Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2010 Proceedings

International Conference on Information Systems (ICIS)

2010

STATE OF SHOPPING AND THE VALUE OF INFORMATION: INSIGHTS FROM THE CLICKSTREAM

Vandana Ramachandran
University of Utah, vandana@business.utah.edu

Siva Viswanathan
University of Maryland, sviswana@rhsmith.umd.edu

Hank Lucas
University of Maryland, hlucas@rhsmith.umd.edu

STATE OF SHOPPING AND THE VALUE OF INFORMATION:

INSIGHTS FROM THE CLICKSTREAM

Completed Research Paper

Vandana Ramachandran
Operations and Information Systems
University of Utah
vandana@business.utah.edu

Siva Viswanathan
Decision, Operations, and Information Technologies
University of Maryland
sviswana@rhsmith.umd.edu

Hank Lucas
Decision, Operations, and Information Technologies
University of Maryland
hlucas@rhsmith.umd.edu

Abstract

A critical challenge for online retailers is to determine what types of product and price information are best suited to influence online conversion. While it has long been known that customers differ in their state of shopping, it is cumbersome to learn about such latent differences offline. The availability of clickstream data however helps us in identifying meaningful segments of sessions on the basis of customers' online behaviors. We examine whether product and price information had different impacts on customers belonging to three states of shopping, and also assess the effect on outcomes within a session and across sessions. Our results question the practice of offering price promotions to all customers of a store, and highlight the value of product information in increasing loyalty for some customers. Depending on the retailer's goal–short term conversion versus longer-term customer relationship–a different information provision strategy is likely to be optimal.

Keywords: clickstream, states of shopping, conversion, price information, product information

Introduction

Provision of targeted information is ubiquitous on the Internet today, and exists in myriad forms across search engines, online social networks, blogs, and various content sites. Much anecdotal evidence points to the positive effects of targeting done correctly- satisfied users and improved conversion rates. In this study, we extend the notion of targeting to *product and price-related information* that retailers can present in *real time* to customers who are *actively visiting* their online store. This micro-level approach has the potential to be highly interactive and complementary to targeting strategies used to attract consumers to e-tailer stores. Real-time targeting requires retailers to present custom information that matches their customers' needs and preferences, which are in turn driven by their shopping state during a session. Consumer's state of shopping is however unobserved, requiring retailers to make inferences on the basis of observed behavior patterns of consumers. A commonly used source of information about consumers in traditional markets, especially for frequently purchased products (such as grocery), is purchase history. 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 especially pronounced for online retailers of durable goods, who face an interesting scenario. On the one hand, these retailers know very little about the customers that visit their (online) store due to the lack of identifiable historical interactions. At the same time, given the nature of the purchase involving big-ticket and high-involvement goods, customers are more likely to conduct extensive pre-purchase search and place greater value on appropriately targeted information that improves the utility of their purchase (Mack 2009; PriceGrabber Consumer Behavior Report 2009). As a result of limited interactions with and a slim dossier on each customer, durable good retailers must seek alternate ways to learn about their customers' needs and preferences in order to target information to them. We explore one source of rich micro-level real-time knowledge about customers- the store-level clickstream.

The goal of our research is twofold. First, we seek to meaningfully characterize consumers' search and navigation behaviors within a session obtained from clickstream data in ways that reflect shopping-relevant underlying differences. We refer to these differences as *states of shopping*. Second, we examine whether product and price-related online information had different impacts on conversion for customers belonging to various *states of shopping*, and whether information varied in its impacts on purchase related behaviors *within a session* (complete purchase) and *across sessions* (return visit and future purchase).

Consumers search online to learn about the assortment of available products, brands and prices across firms. Even when consumers don't buy within a session, they take away useful knowledge about available alternatives – information that is likely to significantly influence their preferences and purchase outcomes later (Mandel and Johnson 2002; PriceGrabber Consumer Behavior Report 2009). Furthermore, several studies report that retailers' web sites trump manufacturer sites and search engines as the information sources cited by consumers as most frequently used for conducting product research online (Compete Online Shopper Intelligence Study 2010; iCrossing Report 2010). It is therefore crucial for firms to both understand the shopping-related needs conveyed by consumers' online search, and act on it by providing the right types of information at the appropriate times to the customer. Fortunately, with the growth of clickstream technologies, there has been a phenomenal improvement in our ability to understand customers. Clickstream data offers the ability to analyze not just the purchase occasion alone, but also the sequence of events that lead to desirable outcomes within a website (Montgomery et al. 2004). Consumers' clicks provide retailers with fine-grained insight ranging from their relative level of interest across categories and their consideration sets, to the types of information accessed and their purchase-related outcomes.

While the retailer can simply provide a variety of purchase-related information, and allow consumers to pick and choose, this is often suboptimal because of concerns involving information overload. For a shopper, a cluttered screen is often a challenge to navigate; and increases the probability that consumers may overlook important pieces of information. On the contrary, consumers may be more receptive to and better served by information that is well targeted to their shopping needs. Prior research has found that a large part of consumers' pre-purchase search activity involves seeking both *price* and *product* information (Diehl et al. 2003; Klein and Ford 2003; Lynch and Ariely 2000). Price information such as free shipping and sales or discounts are no doubt important in helping consumers consummate their purchase; but promotions are frequently margin-eroding (e.g., Gelb et al. 2007). Moreover, it is unclear whether they help build a loyal customer base that will continue to return to the store, or whether consumers that respond to price information are price-sensitive shoppers who continue to seek the lowest price across retailers. If the latter was true, focusing on pricing alone to attract and convert customers may not be the best long term strategy. As online shopping matures consumers are also increasingly seeking to research and understand the available product assortments better. Retailers, in turn, are responding by investing costly dollars in providing a rich multimedia experience consisting of visual merchandising, product configurators, and buying guides for their online customers in an attempt to differentiate from other retailers (Tedeschi 2006). Whether product-related information can turn casual visitors into repeat visitors, and increase their propensity to purchase remains untested. If it did, what types of customers are most likely to benefit from the availability of rich product information? We lack understanding of when consumers value price-related information more than product-related information, and vice versa. Relatedly, are there certain shopping states when product (rather than price) information would help move consumers further along the shopping process and closer to conversion? We explore answers to these questions.

We develop cookie-level panel models to describe and assess the impacts of online information on purchasing both within and across sessions for consumers belonging to different latent states of shopping defined on the basis of their observed session-level behaviors. Our model allows consumers to belong to different states across sessions, and can therefore capture state transitions through time for a subset of consumers with repeat visits. We estimate our models using clickstream data from a leading click-and-mortar retailer in the U.S. market that covers visits from 77,574 customers to four best-selling durable products carried by the retailer in late 2006. We find that a three-state model comprised of *directed shoppers, deliberating researchers* and *browsers*, best describes the latent shopping-relevant differences across customer sessions in our data. We then uncover important differences in the effects of information across the three types of sessions. When examining the impacts on purchase conversion *within a session*, product information had the strongest impact for deliberating researchers, while price information about a category-level discount proved useful for both directed shoppers and browsers. Price information related to site-wide free shipping had a positive impact across a broad set of sessions, highlighting the value placed on free shipping by consumers who shop online. More surprising were the two negative effects of information that led online customers to delay a purchase or abandon a session. We found that discounts or sales that apply to all products in a given category had a negative effect on deliberating researchers, while rich product information that highlights various features of alternatives in a category hampered the purchase process of directed shoppers.

Our next set of findings highlight important tradeoffs in the effects of product and price information on *within-session conversion* versus two *across session outcomes* - future purchases and the likelihood of repeat visits. Whereas price-related information had positive impacts on within-session conversion for a larger set of sessions, both types of price information negatively influenced purchase for returning visitors. When online customers did not purchase upon receiving price or promotion information, they were in general less likely to purchase in future sessions if they returned. Additionally, we observe that price information had contrasting effects on customers' within-session conversion and inclination to revisit the store. In contrast, product information positively influenced a smaller set of customers to convert within a session, but had a strong impact on across-session purchase behaviors, influencing consumers to both revisit and purchase in later sessions across all three states of shopping.

Next, we survey existing research and discuss our conceptual framework. Then we describe our data and develop our empirical strategy for uncovering latent states of shopping. We use the findings about latent states to augment a cookie-panel model that allows us to examine the influence of information within and across sessions. Finally, we examine the robustness of our main results and conclude with a discussion of the findings.

Background and Conceptual Framework

We draw from two main streams of literature. The first focuses on characterizing consumers' clickstream as a new source of insight into their shopping needs and intentions, and the second focuses on understanding how different types of information affect consumers' purchase-related outcomes.

In the first stream, a large body of existing work spanning computer science, information systems and marketing has been devoted to studying consumers' search and navigation behaviors in online channels, and broadly suggest that search paths and patterns can predict outcomes. Scholars in CS and IS in the domain of web usage mining have developed descriptive measures to characterize search into meaningful or 'interesting' patterns (Canter et al. 1985; Catledge and Pitkow 1995; Tauscher and Greenberg 1997; Yang and Padmanabhan 2007; among others) and learn their associations with desired outcomes (Cooley et al.1999; Srivastava et al. 2000). Recent work also incorporates user intention (Jin et al.2004) and contextual information (Adomavicius et al. 2005; Palmisano et al. 2007) into the study of user search paths. Other studies have used paths within the context of e-commerce to construct micro-conversion metrics based on look-to-click rate, click-to-basket rate, and basket-to-buy rate (e.g., Gomory et al. 1999) and to compare the navigation patterns of customers to those of non-customers (e.g., Spiliopoulou et al. 1999). In marketing research, scholars have examined paths taken by consumers across websites (e.g., Johnson et al. 2004; Park and Fader 2004); while others 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). Some have examined search within a session (e.g., Moe 2003; Bucklin and Sismeiro 2003) 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 seen by consumers - which is likely to have influenced a large proportion of outcomes.

Studies in the second stream have examined the impacts of product- and price-related information obtained in online channels, and find that both aid in the reduction of uncertainty or costs associated with purchase (Diehl et al. 2003; Klein and Ford 2003; Lynch and Ariely 2000). Existing works however typically do not distinguish the impacts of information across different types of consumers (e.g., Hodkinson and Keil 2003; Ratchford et al.2003; Viswanathan et al. 2007; Zettelmeyer et al. 2005). It is possible that certain types of consumers benefit more from product information than price information and vice versa. Further, while much of the existing work has examined the final purchase outcome, it is useful to understand whether and how information impacts other related shopping behaviors such as and adding products to the shopping cart and returning to visit the store.

Our research extends these streams of work to understand how clickstream behaviors can be used to characterize customer sessions that differ in their needs or state of shopping, and their response to product vs. price information. In this vein, our work is closest in spirit to a limited number of existing works that study consumers' responses to online marketing communications and/or prescribe strategies to optimally target messages to individual customers or segments. 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 past purchase history. Chatterjee et al. (2003) model consumers' likelihood of clicking on web-based advertisements using clickstream data, but do not examine whether clicks led to purchase outcomes. Manchanda et al. (2006) study the conditional effect of banner advertising on purchasing behavior only for those consumer visits to the site that resulted in a purchase. But they do not observe the content of the ads and thus cannot distinguish between the impacts of product and price information. In contrast to these existing works, we use a rich clickstream dataset to distinguish among different types of product and price information made available by retailers, and identify the exposure to information in an individual session. We then examine the impacts of information by combining techniques from the clickstream modeling of user paths and econometric modeling that allows us to account for both session-level and cookie-level heterogeneity in unobservables.. Next, we describe techniques to identify consumers that differ in their shopping state or orientation, and three types of product and price-related information provided by online retailers.

Consumers' Latent States of Shopping

It is now well understood that not all shoppers are in the same state or mindset when shopping for products, and these underlying differences are known to be reflected in their offline search behaviors (Cox 1967; Olshavsky 1985; Putsis and Srinivasan 1994; Payne et al. 1993). Such variances are likely to translate into the online market as well. Clickstream data, in particular, are composed of navigation trails from a diverse set of customers, who have varying purchasing needs and goals (Bucklin and Sismeiro 2003; Moe and Fader 2004). Treating consumers as homogenous may thus be erroneous. Hoffman and Novak (1996) and Dholakia and Bagozzi (2001) divided online customers' search processes into goal-directed and experiential, but did not discuss the effects of the mind-sets on purchase outcomes. Building on this stream, Moe (2003) divided online customers into four categories based on customers' search behaviors (directed versus exploratory) and purchasing horizon (immediate versus future). She categorized consumers as buying, deliberate-searching, browsing, and knowledge-building. Following existing works, we expect that online customers will fall within a continuum of latent states extending from *exploratory* browsers to *directed* shoppers. Exploratory browsers are undirected, less-deliberate and stimulus-driven (Janiszewski 1998). This type of search may not necessarily be motivated by a specific goal, and consumers derive utility not from the outcomes of search, rather, from the process of searching and visiting a site. Experiential behaviors are often part of a consumer's ongoing search process (Hoffman and Novak 1996; Wolfinbarger and Gilly 2001). By contrast, directed searchers are focused in their search and are driven by a goal (Janiszewski 1998). Consumers who conduct directed search obtain utility by clicking and traversing through paths that allow them to gather information related to a product of interest or an impending purchase (Childers et al. 2001; Titus and Everett 1995; Hoffman and Novak 1996).

Online information

While information in online markets may assume a variety of forms, we focus on information that is directly provided by a firm to active visitors at its online store and whose content and availability is under the control of the firm[1]. In this study, we specifically consider three types of information- category-specific product information, category-specific price information and generic price information. *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). Superior product knowledge may help consumers to lower their uncertainty and increase the attractiveness of products in that category. In online markets, such non-price information may include the use of rich multimedia and microsites that provide product configurators, buying and comparison guides, and video/audio demonstrations of features.

We identify two types of price information. C*ategory-specific price information* offers consumers price incentives to purchase products from select product categories (such as "Huge savings on home furnishing-10% off", "Special values on refrigerators"). *Generic price information* includes a reduction or discount that may be applied to any purchase at the website such as offers on shipping and delivery (such as "free shipping on orders over $X", or "free shipping today"). From the retailer's point of view category-level product and category-specific price information increase the attractiveness of all products in a category whereas generic price information increases the utility for any product in the web store. Examining the impacts of information at this level is consistent with our interest in purchase incidence rather than brand choice. Though not comprehensive, our categorization of online information is a useful starting point for teasing apart the effects of information on different types of customers.

Discussion

We use a data-driven approach to empirically determine the optimal number of states of shopping. Aside from resulting in differences in observed search and navigation behaviors, membership in various latent states is likely to differentially impact the likelihood of purchase (Moe 2003). In the traditional channel, researchers have described the existence of a purchase funnel that consists of a sequence of increasingly directed or focused stages that consumers progress through when making purchase related decisions (see Lee and Ariely's (2006) shopping goals theory; and Trope and Lieberman's (2003) construal level theory among others). We expect to find that on average customers who are browsing will have a lower baseline purchase propensity than customers who are closer to the directed shopper end of the state spectrum and often further ahead in the planned purchase process. Controlling for their baseline purchase propensities, we are interested in the impacts of online price and product information.

Directed buyers have typically completed their research, and are closer to finalizing their purchase. They shop around retailers and price comparison websites to determine the locations of acceptable low prices (Wolfinbarger and Gilly 2001). We expect that they will benefit most from the availability of price than product information because it offers the best utility for their already selected product(s). A sale in a category of interest or a shipping offer can be extremely successful in incentivizing directed consumers to complete the purchase, and preventing them from delaying or abandoning the site in search of better deals elsewhere. At the other extreme are consumers who are browsing several product departments or categories, often, having started the session without a particular product purchase in mind. Some subset of browsers may also be seeking knowledge about a category that they are interested in but perhaps not considering making a near-time purchase. Thus we may observe unplanned or impulse purchases from this group when they obtain information that renders a purchase sufficiently attractive. In an industry study sponsored by the Yankee Group and Ernst and Young (2002), the top two factors that contributed to such a spontaneous impulse purchase indicated by survey respondents were a special sale price (75% respondents) and free shipping (49% respondents). Rich product stimuli, on the other hand, can engage the browsing customer, and help them discover new categories. Existing work has found that spontaneous purchases can be driven by strong emotional reactions to products (Rook, 1987; Rook and Gardner, 1993) evoked when consumers become involved in the product category (Bloch and Bruce, 1984; Laurent and Kapferer, 1985; Schmidt and Spreng 1999). The relative effect of product vs. price information on browsers remains to be tested.

Consumers whose state of shopping lies in between directed buying and browsing actively seek to obtain information to learn about available brands/features and decide amongst alternatives. Thus, product information in the form of buying guides, configurators and rich multimedia tools can be useful in educating the consumer and enhancing their product experience, while also helping move them closer to completing the purchase or becoming directed buyers. Specific and generic price-related information, however, do not help consumers make choices

[1]User-generated content (reviews) is not included since its content is not under the control of the retailer. We also do not include advertising sent to passive consumers (via email or banners) whose goal is to induce consumers to visit.

among or compare alternatives since they render all alternatives in a product category more attractive. Finally, we are interested in studying the relative effects of product and price information on purchase oriented outcomes within a session versus across sessions. Next, we describe our data and empirical strategy.

Data

We use a unique clickstream dataset obtained from consumer visits to four best-selling (focal) products during a 30-day period in late 2006 at a leading U.S. click-and-mortar retailer of durable goods. A session consists of an ordered and time-stamped sequence of clicks at the website. The data provides detailed insight into the type of pages viewed including category, product, and informational pages, promotions, customer service, catalogs etc. It also tracks consumers' use of search tools and decision-aids to refine and screen alternatives using price, brand, and features. However, clickstream data is in text form and requires extensive pre-processing. We built a custom PERL parser to filter and convert the information content of each click into numeric form. Further, in order to accurately encode the content, we also downloaded all relevant pages from the retailer's online site during the period of data collection.

The total number of sessions in the dataset identified using cookie ID and session ID is 86,231. We eliminate sessions which included only one page view[2]. We also cannot determine with certainty what product was purchased when we do not observe the product that the consumers clicked on to add to the cart. We therefore limit our examination to sessions where a focal product was viewed. Finally, to reduce the loss of first sessions from customers who might have made their first visit in the days preceding our data collection, we dropped all sessions in the first two days. This choice is supported by findings from a study of over 150 million online transactions across 800 retailers that found that among shoppers who left an online store due to concerns about security, trust, and the need to price-compare, nearly 80% of those who return did so within 1-2 days (McAfee 2009). Our final sample consists of 40,740 sessions from 36,636 unique users. 7,104 sessions (17.44%) are from repeat visitors. 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%.

Measures

Outcome and Information: A binary measure indicates whether a consumer made a purchase ($Purchase$) and added a product to the shopping cart ($Cart$). Product-related information ($ProdInfo$) measures whether consumers, during the online session, clicked on multimedia/rich media content that offer information on product features, ideas for their usage or applications, and guides for buying and product comparisons. As an example, in the case of appliances, the online retailer we study offered how-to documents, audios, videos 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. Specific price information ($SPriceInfo$)refers to information on the sales and promotions available for products in specific product categories. During the time of the data collection, there were category-specific price sales for three product categories. Generic price information ($GPriceInfo$) includes an offer of free shipping 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 the price and non-price promotions to customers. All are binary variables.

States of Shopping: Observed search and navigation patterns within a session form the basis for categorizing consumers' latent state of shopping. Past work has used breadth, depth, and intensity of search to differentiate between directed versus browsing behaviors (e.g., Moe 2003; Wolfinbarger and Gilly 2001). The breadth of search is defined using two measures[3]. The number of unique product departments ($DeptBreadth$)viewed refers to search across highly unrelated product categories or departments- e.g., clothes, food and decorations, while the number of unique product categories viewed ($CatBreadth$) refers to search within a department and across related product categories- e.g., men's clothes, women's clothes, and accessories. The depth of search reflects the extent of hierarchical search within the product category of the focal product ($Depth$), and is measured as the maximum number of times the customer drilled-down the search results. This measure is normalized since the four product categories allow for a different maximum depth by design of the category. Intensity of search is given by a set of variables that measure the level of involvement the shopper experienced in a given session. We measure the total time spent in minutes ($TotalTime$)and the number of pages ($TotalPages$) viewed in the session. We also include

---

[2] They could be hits from search engines, where the consumer immediately abandoned the session.
[3]The store is organized as Departments (Appliances)→Categories (kitchen app)→ Specific categories (refrigerators)→ products.

squared values of these two measures. We include a count of the total number of unique product detail pages viewed by the consumer ($TotalProducts$). 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 and searched only a few categories and/or that she viewed many product pages.

Controls: The first set consists of counts of consumers' use of electronic decision aids to screen and refine the available assortment of products in a category. By changing the composition of considered alternatives, the use of these tools may shape consumers' decision processes and have significant impacts on consumers' purchase behaviors (see Alba et al. 1997; Diehl et al. 2003; Hoffman and Novak 1996; Lynch and Ariely 2000; Winer et al 1997). We distinguish between refining criteria focused around price such as "under X dollars", "between X and Y dollars" ($\boldsymbol{PriceFacetedSearch}$) versus product attributes such as brands and features ($\boldsymbol{ProdFacetedSearch}$). Another tool is text-based search, using which consumers can directly search and locate items of interest using a textbox ($\boldsymbol{TextSearch}$). This is opposed to searching for products by using hierarchical drill-down. Consumers were also able to conduct comparisons of selected products using a side-by-side comparison matrix ($\boldsymbol{CompMatrix}$).

Our final set of controls are: $Date$ to account for daily fluctuations in the online environment not observed by us, $TimeofDay$of the session measured using dummies for morning, afternoon, or evening/night, whether the session was started on a $Weekend$, whether the $MonthofVisit$month was September or October, whether the session was a $RepeatVisit$, the order of the session within a cookie $OrdSession$, $ProductType$ dummies to indicate which of the four products the consumer searched for, and whether the consumer logged into a user account ($Account$). We also track the number of views of the following pages: home and sitemap ($HomePage$), local retail store and catalog pages ($StorePages$), external pages linked to from the retailer's site ($ExternalPages$), user generated content such as reviews and ratings ($UGCReviews$), and pages marked with error messages ($ErrorPages$).

Modeling the Latent States of Shopping

In order to allow consumers to funnel or belong to different latent states across multiple sessions, we categorize behavior at the level of a session. Here, we specify a model-based approach using finite mixtures to determine consumers' shopping states. While the observed purchase outcome is binary, we assume the existence of an underlying continuous latent variable $\boldsymbol{y_i^*}$ that is a linear function of observed characteristics as shown in eq (1).

$$\boldsymbol{y_i^*} = \boldsymbol{ProdInfo_i'\beta_{prod}} + \boldsymbol{SPriceInfo_i'\beta_{sprice}} + \boldsymbol{GPriceInfo_i'\beta_{gprice}} + \boldsymbol{BDI_i'\beta_{bdi}} + \boldsymbol{x_i'\beta_x} + \boldsymbol{u_i} \qquad (1)$$

$i = 1, \dots\dots, N$ sessions

Where $\boldsymbol{y_i^*}$ is the unobserved continuous random outcome variable (utility perceived by consumer)

$\boldsymbol{\beta_{prod}}, \boldsymbol{\beta_{sprice}}, \boldsymbol{\beta_{gprice}}$are the coefficients of online information

$\boldsymbol{\beta_{bdi}}$ are the coefficients of breadth, depth and intensity measures used to characterize latent states

$\boldsymbol{\beta_x}$ are coefficients for observed session-level explanatory variables

$\boldsymbol{u_i}$ is the IID error term; $u_i \sim N(0, \sigma_u^2)$; σ_u^2 normalized to 1 for identification

y_i^* is related to the observed binary variable as follows: $\boldsymbol{y_i} = \boldsymbol{Purchase_i} = \begin{cases} \boldsymbol{1} \ \boldsymbol{if} \ \boldsymbol{y_i^*} > 0 \\ \boldsymbol{0} \ \boldsymbol{if} \ \boldsymbol{y_i^*} \leq \boldsymbol{0} \end{cases}$ (2)

We anticipate the presence of response heterogeneity- that consumer sessions belonging to different latent states of shopping may experience varied impacts of product and price information obtained during the session. Finite mixture models are a useful tool to segment or group observations by differences in the effects of covariates on the dependent variable (McLachlan and Peel 2000; Wedel and Kamakura 2000. Each support of a heterogeneity distribution can be interpreted to represent a subset of sessions in the online store, and can be used to differentiate among them. The finite mixture model uses a discrete mixing distribution of the parameters and simultaneously estimates both consumers' *membership* in latent states and their *session-level* response parameters to improve the identification of states and model fit across the states. Other alternatives to modeling heterogeneity are available (e.g., cluster analysis, random coefficient models, and hierarchical bayesian models), but given our interest in characterizing the latent states, a finite mixture model is especially useful and managerially appealing in our context.

As a middle ground approach between pooled and individual heterogeneity models, we assume that the observations y_i are drawn from a G-component density **f**, and the mixture distribution is given by the weighted sum across the **g** components.

$$\mathbf{Pr}\ (\boldsymbol{y_i}\ \in \mathbf{population\ g}) = \boldsymbol{\pi_g} \tag{3}$$

The g-component mixture density is given by:

$$\boldsymbol{f}(\boldsymbol{y_i}|\boldsymbol{x_i}, \boldsymbol{ProdInfo_i}, \boldsymbol{SPriceInfo_i}, \boldsymbol{GPriceInfo_i}, \boldsymbol{BDI_i};\ \boldsymbol{\Omega_1}, \dots \boldsymbol{\Omega_G};\ \boldsymbol{\pi_1}, \dots, \boldsymbol{\pi_G}) = \sum_{g=1}^{G} \boldsymbol{\pi_g} \boldsymbol{f_g}(\ \boldsymbol{y_i}|\boldsymbol{x_i}, \boldsymbol{ProdInfo_i}, \boldsymbol{SPriceInfo_i}, \boldsymbol{GPriceInfo_i}, \boldsymbol{BDI_i};\ \boldsymbol{\Omega_g}) \tag{4}$$

$\boldsymbol{\pi_g}$ is the prior probability that observation y_i belongs to component $\mathbf{g}$. $\mathbf{0} < \boldsymbol{\pi_g} < 1$ and $\sum_{g=1}^{G} \boldsymbol{\pi_g} = \mathbf{1}$. $\boldsymbol{\Omega_g}$ are the component parameters that are to be estimated by maximizing the following log likelihood

$$\mathbf{max}_{\boldsymbol{\pi},\boldsymbol{\Omega}}\ \boldsymbol{LL} = \sum_{i=1}^{N} \left(\mathbf{ln}\left(\sum_{g=1}^{G} \boldsymbol{\pi_g} \boldsymbol{f_g}\left(\boldsymbol{y_i}|\boldsymbol{x_i}, \boldsymbol{ProdInfo_i}, \boldsymbol{SPriceInfo_i}, \boldsymbol{GPriceInfo_i}, \boldsymbol{BDI_i}\ ;\ \boldsymbol{\Omega_g}\right)\right)\right) \tag{5}$$

The density is estimated using maximum likelihood techniques using 10 sets of random starting values to avoid local maxima and incorrect or unstable parameters. Given estimates for $\boldsymbol{\delta}$ and our knowledge of $\boldsymbol{BDI_i}$ we can calculate the posterior probability that observation $\boldsymbol{y_i}$ belongs to component $\mathbf{g}$ given by Bayes theorem conditional on covariates and outcome. Each session is assigned to the group for which it has the largest (posterior) probability.

$$\mathbf{Pr}\ (\boldsymbol{y_i}\ \in\ \mathbf{population\ g}\ |\boldsymbol{x_i}, \boldsymbol{ProdInfo_i}, \boldsymbol{SPriceInfo_i}, \boldsymbol{GPriceInfo_i}, \boldsymbol{BDI_i}, \boldsymbol{y_i};\ \boldsymbol{\Omega}) = \frac{\boldsymbol{\pi_g} \boldsymbol{f_g}(\boldsymbol{y_i}|\boldsymbol{x_i}, \boldsymbol{ProdInfo_i}, \boldsymbol{SPriceInfo_i}, \boldsymbol{GPriceInfo_i}, \boldsymbol{BDI_i}, \boldsymbol{\Omega_g})}{\sum_{g=1}^{G} \boldsymbol{\pi_g} \boldsymbol{f_g}(\boldsymbol{y_i}|\boldsymbol{x_i}, \boldsymbol{ProdInfo_i}, \boldsymbol{SPriceInfo_i}, \boldsymbol{GPriceInfo_i}, \boldsymbol{BDI_i}, \boldsymbol{\Omega_g})} \tag{6}$$

Results: Identifying and Characterizing the Latent States

We estimated our model by increasing the components from 1 to 4. Following prior work, we use information criteria to select the best model - AIC and AIC3 (Bozdogan 1987), and BIC (Schwarz 1978). There is recent evidence that the AIC3 measure is more appropriate for discrete data (Andrews and Currim, 2003), and has shown remarkable performance in identifying the true model with only minor overfitting in Monte Carlo studies (e.g., Dias and Vermunt 2007). AIC on the other hand tends to choose the model with more parameter complexity, while BIC places a heavy penalty on complexity. We use AIC3, and choose 3-component solution as the best fit (see Table 1).

Table 1. Examining fit across multi-component models

# components	LL	AIC	AIC3	BIC
1	-3295.526	6703.051	6759.051	7185.484
2	-3196.634	6619.268	6732.268	7592.748
3	-2976.626	6299.252	6472.252	7789.624
4	-2919.453	6292.906	6519.906	8248.481

Our goal in using mixtures is to uncover underlying differences across consumers. From a managerial perspective as well, the usefulness of segmentation lies in its ability to uncover meaningful groupings that obtain different benefits from product and price related messages. In characterizing the sessions as described in Table 2a, we examine the nature of their search behaviors across departments and categories to measure the extent to which their search was focused or dispersed. We begin by assessing the breadth of interest displayed by consumers in the session. Sessions in State 1 had the lowest number of unique department and category breadth – that is they visited very few 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 (depth of search starting from departments to categories to products) while customer sessions in State 3 contained the fewest.

Next, we assess an array of variables that indicate the level or intensity of focus displayed by a customer in their session. 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, and a majority included pages related to the store, promotions, purchase and retailer policies – suggesting that they may be considering or making a purchase. Another related distinguishing variable is the ratio of product level to category level pages which is the highest for sessions in State 2, followed by sessions in State1 and then State 3. This variable provides one measure of the intensity of product search conducted within (focal) product categories. For sessions in State 3 this lower number indicates either that they viewed fewer product pages or conducted a dispersed search across many categories. Customers in State 2 and State 1 viewed significantly fewer product pages per minute than customers in State 3 did indicating that the former may have spent more focused time engaging with (reading about) products.

On the basis of the above characterization, we conclude that sessions in State 1 resemble directed buyers who are the most focused in their search and purchase activities (Moe 2003), sessions in State 2 are similar to searching and deliberating users who are conducting research and learning about products in the focal category, and sessions in State 3 are best described as browsers or experiential window shoppers whose interests were not focused. We refer to these three states of shopping as ***directed shoppers*** (DS), ***deliberating researchers*** (DR) and ***browsers*** (BR) henceforth. While not necessarily perfect, the categorization highlights the most prominent behaviors observed across these clusters.

Table 2a. Variables used to characterize latent states of shopping

	State 1		State 2		State 3	
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

We should also expect the purchase outcomes to differ across states if our labeling was reasonable. In Table 2b, we list variables that help validate our characterization including purchase outcomes.

Table 2b. Validating the latent states of shopping

	State 1		State 2		State 3	
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
SPriceInfo	0.310	0.462	0.284	0.451	0.263	0.440
GPriceInfo	0.507	0.500	0.504	0.500	0.503	0.500

Sessions from directed shoppers had the highest proportion of users that both added products to the cart (14.49%) and completed the purchase (5.13%), whereas sessions from browsers had the lowest proportions for both purchase related behaviors. Interestingly, we observe that conditional on adding a focal product to the shopping cart, browsers had a higher likelihood of completing the purchase (20.2%) than deliberating researchers (17.2%), but lower than directed shoppers (35.4%). Consumers across the different states perhaps use the shopping cart in different ways. State 1 sessions had the highest likelihood of being a repeat visit for a cookie.

Further, we find that customers conducting research were the least likely group to use online decision aids, indicating their greater reliance on compensatory choice processes in building their consideration sets. Whereas, directed shoppers and browsers displayed greater non-compensatory search through the use of decision aids to quickly narrow down the available assortment. Directed shoppers displayed a high usage of text search to directly find products they wanted, and also were most likely to use the product comparison matrix which allows users to side-by-side compare up to 4 chosen products.

In Table 3, we assess the transitions between the latent states uncovered by the mixture model for cookies with multiple or return visits. 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 3. State transitions for repeat visitors (excluding last session)

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

Modeling the Effects of Online Information on Purchase Behaviors

In this section, we develop a cookie-level panel model that allows us to examine both the *within-session* and *across-session* influence of online product and price information. Within-session refers to the impact of information obtained in a session on purchasing in the same session, while across-session refers to the impact of information obtained in a session on purchase-relevant behaviors in future session(s). The panel specification allows us to account for two forms of heterogeneity. The first is cookie-level unobserved heterogeneity which is stable within a cookie and 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 to represent the state following the categorization determined earlier. We are primarily interested in the varied effects of online information across the three states, and characterize them using interactions. To capture the effects of state transitions, we include dummies to represent the state in the preceding session. The across-session impacts of information are estimated from repeat customers. We examine the impacts of price and product information accumulated in past sessions on purchase in the current visit.

Model

$$\begin{aligned}\boldsymbol{y^*_{it}} = {} & \boldsymbol{PastProdInfo'_{it}\,\gamma_{pProd} + PastSPriceInfo'_{it}\,\gamma_{pSprice} + PastGPriceInfo'_{it}\gamma_{pGprice}} \\ & \boldsymbol{+\ PastCart'_{it}\gamma_{pCart} + ProdInfo'_{it}\,\gamma_{prod} + SPriceInfo'_{it}\,\gamma_{sprice} + GPriceInfo'_{it}\,\gamma_{gprice}} \\ & \boldsymbol{+ LATSTATE'_{it-1}\,\gamma_{pLatstate} + LATSTATE'_{it}\,\gamma_{latstate}} \\ & \boldsymbol{+ [ProdInfo'_{it} + SPriceInfo'_{it} + GPriceInfo'_{it}] * LATSTATE'_{it}\,\gamma_{info*latstate}} \\ & \boldsymbol{+ BDI'_{it}\,\gamma_{bdi} +\ x'_{it}\,\gamma_x +\ \alpha_i +\ \eta_{it}}\end{aligned}$$

$$where\ i = 1, \ldots\ldots, N\ \ cookies\ \ t = 1, \ldots.T\ sessions; \qquad \boldsymbol{y_{it} = 1(y^*_{it}} > 0) \tag{7}$$

$\boldsymbol{\gamma_{pProd}, \gamma_{pSprice}, \gamma_{pGprice}}$ are the coefficients of online information accumulated in the past
$\boldsymbol{\gamma_{pCart}}$ is the coefficient of the number of cart adds in past sessions
$\boldsymbol{\gamma_{prod}, \gamma_{sprice}, \gamma_{gprice}}$ are the coefficients of online information obtained in the current session
$\boldsymbol{\gamma_{pLatstate}}$ are the dummies that represent the (latent) 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 other observed session-level explanatory variables mentioned under measures
$\boldsymbol{\alpha_i}$ is the unobserved cookie-level random effect assumed to be uncorrelated with the covariates.

Let **z** contain the covariates in equation (7). We estimate the above model by maximizing the log-likelihood:

$$\boldsymbol{LL = \textstyle\sum_i \sum_t y_{it}\, ln\ F(z'_{it}\gamma) + (1 - y_{it})\ \mathrm{ln}\ (1 -\ F(z'_{it}\gamma))} \tag{8}$$

Results

The results from the cookie-panel model are presented in Table 4, column (1).

Table 4. Impacts of information and shopping state on purchasing in a session

	Covariates	(1) Full sample β	(1) Full sample s.e.	(2) Interested sample β	(2) Interested sample s.e.
Latent state in current session	LatState_DS	-2.725***	0.404	-0.998*	0.507
	LatState_DR	-0.126	0.147	-0.119	0.178
	LatState_BR	-1.025***	0.225	-1.328***	0.263
Latent state in past session	PastLatState_DS	-0.323**	0.109	-0.445***	0.122
	PastLatState_DR	0.212*	0.085	0.019+	0.093
	PastLatState_BR	0.225*	0.090	0.111+	0.099
Info gathered in the past	PastProdInfo	0.180***	0.040	0.175***	0.043
	PastSPriceInfo	-0.401***	0.095	-0.371***	0.099
	PastGPriceInfo	-0.115*	0.048	-0.128*	0.051
Info in current session for DS	ProdInfo	-0.352***	0.103	-0.465***	0.122
	SPriceInfo	0.753***	0.148	0.997***	0.179
	GPriceInfo	0.802***	0.197	0.932***	0.243
Info in current session for DR	ProdInfo	0.469***	0.077	0.402***	0.091
	SPriceInfo	-0.259*	0.127	-0.285+	0.149
	GPriceInfo	0.437*	0.169	0.398+	0.214
Info in current session for BR	ProdInfo	0.319***	0.098	0.274*	0.115
	SPriceInfo	1.187***	0.203	1.495***	0.236
	GPriceInfo	1.242***	0.238	1.471***	0.286
	σ_α	0.517	0.031	0.537	0.046
	ρ	0.211	0.020	0.224	0.030
	N	40736		11408	

Note: DV is Purchase. Model has full set of covariates with panel robust standard errors. *** p<0.001, **p<0.1, *p<0.5, + p<0.1

States of shopping: Purchase is a rare event in our data - reflected by the negative coefficients for all three states of shopping, with the highest rate for directed shoppers ($LatState\_DS$) at 5.12%, followed by researchers ($LatState\_DR$) at 1.80% and browsers ($LatState\_BR$) at 1.29%, controlling for covariates.

State transitions: In this analysis, the baseline consists of sessions without a past state. As seen in Table 4 column 1, the coefficient of $PastLatState\_DS$ is negative, while the coefficients of $PastLatState\_DR$ and $PastLatState\_BR$ are positive and significant. Returning to shop after being in a directed buying state had a significant negative effect, while returning to shop after having been in either deliberating or browsing states had significant positive effects on purchase. Sessions abandoned by directed shoppers are thus a costly loss - when these customers leave without purchasing, their likelihood of doing so when they return is significantly lowered. This result suggests that retailers should focus on trying to convert directed shoppers in the current session itself. In Table 5, we examine further details to help understand the impacts of transitioning between latent states of shopping. If directed shoppers return as researchers, the conversion rate is 4.57%, whereas it drops to 3.52% and 2.70% when they return as directed shoppers and browsers. On the other hand, when consumers' transition into the directed state of shopping after being in the other two states the results are optimistic. For instance, for 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 deliberating researchers and browsers we find that transitions to the directed buying state had the highest likelihood of converting in the next session, followed by transitioning to researching and last, browsing, providing some evidence of a funnel.

Table 5. The conditional effects of state and state transitions on completing a purchase

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

Past online information: We turn our attention to the impacts of online information accumulated in past sessions. In Table 4 column 1, the cumulative effects of product information obtained in earlier sessions ($PastProdInfo$) had a positive impact on purchasing in a given session. By contrast, the accumulated effects of category specific price promotions ($PastSPriceInfo$) and to a lesser extent generic price promotions ($PastGPriceInfo$) obtained in the past sessions had a negative effect on purchase in a given session. This finding highlights the potential negative future effects of promotions when consumers expect them but they may no longer be available. Next, we assess the contemporaneous or within-session impacts of information obtained by a consumer during the session itself

Product information ($ProdInfo$**)** had the strongest impact on within-session purchase for deliberating researchers, followed by browsers. Consumers who are conducting research, gathering information and deliberating about a product category are the ones that display the greatest positive response to the information contained in product buying and use guides, how-to documents, and multimedia demonstrations of product features. This information enables consumers to assess and compare products for an impending purchase, and to form preferences. This result provides empirical confirmation of an intuitive relationship. The impact on browsers is interesting. We find that product information had a positive impact on customers who were not necessarily focused on the particular category, suggesting that such information may have attracted customers to a product category. Browsing customers who may have had a general interest in the product category but not necessarily considering a near-term purchase and received product information appeared to purchase more often than browsers who did not receive product information.

In contrast to the impact on deliberating and browsing customers, product information appeared to lower the likelihood of purchase in a given session when presented to directed buyers. The negative effect on directed buyers is surprising. One possible explanation is that receiving detailed product information at this stage creates ambivalence or distraction when such information contradicts consumers' original impressions or preferences, especially if it highlights product-relevant aspects that the consumer may have overlooked or ignored before. Some past works have found that under certain circumstances, the use of decision tools and recommendation agents in online settings may provide suggestions that are counter to the preferences of users, thereby causing negative reactance (e.g., Fitzsimons and Lehmann 2004). We expect that such reactions might have caused directed buyers to delay (or abandon) their purchase upon obtaining product related information. Unlike browsers who have most likely not engaged in behaviors that create a commitment to purchase, directed buyers have in the recent past invested active time and effort (perhaps elsewhere) in considering the purchase. Thus obtaining product related information in the form of help and buying guides, and multimedia demonstrations appears to distract the latter type of buyer, while it attracts the former.

Category specific price information ($SPriceInfo$) had significant positive impacts on both directed shoppers and browsers, leading them to convert more often than in its absence. Consumers who display directed behaviors at a website are typically highly focused on a product category, and have usually narrowed down their consideration sets and are not seeking more product-related information (Moe 2003). Such consumers may price-shop across retailers as they look for deals or promotions on the specific product(s) that they are considering. Obtaining informattio on relevant promotions can therefore incentivize them to purchase from the said retailer. Browsers, on the other hand, are likely to convert upon receiving promotions if they consider the purchase price to be an attractive deal. Unlike directed buyers, browsers displayed a broader interest across product departments and categories during their session – suggesting that while they were interested in the focal product and its category, they may not have been actively seeking out related information. Obtaining price information on an attractive promotion or sale in a focal product category may therefore serve to generate interest and influence consumers to respond with an impulse purchase. In some cases, an impulse purchase may be driven because the browsing customer encounters information that stimulates their memory and reminds them about a product(s) that he/she had planned long before to purchase but had postponed or delayed it while awaiting to gather more information (perhaps about sales).

Interestingly, specific price information did not induce similar effects on deliberating researchers, and had a negative effect on their purchase behavior. This appears counter-intuitive at first, but to see why recall that these consumers are still conducting research, deliberating and forming consideration sets. Specific price information increases the utility of all products within a focal product category, without affecting their relative attractiveness. Obtaining information about a category price promotion improves the valuation of all products in a category, but this increased attractiveness of products might also mean that more alternatives now satisfy the feasibility constraints of a shopper. Thus, rather than help the customer move closer to making a purchase; in fact specific price information may delay their decision-making by increasing the number of alternatives whose (sale adjusted) values are now acceptable. In a wide variety of settings (e.g.,Chernev 2003; Dhar 1997; Dhar and Simonson 2003; Gourville and Soman 2005; Iyengar and Lepper 2000) it has been observed that consumers, when required to negotiate difficult trade-offs between alternatives, delay purchasing. Thus, adding more alternatives to the choice set caused choice overload, increased choice conflict and resulted in choice deferral. The deferral 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 is likely to happen when, for example, more appliances fall into a consumer's feasible set of alternatives, but they include machines that vary in the availability of features or attributes, thereby making comparison among them more difficult for the consumer. Overall, this finding suggests that for consumers in the deliberation state, price information by itself is insufficient to motivate them to purchase. They are instead likely to continue researching and gathering knowledge about products in the focal category.

Generic price information ($GPriceInfo$) or promotions related to shipping fees tended to have across-the-board positive impacts on consumer sessions belonging to all three states, suggesting that shipping offers continue to be highly valued by online buyers. In other words, the absence of free shipping or related promotions appears to lower the purchase likelihood for all consumers. This result is also supported by recent studies conducted by PayPal and comScore (2009) that found that the leading cause of shopping cart abandonment cited by 46% of respondent was high shipping charges. The strongest positive impact of shipping related price offers is interestingly observed for browsers, followed by directed shoppers and then researchers. Directed shoppers are those who are close to finalizing their purchase and have a deeper commitment to the purchase than browsers. This result suggests that directed buyers are less likely than browsers to abandon their purchase when a shipping offer is not available. Researchers who have not yet completed their evaluations and formed their preferences are only weakly (nevertheless significantly) influenced by free shipping offers.

Additional analyses using a restricted sample

One limitation of using clickstream to study consumer's purchase outcomes is that we cannot ascertain the true intent or motivation of consumers. While consumers may have visited a product page sometime during the session, it may not translate into true interest in the product and need not suggest that the product was considered for purchase by the consumer. We assess the validity of our results using an additional restricted sample where we place a stronger restriction on the customers whose sessions will be included. We require the customer to have displayed "substantial interest" in one of the focal products during at least one of his/her visits to the store- demonstrated either by adding a focal product to the shopping cart or in its absence, viewing the focal product multiple times. This results in an *interested sample* of 11,408 sessions from 8,842 unique cookies. The results are in column 2 in Table 4, and are broadly consistent with column 1. This engenders in our main findings about the effects of information.

Tradeoffs between the within-session and across-session impacts of information

Together, the results of the influence of three types of information underscore an important observation – that there is a tradeoff between the effects of product and price related information on purchase outcomes within a session and across sessions. Information about products in a focal category aid consumers in researching product alternatives, learning about product features, uses and applications, and has significant within-session influence on purchase behaviors for deliberating researchers (to a smaller extent browsers in some models). But it negatively influences the within-session conversion of directed shoppers. Exposure to product information, however, had the strongest positive influence on purchase decisions for returning consumers irrespective of their state of shopping in previous sessions. In contrast, both types of price information displayed strong positive effects on within-session purchase for customers in two latent states of shopping, but had weak to strong negative impacts on purchase for returning customers who abandoned sessions previously. We explore this tradeoff further. In addition to the above demonstrated contrasting effects on purchase that product and price information obtained in the past have, we examine whether online information influenced consumers who do not purchase the focal product to *return to visit the online store* in the future. In Table 6, we model this probability as a function of the information received in the current session. We specify a panel model to control for cookie-level unobservables, and the shopping state is modeled using dummies and interactions as before, for the full (column 1) and interested samples (column 2).

Table 6. Impact of information obtained within a session on return visit for non-purchasers

Latent state	Information	(1) Full sample		(2) Interested sample	
		β	s.e.	β	s.e.
Directed shopper	ProdInfo	0.434***	0.082	0.368**	0.143
	SPriceInfo	-1.532***	0.072	-1.464***	0.130
	GPriceInfo	-1.836***	0.105	-2.362***	0.247
Deliberating researcher	ProdInfo	0.295***	0.078	0.297*	0.136
	SPriceInfo	0.186**	0.060	0.346***	0.098
	GPriceInfo	-0.166	0.094	-0.616**	0.230
Browsers	ProdInfo	0.168*	0.080	0.160*	0.071
	SPriceInfo	0.123*	0.057	0.225*	0.099
	GPriceInfo	-0.066	0.094	-0.509*	0.229
	σ_α	0.243		0.352	
	ρ	0.056		0.110	
	N	39897		10569	

Note: DV is return visit. Model has full set of covariates with panel robust standard errors. *** p<0.001, **p<0.1, *p<0.5

For directed shoppers the coefficient of $SPriceInfo$ is negative. This suggests an interesting tradeoff, whereby it has a significant positive effect on helping directed shoppers to complete a purchase within a session, but when such a customer does not purchase and leaves (perhaps, in search of better deals or prices), she is also less likely to return. A similar pattern of tradeoff effects is observed for shipping related information $GPriceInfo$ on directed shoppers. Deliberators and browsers however experienced a different pattern of effects of price on the propensity to return. In Table 6, the coefficients of $SPriceInfo$ are significantly positive for both, while the coefficients for $GPriceInfo$ are insignificant. Thus, we observe a tradeoff between the within-session and across-session impacts of $SPriceInfo$ for deliberating researchers, but in a direction opposite to that experienced by directed shoppers. Information about a specific product category on sale did not convert deliberators into purchasers within a given session (in fact it had a negative effect), but it increased their likelihood of returning to the store. For browsers, $SPriceInfo$ had a positive effect on both buying within a session and returning to visit. Finally, while $GPriceInfo$ had broad positive effects on within-session purchase behaviors of consumers, it failed to influence researchers and browsers to return.

In contrast to these effects of $SPriceInfo$ and $GPriceInfo$, our results suggest that customers belonging to all three states of shopping who obtain and view product related information are more likely to return to visit the retailer after they abandon the session. This finding is relevant because it highlights the value of $ProdInfo$ in helping engage the customer and in building a relationship with them that extends beyond a given session. Given concerns echoed by several 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 a rich product experience for their customers. Product information helps to attract consumers back to the online store, and also has significant impacts on converting deliberating researchers within the session.

Robustness Checks

Endogeneity in Product Information

$SPriceInfo$ and $GPriceInfo$ are available to all customers in our dataset who view the focal product category on the days that the related discount and shipping offers were provided by the retailer. Therefore, as is to be expected, there is no significant difference in the amount of price information obtained across the three types of customer sessions (see Table 2b). However, the impact of $ProdInfo$ on purchase may suffer from endogeneity bias if consumers who are more likely to purchase were also the ones more likely to seek and obtain product information. Upon comparing the proportion of customers in each state (see Table 2), we observe that fewer deliberating buyers (the group with the strongest $ProdInfo$coefficient inTable 5) obtained product information than directed shoppers, suggesting that endogeneity may not be a concern. Yet, to more formally address this concern, we use the matching method to estimate the effects of $ProdInfo$.

Due to space constraints, we do not present a detailed discussion here. We use propensity score matching to estimate the sample average treatment effect on the treated (SATT) by comparing the purchase 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 ($ProdInfo$). Matching on such a score serves to simulate random assignment of treatment when: a) the observed covariates used to construct the score are balanced, and b) there is no bias from unobserved covariates (Abadie and Imbens 2002; Rosenbaum and Rubin, 1983). We check that condition a) holds, and restrict matching over the common support region. Since we have determined that important unobserved or latent differences exist across sessions, we additionally match using the latent state of shopping. This allows us to again separately identify the effect of (product) information on outcomes across different types of consumer sessions. This provides us with one way, albeit imperfect, in which to account for relevant unobservables. We use both caliper or radius matching ($r = 0.1$), and a block-stratified matching algorithm (results not shown), and find that our primary results remain robust. The coefficient of $ProdInfo$ is positive and significant for researchers and browsers; whereas it is negative to insignificant for directed shoppers.

Price vs. Brand Sensitivity of Consumers

We examine an alternate explanation for the differences in consumer purchase behaviors observed earlier that we attribute to states of shopping. Were consumers who completed the purchase when provided with price related (vs. product related) promotions merely more price-sensitive (or brand/feature-sensitive)? An important feature of shopping online is the availability of refining and screening tools which offer consumers the ability to alter the products that they see, and thereby affect the consideration sets that they build, and the final products that they choose to buy (e.g., Alba et al. 2000, Haubl and Trifts 2000; Lynch and Ariely 2000). We are specifically interested in consumers' use of price ($PriceFacetedSearch$) vs. product ($ProdFacetedSearch$) attributes to screen alternatives. We use this as a proxy for consumers' price vis-à-vis product sensitivity for purchases in the focal product category, and examine whether it influenced the main results. If this were the case, we should expect to see that deliberating researchers are more product-sensitive than directed buyers; and that directed buyers and browsers are more price-sensitive than deliberating researchers. Based on Table 2, browsers conducted more price-based than feature/brand-based refining and screening operations, and directed shoppers were equally likely to refine using both types of attributes, suggesting that they displayed greater non-compensatory search through the use of aids to quickly narrow down the available assortment. As a group, researchers were least likely to use either refining criterion (but relied slightly more on price-based screening), indicating their greater reliance on compensatory choice processes in building their consideration sets.

In Table 4, we had controlled for the extent of price vs. product refining conducted by a consumer in a session. However, we found that neither type of refining significantly influenced consumers' likelihood to purchase (not shown), whereas the coefficients for the states of shopping and information were significant. In order to assess whether the three latent states were masking consumer's price sensitivity, we include interaction terms between the three types of information and both types of refining and screening to additionally separate and control for their effects. We again find (not shown) that our main results for the effects of information obtained within a session and in the past sessions on conversion remain robust. The latent states of shopping remained a significant predictor of whether product or price information would influence shoppers to purchase.

Conclusions and Discussion

We began this study with the goal of determining how firms and retailers should manage the provision of online price and product-related information to customers who are actively visiting their online store. We derived a segmentation of customer sessions, more appropriately termed *states of shopping*, and assessed whether three types of commonly available information differently influenced purchase outcomes across the states. We found interesting tradeoffs in the effects of information on within-session vs. across-session purchase behaviors.

Implications: Our study and its findings provide firms with the knowledge that can be a useful starting point for differentiating sessions from anonymous customers in meaningful ways, and determining the optimal provision of product and price information to these different types of customers. In the absence of identifying information that is typically available in offline channels and for frequently purchase goods, durable good retailers have to devise alternate ways to distinguish their customers. A particularly interesting aspect of our study is the use of observed and easily available search and navigation activity on the website itself to generate the background covariates required to determine latent shopping states. Since historical purchase behaviors are not used, these techniques may allow retailers to actively target and interact with even *new* visitors to their web store. Our model of targeted information 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.

Our study questions the current common practice of offering promotions such as free shipping and category discounts to all customers visiting a store, and provides empirical evidence to support this intuition. 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 shopping states, retailers can instead optimally target product and price information to customers who are less likely to complete a purchase in the absence of such information, thereby increasing the lift created by online information. Moreover, depending on the retailer's goal – immediate conversion in the short term, i.e. before the customer ends a session, versus ensuring that the customer develops a longer-term relationship with the retailer and returns to the site over time – a different information provision strategy is likely to be optimal. This implication is driven by the tradeoffs or contrasting effects generated by our model for product and price information on purchase related behaviors within and across sessions.

Limitations and future work: Our study adds to a growing stream of research that suggests ways in which firms can improve their customer's online experience by making websites more usable and navigable (Agarwal and Venkatesh 2002; Palmer 2002; Venkatesh and Agarwal 2006), and how retailers can aid in consumers' online search and purchase decisions (Hauser 2009; Novak et al. 2000). Our current work is based on a sample observed over a short period that precludes us from studying purchases that may have occurred beyond 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 different effects of buying guides vs. other multimedia demonstrations, for instance. This study should also be extended to study the effects of UGC such as reviews that is becoming wildly popular in online shopping contexts.

A modeling limitation of our current study is the separation of the tasks of identifying latent states at the session level and estimation of information effects using a cookie-panel. While combining them would require us to make several additional assumptions about the distribution of unknown parameters (that drive the latent state and state transitions) that may not necessarily be realistic, it can help validate the robustness of our findings. In future studies, it will be useful to examine the pathways of influence – how product vs. price information 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 (and advance from being a browser to a deliberating researcher to a directed shopper)? 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 even more specific decisions related to optimal provision of online information.

References

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

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), pp.57–78.

Andrews, R. L., and Currim, I.S. 2003. "A Comparison of Segment Retention Criteria for Finite Mixture Logit Models," *Journal of Marketing Research* (40), pp.235-243

Bloch, P. H. and Bruce, G.D. 1984. "Product Involvement as Leisure Behavior," in *Advances in Consumer Research*, Vol. 11, ed. Thomas C. Kinnear, Ann Arbor, MI: Association for Consumer Research, pp.197- 202.

Bradlow, E., Huber, J., Johnson, E., Khan, B., Little, J., and Schkade, D. 1999. "Agents to the rescue?" *Marketing Letters* (10: 3), pp.285–300.

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* (40: August), pp.249-267

Canter, D., Rod, R.., and Storrs, G. 1985. "Characterizing user navigation through complex data structure", *Behaviour and Information Technology* (4:2), pp. 93-102

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.

Chatterjee, P., Hoffman, D.L., and Novak, T.P. 2003. "Modeling the Clickstream: Implications for Web-Based Advertising Efforts," *Marketing Science* (22:4), pp.520-541.

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.

Compete Online Shopper Intelligence Study. 2010. http://blog.compete.com/2010/02/22/online-shopper-intelligence-study-released/

Cooley, R., Tan, P-N., and Srivastava, J. 1999. "WebSIFT: The web site information filter system," In *Proceedings of KDD' 99 Workshop on Web Usage Analysis and User Profiling* WEBKDD'99, San Diego, CA, Aug. 1999. ACM (Masand and Spilipoulou eds.)

Cox, D. F. 1967. *Risk taking and information handling in consumer behavior*, Cambridge, MA: Harvard University Press.

Danaher, P. J., Wilson, I. W., and Davis, R.A. 2003. "A comparison of online and offline consumer brand loyalty," *Marketing Sci*ence (22), pp.461–476

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.

Dhar, R. and Simonson, I. 2003. "The Effect of Forced Choice on Choice," *Journal of Marketing Research* (40: May), pp.146–60.

Dholakia, U. M., and Bagozzi, R. P. 2001. Consumer behavior in digital environments. In J. Wind & V. Mahajan (Eds.), *Digital marketing:Global strategies from the world's leading experts* (pp. 163-200). New York: Wiley

Dias, J.G., and Vermunt, J.K. 2007. "Latent class modeling of website users' search patterns: Implications for online market segmentation," *Journal of Retailing and Consumer Services* (14), pp. 359-368

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

Gelb, B.D., Andrews, D., and Lam, S.K. 2007. "A Strategic Perspective on Sales Promotions," *Sloan Management Review*,http://sloanreview.mit.edu/the-magazine/articles/2007/summer/48401/a-strategic-perspective-on-sales-promotions/

Gomory, S., Hoch, R., Lee, J., Podlaseck, M., and Schonberg, E. 1999. "Analysis and visualization of metrics for online merchandizing," In *Proceedings of KDD' 99 Workshop on Web Usage Analysis and User Profiling* WEBKDD'99, San Diego, CA, Aug. ACM (Masand and Spilipoulou eds.)

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

Greenleaf, E. A. and Lehmann, D.R. 1995. "Reasons for Substantial Delay in Consumer Decision Making," *Journalof Consumer Research* (22: September), pp.186–99.

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

Gupta, S. and Chintagunta, P.K. 1994. "On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models."*Journal of Marketing Research* (31:1), pp. 128-36.

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.

Hauser, J.R., Urban, G.L., Liberali, G. and Braun, M. 2009."Website Morphing," *Marketing Science* (28:2), pp.202-223

Hodkinson, C., and Kiel, G. 2003. "Understanding Web Information Search Behavior: An Exploratory Model," *Journal of End User Computing* (15:4), pp. 27-48.

Hoffman, D. and Novak, T. 1996. "Marketing in hypermedia computer-mediated environments: conceptual foundations," *Journal of Marketing* (60: July), pp.50–68.

iCrossing. 2010. " How America Searches: Online Retail Report," http://www.icrossing.com/articles/How%20America%20Searches%20-%20Online%20Retail.pdf

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:), 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.

Krishnamurthi, L., Mazumdar, T., and Raj, S.P. 1992. "Asymmetric Response to Price in Consumer Brand Choice and Purchase Quantity Decisions," *Journal of Consumer Research* (19:3), pp. 387-400

Koufaris, M., Kambil, A., and LaBarbera, P.A. 2001. "Consumer Behavior in Web-Based Commerce: An Empirical Study," *International Journal of Electronic Commerce* (6:2), pp. 131-154.

Laurent, G. and Kapferer, J.N. 1985. Measuring Consumer Involvement Profiles. *Journal of Marketing Research* (22:1), pp. 41–53

Lee, L. and Ariely, D. 2006. "Shopping Goals, Goal Concreteness, and Conditional Coupons." *Journal of Consumer Research* (33), pp. 60-70.

Li, Kairong, Kuo, C. and Russel, M. G. 1999. "The Impact of Perceived Channel Utilities, Shopping Orientations, and Demographics on the Consumer's Online Buying Behavior," *Journal of Computer Mediated Communication* (5:2), online

Lucas, P. 2001. "How Product Demos are Helping Shoppers Kick the Tires on the Web," *Internet Retailer Featured Article*, Feb, 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. 2000. "Wine Online: Search Cost and Competition on Price, Quality, and Distribution," *Marketing Science* (19:Winter), pp. 83–103.

Mack, D. 2009. "Are People Shopping Online for Big-Ticket Items?," http://blogs.adobe.com/scene7/2009/06/are_people_shopping_online_for.html

Maddala, G. S. 1987. "Limited Dependent Variable Models Using Panel Data." *Journal of Human Resources* (22:3), pp.305-338.

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.

Mandel, N., and Johnson, E.J. 2002. “When web pages influence choice: Effects of visual primes on experts and novices,” *Journal of Consumer Research* (29), pp.235-245

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, 2004. 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), 2004, pp. 326-335.

Montgomery, A., Hosanagar, K., Krishnan, R., and Clay, K. B. 2004. “Designing a Better Shopbot,” *Management Science* (50:2), pp. 189-206.

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.

Novak, T., Hoffman, D. and Yung, Y. 2000. “Measuring the customer experience in online environments: a structural modelling approach”, *Marketing Science* (19:1), pp.22–42.

Olshavsky, R.W. 1985. “*Perceived quality in consumer decision-making: an integrated theoretical perspective*, Lexington. MA: DC Heath and Co.

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.

Palmer, J. W. 2002. "Web site usability, design, and performance metrics," Information Systems Research (13:2), pp.151–167.

Palmisano, C., Tuzhilin, A., and Gorgoglione, M. 2007. "User Profiling with Hierarchical Context: An e-Retailer Case Study," *Context*, pp. 369-383

Park, Y-H. and Fader, P.S. 2004. “Modeling Browsing Behavior at Multiple Websites,” *Marketing Science* (23:3), pp. 280-303.

Payne J.W., Bettman J.R., and Johnson E.J. 1993. *“The adaptive decision maker”*. Cambridge, NY: Cambridge University Press.

Perkowitz, M., and Etzioni, O. 1998. "Adaptive Web Sites: Automatically Synthesizing Web Pages," *Proceedings of the 15th National Conference on Artificial Intelligence*, pp.727-732

PriceGrabber Consumer Behavior Report. 2009. https://mr.pricegrabber.com/Online_Home_Furniture_Appliance_and_At_Home_Entertainment_Trends_May_2009_CBR_1.pdf

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 buying’s affective antecedents”, In *Research in consumer behavior* vol 6, (Janeen A-C, Russell WB, editors), Greenwich, CT: JAI Press

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

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

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.

Schmidt, J. B. and Spreng, R. A. 1966. “A Proposed Model of External Consumer Information Search,” *Journal of the Academy of Marketing Science* (24:3), pp. 246-256.

Schwarz, G. 1978 “Estimating the dimension of a model*,” Annals of Statistics (*6), pp. 461-464.

Shim, S., Eastlick, M.A., Lotz, S.L., and Washington, P. 2001. “An Online Pre -purchase Intentions Model: The Role of Intention to Search,” *Journal of Retailing* (77), pp.397-416

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), pp.306-323

Spiliopoulou, M., Pohle, C., and Faulstich, L.C. 1999. "Improving the effectiveness of a web site with web usage mining," In *Proceedings of KDD' 99 Workshop on Web Usage Analysis and User Profiling* WEBKDD'99, San Diego, CA, Aug. ACM (Masand and Spilipoulou eds.).

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), 12-23.

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.

Tedeschi,B. 2006. "Buying a new fridge? Watch the interactive video first" nytimes.com/2006/05/15/technology/15ecom.html

Titterington, D., Smith, A., and Makov U. 1985. "Statistical Analysis of Finite Mixture Distributions," John Wiley & Sons.

Titus, P.A. and Everett, P.B. 1995. "The consumer retail search process: a conceptual model and research agenda," *Journal of the Academy of Marketing Science* (23:2), pp.106–119.

Trope, Y., and Liberman, N. 2003. "Temporal Construal," *Psychological Review* (110), pp.403-421.

Urbany, J.E., Dickson, P.R., and Kalapurakal, R. 1996. "Price Search in the Retail Grocery Market," *Journal of Marketing* (60:2), pp. 91–104.

Venkatesh, V., and Agarwal, R. 2006. "From Visitors to Customers: A Usability-Centric Perspective on Purchase Behavior in Electronic Channels," *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.

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

Wedel M., and Wagner, K. 2002. "Introduction to the special issue on market segmentation," *International Journal of Research in Marketing* (19:3), pp. 181-183.

Winer, R.S., Deighton, J., Gupta, S., Johnson, E.J., Mellers, B., Morwitz, V.G., O'Guinn, T., Rangaswamy, A. and Sawyer, A.G. 1997. "Choice in Computer-Mediated Environments", *Marketing Letters (*8:3), pp. 287–96.

Wolfinbarger, M. and Gilly, M. 2001. "Shopping online for freedom, control, and fun", *California Management Review* (43:2), pp.35–55.

Wu, J., Cook, V.J. Jr., and Strong, E.C. 2005. "A Two-Stage Model of the Promotional Performance of Pure Online Firms," *Information Systems Research* (16:4), pp.334-351

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.

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