.025	0.014	-1.796
Deliberating researcher	1724	14525	0.048	0.006	8.234
Browser	1227	15921	0.026	0.005	4.828

Table 9. Robustness Checks

	(1)		(2)		(3)		(4)		(5)	
Dependent Variable	**Purchase**		**Purchase**		**Purchase_cnt (Poisson)**		**Purchase_cnt (Ordered probit)**		**Add to cart**	
Robustness to	**Price- product sensitivity**		**Correlated random effects**		**Alternate DV**		**Alternate DV**		**Alternate DV**	
	β	**s.e.**	***β***	**s.e.**	***β***	**s.e.**	***β***	**s.e.**	***β***	**s.e.**
PAST SESSIONS										
Past_ProdInfo	0.179***	0.040	0.100**	0.036	0.255***	0.068	0.055*	0.024	0.028	0.035
Past_SPriceInc	-0.391***	0.095	-0.451***	0.096	-0.330***	0.091	-0.166**	0.057	-0.151*	0.067
Past_GPriceInc	-0.122*	0.048	-0.170***	0.048	-0.242***	0.056	-0.102**	0.036	-0.152***	0.042
WITHIN SESSION: Directed shopper										
ProdInfo	-0.306*	0.127	-0.341**	0.103	-0.512**	0.174	-0.015	0.076	-0.258**	0.098
SPriceInc	0.559**	0.178	0.742***	0.147	1.266***	0.290	0.446***	0.122	0.664**	0.211
GPriceInc	0.301**	0.108	0.853***	0.212	1.265***	0.355	0.166	0.150	0.427**	0.139
WITHIN SESSION: Deliberating researcher										
ProdInfo	0.684**	0.051	0.475***	0.077	1.368***	0.120	0.811***	0.051	0.639***	0.069
SPriceInc	-0.456**	0.168	-0.233*	0.099	-0.254*	0.099	-0.402***	0.086	0.340+	0.190
GPriceInc	-0.092	0.193	0.504**	0.187	0.479+	0.288	0.073	0.124	0.140	0.097
WITHIN SESSION: Browser										
ProdInfo	0.080	0.069	0.318**	0.097	1.272***	0.174	0.427***	0.071	0.017	0.091
SPriceInc	1.043***	0.222	1.199***	0.200	2.444***	0.276	0.564***	0.121	0.939***	0.204
GPriceInc	0.812**	0.260	1.300***	0.251	1.919**	0.343	0.290*	0.147	0.851***	0.126

Note: We estimate cookie-panel models with cluster robust standard errors using the full sample. In (1) and (2) the dependent variable is a binary indicating whether the customer purchased a focal product during the session. All the models contain the full set of covariates from table 6. Additionally, in model (1), we include the interactions between the two types of faceted search and their interactions with information/incentives. In model (2), we add seven time/group means of the BDI measures to test correlated effects. In models (3) and (4) the dependent variable is Purchase_cnt estimated using panel poisson and ordered probit, respectively. In model (5) the dependent variable is a binary indicating wherger the consumers added a focal product to the shopping Cart.
*** p<0.001, **p<0.01, *p<0.05, + p<0.1

APPENDIX A. STATE TRANSITIONS

In this appendix, we assess how consumers with multiple visits to the e-tailer transitioned through the three latent states of shopping uncovered by the mixture model. As seen in Table A1 below, we observe a high level of inertia for directed buyers (66%) and information gatherers (63%), where consumers are likely to continue in the same state. For browsers, the likelihood of returning as a browser is close to 50% and as a deliberating researcher is 41%.

Table A1. State transitions for repeat visitors (excluding last session)

	Directed shopper	Deliberating researcher	Browser	Total
Directed shopper	**65.93%**	21.79%	12.28%	904
Deliberating researcher	9.60%	**63.34%**	27.06%	1855
Browser	8.86%	41.25%	**49.88%**	1343
Total	21.77%	46.95%	31.28%	4102

Next, we examine further details to help understand the impacts of transitioning between latent states of shopping on purchase outcomes in Table A2. We find that among directed shoppers that do return, the likelihood to complete the purchase drops sharply. If they return as researchers, this conversion rate is 4.57% and decreases to 2.70% when they return as browsers. On the other hand, when consumers transition into the directed state of shopping from the other two states the results show increased conversion. In sessions where consumers transition from deliberation to directed shopping, the purchase likelihood jumps to 34.83% and for sessions where consumers proceed to directed buying after browsing, this number improves to 29.41%. Overall, for deliberators and browsers we find that transitions to a directed shopper had the highest likelihood of conversion, followed by transitioning to deliberating and last, browsing. These results merely indicate correlations. We also refrain from modeling the impact of information on state transitions because consumers may have very well interspersed offline store visits with online visits to the click-and-mortar retailer.

Table A2. The conditional effects of state and state transitions on completing a purchase

Previous state → **Current State**	**Directed shopper**	**Deliberating researcher**	**Browser**
Directed shopper	3.52%	4.57%	2.70%
Deliberating researcher	34.83%	2.72%	4.18%
Browser	29.41%	4.33%	3.88%
Overall	5.12%	1.80%	1.29%

APPENDIX B. PROPENSITY SCORE MATCHING

The literature on treatment effects defines the treatment effect of a binary treatment as the difference in outcome when units (here sessions) are treated (receive $ProdInfo$) and when those same units are not treated. However, we only observe sessions in either the treated or the non-treated condition, and therefore must construct the necessary missing counterfactuals for the sessions. Propensity score matching allows us to estimate average treatment effects by comparing the outcomes of treated and control groups that have been matched on the breadth, depth and intensity covariates instrumental in determining the likelihood of receiving treatment[9]. We construct a stratified or matched sample of observations that consists of treated and control groups that are balanced across these observed covariates – and therefore, on average observationally identical. The propensity score is the conditional probability of receiving the treatment rather than being part of the control group given the relevant observed covariates W (Rosenbaum & Rubin, 1983). It is estimated using a probit model as follows where the treatment is $ProdInfo = 1$ and W contains variables that describe breadth, depth and intensity of search.

$P(ProdInfo = 1|W) = \Phi\{h(W)\}$ where Φ is the normal c.d.f. (B.1)

Matching on such a score serves to simulate random assignment of treatment when two conditions hold: a) the observed covariates used to construct the score are balanced, and b) there is no bias from unobserved covariates. We check that condition a) holds, and we restrict the matching to be performed over the common support region – that is using observations whose propensity scores belongs to the intersection of the supports of the propensity scores of the treated and control sessions. Condition b) is the Conditional Independence or Unconfoundedness Assumption that treatment assignment is ignorable (independent of the potential binary outcomes for purchase Y(0) or not Y(1) in a session) conditional on observed covariates - a critical assumption in matching models (Abadie and Imbens 2002).

$$P\left(ProdInfo = 1 \middle| W, Y(0), Y(1)\right) = P(ProdInfo|W) \quad \text{(B.2)}$$

Identification is achieved when the probability of assignment of treatment is bounded away from zero and one, known as the Overlap assumption (Abadie and Imbens 2002): $0 < P(ProdInfo = 1|W) < 1$

When these regularity conditions hold, then imbalances in pretreatment covariate levels can be controlled by adjusting the unidimensional propensity score calculated in (12) such that comparisons of outcomes occur between treated and control groups that differ only in their exposure to treatment (Rosenbaum and Rubin 1983). The treatment effect for an individual is

$$TE_i = Purchase_i\,(ProdInfo_i = 1) - Purchase_i\,(ProdInfo_i = 0) \quad \text{(B.3)}$$

[9] Additionally, W is chosen to satisfy the Balancing Hypothesis of matching estimators.

Then aggregate impact of product information on outcomes is calculated as the sample average treatment effect on the treated ($SATT$) given by

$$SATT = \frac{1}{n_T} \sum_{i \in T} TE_i \qquad \text{(B.4)}$$

Where $n_T = \sum_{i=1}^{n} T_i$ is the number of *treated* units with observed treatment $ProdInfo_i = 1$.

An important concern in using propensity score matching methods to estimate treatment effects is the potential violation of condition b) above. While the model accounts for selection on observables, consumers' choice to visit online product information pages such as buying guides is likely to covary with important unobservables in the study. The main results in the paper suggest that the varied effects of several relevant variables (breadth, depth and intensity of search and navigation behaviors) on the purchase likelihood are summarily captured in the latent state of shopping. If the latent state simultaneously affects assignment into treatment and the outcome variable, a *hidden bias* might arise to which matching estimators are not robust (Rosenbaum 2002). Additionally, given our interest in separately identifying the effect of (product) information on outcomes across consumer sessions belonging to different latent states, we construct propensity score matching estimates for each latent group separately, in effect using a latent state dummy as a matching covariate in addition to W. This provides us with one way, albeit imperfect, in which to account for unobservables.

REFERENCES

Rosenbaum, P., and Rubin, D. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* (70), pp.41–55.

Abadie, A. and Imbens, G. 2002. "Simple and Bias-Corrected Matching Estimators for Average Treatment Effects", *NBER Technical Working Papers*.

Rosenbaum, P.R. 2002. *Observational Studies*. 2nd ed. New York, NY: Springer.

APPENDIX C. JOINT ESTIMATION OF PURCHASE AND CART OUTCOMES

We jointly assess the effects of covariates on $Purchase$ and $Cart$ using a bivarite model. The mean correlation between the standard errors across the two outcomes is estimated to be 0.9935 ($\chi^2(1) = 304.98, p = 0.00$) - this high number indicates that there is high level of similarity in the unobservables that affect a consumer's decision on both outcomes. We calculate the predicted probabilities after controlling for several relevant covariates (as used in Table 6). At a joint predicted probability $\Pr(Purchase_i = 1, Cart_i = 1)$ of 4.44%, the conversion rate is the highest for directed shoppers followed by deliberating researchers (1.73%) and browsers (1.30%). The groups were ranked in the same order for the marginal predicted probabilities of both outcomes - adding to the shopping cart and completing the purchase. The conditional probability $\Pr(Purchase_i = 1 \mid Cart_i = 1)$ tells a slightly different story. Conditional on having added products to the shopping cart, directed shoppers had the highest rate of conversion (35.35%), while deliberators had the lowest (18.80%). This suggests that consumers across the different states perhaps use the cart for different reasons. The lower conditional rate of conversion of shopping carts for researchers underscores the importance of recognizing that some consumers may not be ready to purchase in the current session even if they add products to their cart. They may use the cart to conveniently hold and compare chosen alternatives. These results shed some light on shopping cart abandonment- a common woe of online retailers (cf., Murthi and Sarkar 2003) - across the three states.