A Bayesian Treed Model of Online Purchasing Behavior

Using In-Store Navigational Clickstream

Wendy W. Moe
Assistant Professor of Marketing
University of Texas at Austin
2100 Speedway, CBA 7.216
Austin, TX 78703
wendy.moe@bus.utexas.edu
(512) 232-2793

Hugh Chipman
Associate Professor of Statistics
University of Waterloo

Edward I. George
Professor of Statistics
University of Pennsylvania, Wharton

Robert E. McCulloch
Professor of Econometrics and Statistics
University of Chicago

April 2002

A Bayesian Treed Model of Online Purchasing Behavior Using In-Store Navigational Clickstream

ABSTRACT

Internet clickstream data allows online retailers to observe their customers as they click from page to page. The ability to infer a customer's motivation and search strategy (and therefore his/her likely response to in-store experiences) from observed clickstream data has tremendous benefits for online marketers when segmenting and targeting these individuals. Toward that effort, we develop a Bayesian treed model that simultaneously (1) groups shoppers based on patterns observed in their online navigational clickstream and (2) examines their purchasing decision as a function of in-store experiences. Bayesian treed models are ideal for this research problem as they can identify structure across observations by accommodating two different sources of variance, using variables for segmentation as well as for prediction within segments. First, we account for heterogeneity by branching the tree into segments depending on how each observation scores along navigational measures. Then, variance in purchasing behavior within each terminal node of the tree is modeled to examine the purchasing processes driving each visit type. Alternative methods such as a latent class logit model are also considered but are seen to be significantly less effective in representing the patterns found in the data.

A Bayesian Treed Model of Online Purchasing Behavior Using In-Store Navigational Clickstream

Introduction

Despite the rapid growth in e-commerce, online purchasing conversion rates have remained chronically low. For a typical online retailer, only 1-2% of visits convert into purchases (Brownlow 2001). This implies that over 95% of visits to a given retail site are made by shoppers who are merely browsing. Increasing these conversion rates has been a primary focus of many e-commerce marketers. But before we can begin exploring methods to achieve this goal, we must first gain a clearer understanding of the differences between browsing and buying visits in terms of their underlying motivations and purchasing processes. Additionally, we must also develop methods and measures with which marketers can differentiate between these visits.

For example, imagine a student who needs to buy a very specific textbook for class. This individual's behavior at an online bookstore will be very different from that of someone who may merely be browsing for the entertainment value of window shopping, both in terms of navigational pattern as well as purchasing response to various aspects of the store visit. In a previous study, online shoppers were found to vary dramatically in terms of their motivations and search strategies and could be identified by their unique clickstream patterns (Moe 2002). In this paper, we develop methods and measures not only to differentiate between different types of store visits but also to model the differences that exist between purchasing processes.

Internet clickstream data allows online retailers to observe their customers as they click from page to page. The ability to infer a customer's motivation and search strategy (and therefore his/her likely response to in-store experiences) from observed clickstream data has tremendous benefits for online marketers when segmenting and targeting these individuals. Toward that effort, we develop a Bayesian treed model (Chipman, George, and McCulloch 2002, hereafter CGM) that simultaneously (1) groups shoppers based on patterns observed in their online navigational clickstream and (2) examines their purchasing decision as a function of in-store experiences. Bayesian treed models are ideal for this research problem as they can identify structure across observations by accommodating two different sources of variance, using variables for segmentation as well as for prediction within segments. First, we account for heterogeneity by branching the tree into segments depending on how each observation scores along navigational measures. Then, variance in purchasing behavior within each terminal node of the tree is modeled to examine the purchasing processes driving each visit type. Alternative methods such as a latent class logit model are also considered but are seen to be significantly less effective in representing the patterns found in the data.

Previous studies (Bucklin and Sismeiro 2000, Novak, Hoffman, and Yung 2000) have explored navigational behavior in terms of the effects of page-depth (the number of pages viewed in a shopping session) and session-duration (the amount of time spent during an online session). However, these studies and the measures used in them have ignored the *content* of the pages viewed and the effect of this content on purchasing. But without knowing *what* the shopper sees from page-to-page, it is difficult to understand the effects of these pages in terms of their role in the navigational process and their influence on the purchasing decision. In this paper, we directly examine the role of page content by using a number of within-session metrics that are suitable for most online retailers and efficiently characterize the unique patterns found in navigational clickstream data. Used in our Bayesian treed model, these measures will allow us to differentiate between store visits and hence model purchasing accordingly.

Additionally, it has been shown that consumers often construct their preferences during their online shopping session according to the page content encountered during the session (Mandel and Johnson 2000). But despite this apparent relationship between in-store navigation and preference construction, very little research, if any, has been conducted to study the effect of pageviews on whether or not the individual purchases. Therefore, at the root of our Bayesian treed model is a logit model of purchasing that examines the effect various aspects of a shopper's in-store experience (characterized by the pages and content viewed) have on the buying decision. But since the model also allows for heterogeneity in motivation and search strategy across store visits, we can measure differences in these pageview effects on purchasing across different types of shopping sessions.

In the next sections, we will review some of the relevant literature on search behavior and discuss the implications they have for modeling in-store navigational behavior and purchasing. Then we will provide a thorough discussion of the type of data and measures that we will use in this study. From there, we describe Bayesian treed models and apply the method to our online shopping problem.

Online Search Behavior

Before we begin developing a model of navigational patterns and purchasing probabilities, we first discuss some of the ways in which online search behavior is expected to vary. Specifically, we differentiate in-store behavior along two dimensions: (1) goal-directed versus exploratory search and (2) stage of purchasing process.

Previous offline research has dichotomized search behavior into goal-directed versus exploratory search (Janiszewski 1998). Others have also applied such a dichotomy to online behavior when segmenting shoppers (Moe 2002) or conceptualizing an internet user's state of *flow* (Hoffman and Novak 1996). In these studies, goal-directed search refers to behavior for which the shopper has a specific or planned purchase in mind. The objective of search, in this situation, is to gather relevant information regarding a specific purchasing decision (Brucks 1985, Wilkie and Dickson 1985). Therefore, it would be expected of a goal-directed shopper that in-store navigation is very focused around a given purchasing decision.

> <u>Assumption 1</u>: Directed search behavior can be characterized by a high degree of *focus* exhibited in a shopper's store visit. The more focused the store visit is around a specific product purchasing decision, the more likely the shopper is engaging in goal-directed search behavior.

Exploratory search, on the other hand, refers to behavior in which the consumer is less deliberate, less focused, and perhaps not even considering a purchase. Instead, navigation tends to be undirected and stimulus-driven rather than goal-driven (Janiszewski 1998). In this case, the shopper derives hedonic utility from the shopping process itself and the environmental stimuli encountered (Babin, Darden, and Griffin 1994, Sherry, McGrath, and Levy 1993, Hirschman 1984). This type of behavior is also sometimes referred to as browsing or ongoing search. Since ongoing search is not motivated by any specific decision making need, in-store activities are expected to exhibit less focus around a specific product decision and more browsing variety across different products and product categories.

Assumption 2: Exploratory search behavior can be characterized by a low degree of *focus* in a shopper's store visit around a specific product decision. The more a shopper's navigational patterns reflect a highly varied set of product considerations, the less likely the shopper has a planned purchase in mind and the more likely the individual is engaging in exploratory search behavior.

One key differentiating factor between goal-directed and exploratory search behavior is the shopper's purchasing orientation at the start of the store visit. For example, exploratory shoppers do not necessarily enter the store with an intended purchase in mind. In contrast, goal-directed shoppers often have a very specific product purchase in mind and, as a result, tend to be predisposed to purchasing upon entering the store. Because of these differing motivations, we expect goal-directed shoppers to be more likely to purchase than exploratory shoppers.

Proposition 1: Goal-directed shoppers, as identified by the level of focus in their navigational patterns, are more likely to purchase than exploratory shoppers.

In addition to differences in navigational behavior and in the overall propensity to purchase between goal-directed and exploratory search, there are also significant difference in the purchasing *process* that each entails. The more commonly studied search behavior in marketing research tends to be goal-directed search where the individual goes through a process of forming a consideration set, deliberating the options, and making a final purchasing decision based on the attributes of each product in the set (Kardes et al 1993, Shocker et al 1991, Ratneshwar et al 1991, Barsalou 1991). The top panel of Figure 1 illustrates the purchasing process of a goal-directed shopper.

FIGURE 1 HERE (The Purchasing Process)

Navigational patterns for a goal-directed shopper will depend on the stage of the purchasing process. Initially, in-store navigation will aim to construct a consideration set by searching across products within a particular category. Once this is done, the behavior turns to the deliberation of options where each item in the consideration set is carefully evaluated based on their product attributes. Depending on the outcome of the deliberation process, purchase may or may not occur.

Putsis and Srinivasan (1994) studied the pre-purchase deliberation process described here and proposed that customers endeavor to gather and accumulate information during the pre-purchase deliberation process. When enough relevant information is acquired to surpass a threshold, purchase occurs (Putsis and Srinivasan 1994, Moe and Fader 2001). In many cases, goal-directed shoppers have already made the decision to buy in a particular product category, and it is simply a matter of choosing a specific product in that category. This final decision hinges on the shopper's ability to gather enough product information pertaining to that purchase. Early stage shoppers (i.e., those who are still building their consideration set) have yet to acquire enough information to make such a decision and are therefore less likely buy. Later in the process, after sufficient product-related information has been collected, more in-depth deliberations are possible, making purchasing more likely.

> Proposition 2: Goal-directed shoppers in the later stages of deliberation are more likely to purchase than those goal-directed shoppers who are in earlier stages of the purchasing process.

Goal-directed shoppers are influenced by different factors depending on the stage of the decision process. Early stage goal-directed shoppers are focused on identifying eligible products and constructing their consideration set. Therefore, these shoppers will respond positively to any in-store experiences that aid in this process. Once a shopper enters the deliberation stage, in-store activities that allow the shopper to more carefully examine a product will increase the individual's likelihood of buying. This could include reviewing detailed product information previously seen during the consideration set formation process but not necessarily closely examined.

> <u>Proposition 3a</u>: Early stage goal-directed shoppers will be positively influenced to purchase by experiences that help them identify eligible products and construct their consideration set.
>
> <u>Proposition 3b</u>: Late stage goal-directed shoppers are positively affected by reviewing any product information that allows them to more closely examine a product in the consideration set. This information has likely been viewed earlier when the consideration set was being constructed.

Exploratory shoppers go through a very different purchasing process (Figure 1). Exploratory search starts not with the objective of forming a consideration set but with the voluntary exposure to environmental stimuli (e.g., viewing a wide variety of products and product categories). One of two outcomes could result from this activity. First, the stimulus encountered may not be of any interest to the shopper and therefore would not influence the shopper to buy. In this case, the shopper exits the store without purchasing. However, the shopper may also encounter interesting information that may stimulate the shopper to consider a purchase. This would drive the shopper to further examine and evaluate a specific product, potentially leading to what is defined as an unplanned or impulse purchase (Janiszewski 1998, Jarboe and McDaniel 1987). For example, imagine an exploratory shopper in an online bookstore. As the shopper browses across categories and across books, he/she will encounter many new products and much product information. If these products are uninteresting to the shopper (non-positive influence), then the shopper will ultimately leave the store without buying. If, however, the shopper comes across a book that looks interesting, he/she may examine the specific book more closely by acquiring detailed product information. Shoppers who reach this later stage of exploratory search are more likely to buy than early stage exploratory shoppers who have yet to find a product worth considering.

> <u>Proposition 4</u>: Exploratory shoppers in the later stages of deliberation (i.e., those who are examining and evaluating product details) are more likely to buy than those who are still just browsing.

We now turn to the factors that influence purchasing among exploratory shoppers. In general, the probability that these stimulus-driven shoppers will find an item worth considering is a stochastic process, where each product-related experience has the potential of generating interest and perhaps resulting in a purchase. We therefore expect that increased exposure to products available at the store increases the probability that the shopper will buy. As an illustrative example, imagine a child in a toy store. The more new toys the child is exposed to in the toy store, the more likely one of the toys will be "irresistible" and ultimately purchased.

> <u>Proposition 5</u>: The more exposure an exploratory shopper has to products in the store, the more likely he/she will be to buy. That is, purchasing is positively influenced by increased product-related experiences.

After a product has been identified as a possible purchase, the shopper enters the deliberation stage. In this stage, the shopper's objective is to further examine a product previously viewed and deemed interesting. Exploratory shoppers in this stage are similar to early stage goal-directed shoppers in that repeat examination of product information helps the shopper further evaluate the purchasing decision, increasing the likelihood that he/she will reach a final decision and buy.

Proposition 6: Late stage exploratory shoppers (who are deliberating a purchase) are positively influenced to buy when reviewing product information for items previously identified as potential purchases.

Data

The Store Site

The context of our study is an online retailer that sells nutrition products such as vitamins, weight loss aids, body-building supplements, etc. The range of their product offerings provides a mix of customer types ranging from casually health conscious consumers, interested in buying daily vitamins and nutrition supplements, to health and body-building fanatics, looking for performance enhancers and protein supplements. These shoppers vary dramatically in terms of their objectives, involvement levels, and expertise, which should lead to very different shopping strategies as reflected by their page-to-page behavior at the site. A relatively small and new site, the store experiences roughly 5,000 to 10,000 visits a month, approximately 80% of which are made by unique visitors. Their conversion rates, the proportion of visits that end with a purchase, are in line with the industry averaging slightly less than 2%.

From the site's homepage, the shopper has a number of options. First, the store visitor may choose to login and manage their account. This includes activities such as registering as a new user, updating your personal profile, monitoring the status of a purchase, etc. Second, the visitor may view informational pages. A common practice of e-commerce retailers is to provide community areas on their sites or advice columns to help shoppers learn more about nutritional and health issues related to the products that they sell. Visitors may also access the customer service pages for information about the shipping, return policies, the company, and privacy issues before deciding whether or not to transact with the site. Last, but definitely not least, are the product-related shopping pages. From the home page, you can drill down by different product categories (e.g., vitamins, nutrition bars, fat loss, etc.) or the shopper may jump directly to a specific product of interest by using the site's search function.

Data Collection

Data collection is done by a market research firm, NetConversions, employed by the store site. NetConversions uses cookies downloaded onto the visitor's computer when they visit the store site to track the shopper's behavior at that site alone. For each shopper, NetConversions records an identification number, the pages viewed, and the precise time those pages were viewed. The action taken on each page is also recorded. For example, when a shopper submits a purchase transaction, that information is recorded. Since NetConversions does not use the store site's customer registration information, no demographic or financial data is recorded. The data collected monitors *only* the observable behavior at the site without knowing the identity of the shopper, thereby protecting their privacy.

Rather than simply recording the URLs viewed by the shopper, NetConversions categorizes each of the pages on a web site and records pageviews as described by these categories. Table 1 provides descriptions of the general page categories used in the data set: (1) administrative, (2) informational, and (3) shopping pages. Of particular interest are the shopping pages[1]. Online stores are typically organized by product category or department. In the case of the nutrition site, for example, shoppers may view a list of the different nutrition bars the store offers. This type of page would be categorized as a *category page*. If the shopper sees a particular product of interest, she can click on the product for additional information. This would be categorized as a

[1] Sessions that contained no shopping pages were omitted from our study.

product page. Additionally, each product page is sub-typed by the product's category and brand. Shoppers can also jump directly to a product page as a result of a search. Since the search engine is available on all pages of the site, the resulting page of matches is coded as the search page.

TABLE 1 HERE (Page Categorizations)

Data Summary

The data used in this study spans six months from June 1, 2000 to November 30, 2000. In this time, 34,585 visit sessions were made by 24,645 unique visitors. This constituted over 138,775 lines of pageview data. Of the 34,585 sessions, only 560 were accompanied by a purchase (427 were unique buyers). This results in a 1.62% purchase conversion rate which is fairly typical in e-commerce retailing.

We will aggregate this page-to-page data to the session level using measures that we will describe next. On average, a shopper viewed 4.01 pages per visit for a total of 452 seconds. However, 20,100 were one-page visits. These one-page visits may represent shoppers that unintentionally entered the store site either by typing in the wrong web address or thinking the store was something it was not. Regardless of the reason for the one-page visit, viewing a single page does not constitute a substantial shopping experience and is therefore not the focus of our study. Though we do not remove these observations from our dataset, the treed model presented later in this paper will separate them from the bulk of the analysis.

Measures

Our objective in this paper is to develop a Bayesian treed model that segments store visits based on navigational patterns and then models purchasing behavior within each segment. Therefore, we need to create two types of measures from our data: (1) navigational measures and (2) purchasing influence measures (see Table 2 for a summary of all the measure described in this section).

TABLE 2 HERE (Summary of Measures)

First, we need a set of measures to characterize the navigational patterns observed within each session. Two general measures which provide a simple summary of the depth of store visit are the total number of pages viewed during the session (PAGES) and the average time spent per page (PGTIME). However, the propositions discussed earlier are on our ability to measure behavior that differentiates goal-directed from exploratory behavior. One characteristic that varies across goal-directed and exploratory behavior is that of navigational *focus* (see Assumptions 1 & 2). Therefore, we develop a set of *focus* measures that reflect the types of pages each shopper chooses to view during the store visit. Specifically, we differentiate between category level and product level pages. A goal-directed shopper is expected to view a large proportion of product level pages for the detailed product information needed in a specific evaluation process. In contrast, exploratory shoppers, especially those in the early stages, may tend to focus on category level pages that offer brief summary information on a wide array of individual products. Therefore, we employ two measures in describing the focus of navigational patterns: (1) the percentage of pages viewed which are at the category level (CATPG) and (2) the percentage at the product level (PRODPG) [2].

In addition to focus measures, we also construct a number of *browsing variety* measures. The first is a product-to-category ratio (PRODCAT). That is, what is the average number of unique products viewed in a given category? A high product-to-category ratio would indicate that the shopper searched many products within a limited set of categories suggesting that the individual may be constructing a consideration set and therefore is a goal-directed searcher. On the other hand, a session that exhibits high browsing variety with a low product-to-category ratio would indicate that the shopper is jumping from one category to another suggesting that the individual may be an exploratory searcher. Additionally, we will examine two other variables that also reflect the level of browsing variety: (1) the percentage of all category level pages viewed that are unique (UNIQCAT) and (2) the percentage of all product level pages viewed that are unique (UNIQPROD).

Aside from differentiating between goal-directed and exploratory behavior, we also wish to assess the shopper's stage in the decision process. Some of the above measures also provide an indication of the shopper's stage of search. But in addition to the focus and browsing variety measures mentioned in this section, we also examine the maximum number of repeat pageviews per product (MAXREP) as an approximation of the shopper's level of deliberation. For example, a shopper who is close to purchasing and is deeply deliberating a particular product is likely to repeat view pages that offer detailed information about that product. Therefore,

[2] All percentage measures in this section represent the percentage of the number of shopping pages viewed in the session and not the percentage of all pages viewed.

shopping sessions with high levels of repeat product page viewings will be considered sessions in the deliberation stage of the purchasing process.

The second set of measures that we will define are those factors that influence purchasing behavior. Figure 2 provides a histogram of the number of visits associated with a given level of pageview depth (i.e., the number of pages viewed in the visit). Also plotted in Figure 2 are the purchasing rates observed by pageview depth. In general, visits in which more pages are viewed are more likely to result in a purchase, suggesting that purchasing probability is positively related to the number of pages viewed. This relationship is seen across a number of different datasets and has also been documented in other research (Bucklin and Sismeiro 2001). However, in this paper, we will show that it is not just the number of pages viewed but the type of pages viewed that affect purchasing behavior.

FIGURE 2 HERE (Visits by Pageview)

Therefore, we define a set of *purchase influence measures* by decomposing pageviews by type. Specifically, we believe that the number of (not the percentage of) search pages, information-related pages, category-level pages, and product-level pages will differentially affect purchasing probabilities. In addition to decomposing pageviews by type, we also differentiate between pages that are viewed for the first time during the session versus those that have been previously viewed in that session, specifically with respect to the category and product pages. This differentiates between the effects that may result from a shopper's exposure to new environmental stimuli and those resulting from the deliberation and repeat viewing of environmental stimuli previously encountered. Table 2 provides more detailed information about the purchase influence measures and how they are defined.

Method: Bayesian Treed models

One objective of this paper is to examine the effects of in-store navigational experiences on buying behavior. Therefore, we can model purchasing incidence as a logit function of the *purchasing influence measures*. Conceptually, each page that the shopper encounters during a given store visit will provide information or stimulus that contributes to the purchasing process, either positively or negatively. The additive effect of the pages drives the final decision of whether or not to buy. But in addition to purchasing influence measures, we also include measures of customer history as explanatory variables in the logit model, such as the number of past visits (PASTVIS) as well as the number of past purchases (PASTPUR) made by the shopper prior to the current visit. Previous research has shown that historical visiting and purchasing behavior can strongly influence future purchasing incidence (Moe and Fader 2001). The inclusion of these measures allow for such dynamics. The probability of purchase, $p(PUR_j)$, during visit, j, can then be written as:

$$p(PUR_j) = \frac{\exp\{\boldsymbol{\beta} \cdot \mathbf{x_j}\}}{1+\exp\{\boldsymbol{\beta} \cdot \mathbf{x_j}\}} \quad \text{where} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_{PASTVIS} \\ \beta_{PASTPUR} \\ \beta_{PSEARCH} \\ \beta_{PINFO} \\ \beta_{PNEWCAT} \\ \beta_{PNEWPROD} \\ \beta_{PREPCAT} \\ \beta_{PREPPROD} \end{bmatrix} \quad \text{and} \quad \mathbf{x_j} = \begin{bmatrix} 1 \\ PASTVIS_j \\ PASTPUR_j \\ PSEARCH_j \\ PINFO_j \\ PNEWCAT_j \\ PNEWPROD_j \\ PREPCAT_j \\ PREPPROD_j \end{bmatrix}$$

An equally important objective of this paper is to differentiate shoppers based on their motivations and navigational patterns. Therefore, we subdivide shoppers into segments with different navigational patterns, and within each segment, use a logit model. The coefficients of the logit model are permitted to differ across segments. Such a model is considered by CGM, who use a tree structure to divide the data into segments, and use distinct linear models within each segment. Here, we consider a logit version of this model. Although it is in general possible to use arbitrary sets of variables in the tree and in the terminal node logit models, in this context it is appealing to segment using navigational variables, and model stimulus (from pageview variables) via logit models within each segment.

Figure 3 displays a tree we have fit to the data using the method of CGM. The tree partitions the observations into disjoint groups based upon the navigational measures. Each partitioning step in the tree splits subsets of the data into subgroups depending on the values of one of the measures. For example, at the top of the tree, we see that the data is first split into three groups depending on whether the measure PAGES is 1, 2-9, or greater than 9. The subset of observations with PAGES greater than 9 is then split into two subsets corresponding to whether or not the measure CATPG is greater than .86 or not. These two subgroups are not split again. The subgroup of observations with PAGES from 2 to 9 is split into three subgroups using the measures PRODPG and PRODCAT. Thus, the tree splits the observations into six disjoint groups. In tree modeling terminology, these 6 groups are referred to as *terminal nodes* of the tree. The interior nodes of the tree correspond to measure based rules for splitting groups of observations into subsets. Note that although only four of the navigational measures are used in the tree although all eight were made available to the tree formation method.

FIGURE 3 HERE (Bayesian Tree Diagram)

The tree is chosen so that a logit model using the influence measures to predict a purchase will work well for observations within a specific terminal node. Since our tree has six terminal nodes we will fit six logit models, one for each terminal node. In particular, if all customer visits reacted the same way to the purchasing influence measures (as captured by the logit model) the tree would consist of a single "terminal node" lumping them all together. The fact that in order to fit the data well, the method of CGM grows the trees supports a basic thesis of this paper. We find subsets (or segments) of visits based on the navigational measures that respond differently to the purchasing influence measures.

It is important to note that the *treed logit* approach employed in this paper differs fundamentally from the usual application of tree-based methods to fit a binary response. In the usual approach, a tree is used to partition the observations so that within each terminal node the observations are homogeneous in the sense that the same probability of a positive response (i.e. a purchase) is assigned to each of them. In our treed logit model, observations within a terminal node are homogeneous in that the same logit model is used to describe how they respond to the influence measures. See Currim, Meyer, and Le (1988) for an example of the usual approach in which coffee brand choice is related to price, feature/display, and brand name.

Detailed description of the treed model method is found in CGM. Here we sketch the approach. First and foremost the approach is Bayesian. This entails (i) a formal definition of the parameter space (ii) a choice of prior and (iii) an approach for computing the posterior.

The parameter is denoted by the pair (T, B), where T denotes the tree and B denotes the logit coefficients. T captures the parent-child structure of the tree nodes and the partitioning rule associated with each interior node. Given T, let b denote the number of terminal nodes. Then $B = (\beta_1, \beta_2, \ldots \beta_b)$ where β_i, denotes the set of coefficients for the logit model fit the subset of observations corresponding to the i^{th} terminal node.

The prior has the form $p(T, B) = p(T)p(B \mid T) = p(T)\prod_{i=1}^{b} p(\beta_i)$. In application we first standardize the explanatory variables so that the same $p(\beta)$ prior can be used independently in each terminal node. This prior is chosen to be multivariate normal with zero mean and variance matrix proportional to the identity. Hence, all we need to choose is a single prior standard deviation for each of the coefficients in each of the terminal nodes. To put a prior on T we define a stochastic process by which a tree grows. We define the probability that a current terminal nodes splits into two children to be $\alpha(1+d)^{-\gamma}$, where d is the depth of the node (how far down the tree from the top node it is). This choice of prior allows us to specify the prior belief that sufficiently deep nodes are unlikely to be split further which, in an overall sense, puts small prior mass on “large” trees. α is the probability that the first split is made.

To calculate the posterior, we use an approximation to an integral and Markov Chain Monte Carlo. Let $p(Y \mid X, T, B) = \prod_{i=1}^{b} \prod_{j=1}^{n_i} p(y_{ij} \mid x_{ij}, \beta_i)$, where n_i is the number of observations in node i, y is the binary indicator for purchase and x denotes the vector of influence measures. Then our posterior is $p(B, T \mid Y, X) \propto p(Y \mid X, B, T) p(B \mid T) p(T)$. To calculate this formidable object we first integrate out B: $p(T \mid Y, X) \propto p(T) \int p(Y \mid X, B, T) p(B \mid T) dB$. Although this integral cannot be done in closed form, accurate approximations are obtained by the Laplace method (see for example, Gelfand and Dey (1994)). We then define a Markov Chain in the space of trees so that the stationary distribution is equal to the marginal posterior of T. Steps are proposed in the tree space and these steps are accepted or rejected using the Metropolis method so that the stationary distribution of the chain is the desired posterior (see Liu (2001)).

The use of Markov Chain to search for high posterior trees T may be viewed as a form of stochastic search. We randomly wander in the space of trees, gravitating towards those that fit the data best. The prior and Bayesian specification naturally restrain our search so that we do not tend to fit a tree that is too large. This search mechanism should be contrasted with the usual greedy algorithm (Quinlan (1983), Karalic (1993)). The greedy algorithm provides a much more limited exploration of the tree space.

The results reported in this paper were obtained using the following choice of prior. We used α=.5 and γ=1. This choice put almost all the prior probability on trees with 5 or less terminal nodes. We used a prior standard deviation of 20 for the logit coefficients, which is roughly non-informative. The Markov Chain was started at the tree with just the root node and run for 5000 iterations. The chain was started 10 times (5000 iterations each time). We selected the tree with the highest integrated likelihood $\int p(Y \mid X, B, T) p(B \mid T) dB$ amongst all those visited.

Results

We first assess the fit of the treed model in comparison to benchmark models. The logit treed model presented in this paper captures two sources of variance that have previously been unstudied. First is the differential effects of pageviews on purchasing. This is captured by the logit model itself through the use of purchasing influence measures instead of total page count as explanatory variables. Table 3 reports the results of estimating a homogenous logit model across all store sessions. In one case, only the total number of pages is used to characterize the in-store experience. The second logit model presented in the table decomposes this page count measure and models page-specific effects on purchasing. The results indicate that allowing for page-specific effects significantly improves fit (BIC=5112.1 vs. BIC=5160.9) over simply modeling purchasing as a function of the total number of pages viewed.

TABLE 3 HERE (Logit Results)

Having established that the decomposition of page by page content significantly adds to our modeling efforts, we address a second source of variance: heterogeneity in navigational behavior. The treed logit model accommodates heterogeneity across visits by grouping together visits according to measures of in-store navigational patterns and then modeling purchasing within each group. But an alternative to using a treed model for this purpose is a latent class logit model, which we will consider as a benchmark.

$$p(PUR_j) = \sum_{s=1}^{S} \pi_s \frac{\exp\left\{\boldsymbol{\beta}^{\mathbf{s}} \mathbf{x}_{\mathbf{j}}\right\}}{1+\exp\left\{\boldsymbol{\beta}^{\mathbf{s}} \mathbf{x}_{\mathbf{j}}\right\}}$$

where π_s is the share of the s^{th} segment ($0 \leq \pi_s \leq 1$, $\Sigma\, \pi_s = 1$).

Figure 4 plots lift charts for the logit treed model, simple logit model, and the two-segment[3] latent class logit model. In these lift charts, visits are rank ordered by their predicted probability of purchase and plotted along the *x*-axis with the rank of one representing the visit with the largest probability. The cumulative number of purchases represented by these visits are plotted along the *y*-axis. The perfect model would predict high purchasing probabilities for all actual purchasing visits and low probabilities for all non-purchasing visits. As a result, a perfect model would be associated with a very steep lift curve, represented by a line that goes nearly straight up. In contrast, a random assignment of purchasing probabilities would result in a 45º line as its lift curve. In other words, the steeper the curve (i.e., the more quickly the curve increases), the better the model captures the actual purchasing process. In general, accommodating heterogeneity either by using the latent class logit or the treed model improves fit over the simple logit model, though the former offers only a slight improvement. It is clear from Figure 4 that the treed model is able to detect purchasing visits more accurately than a latent class model. Additionally, the tree structure is able to capitalize on observable navigational patterns when classifying visits and is not simply grouping observations based on purchasing response to pageviews, as is the case with the latent class logit model. The gains chart plotted is for the full dataset; similar patterns are seen if the data are subdivided into train and test sets, and gains measured on the test set.

[3] A three-segment latent class model does not significantly improve model fit. BIC=4208.6 for the 2 segment model compare to BIC=4433.0 for the 3 segment model.

FIGURE 4 HERE (Lift Chart)

From the navigational patterns of each visit, the treed model effectively groups sessions in a manner consistent with many of the concepts and propositions offered in section two. Figure 3 offers a pictorial representation of the logit treed model results while table 4 provides a profile for each terminal node, in terms of the average value of each navigational measure. Table 4 also presents the logit coefficients that result from the purchasing model within each node. The final tree generated by the model was one with six terminal nodes that were segmented based only on four variables[4]. Note that even though all eight navigational measures were provided in the model estimation process, the algorithm found it only necessary to utilize four of them when branching the tree: SHOPPG, CATPG, PRODPG, and PRODCAT. These four measures best differentiate between the navigational patterns found in observed in-store behavior. Additionally, several nodes had very low numbers of purchase transactions (nodes 1, 4, and 6). The homogeneity of purchasing patterns in these nodes (i.e., no purchases are likely) suggest that the shoppers are insensitive to any purchasing model variables and are simply unlikely to buy. Therefore, we do not report any logit coefficients for these two nodes. We can, however, still test many of our propositions using the overall purchasing conversion rates.

TABLE 4 HERE (Node Profiles)

[4] The original tree that resulted from the model estimation actually contained seven nodes. Node 4 in figure 5 further splits into two children nodes based on UNIQPROD. However, these two nodes had similar navigational profiles and both had negligible purchasing activity. Therefore, in the final analysis, we merged these two terminal nodes.

Before describing the tree results in relation to the propositions, we would like to briefly discuss those shopping sessions in node 1. The defining characteristic of the visit sessions categorized in this node is that the shoppers viewed only one page. In other studies of online browsing behavior, researchers have typically dropped observations with only one pageview as these visits do not represent the behavior of interest and often skew results (Bucklin and Sismeiro 2000). In our analysis, however, we retained these observations and allowed the model to determine whether these observations were significantly unique to warrant a separate classification. As a result, node 1 was created by the model and consist exclusively of one-page visits. The fact that the model was able to identify this group of visits and keep the analysis of the remaining observations separate without researcher intervention serves as a strong testimonial to the flexibility of the model.

Proposition 1 argues that goal-directed shoppers are more likely to buy than exploratory shoppers. But before we can determine whether this is the case, we need to identify each terminal node as either goal-directed or exploratory. Recall that we define goal-directed shoppers as those with a high degree of focus centered around a product purchasing decision while exploratory shoppers navigate through the store site with a highly varied set of product considerations. With this in mind, we categorize each terminal node as either goal-directed or exploratory based on their navigational profiles.

Node 2 shoppers examine a high percentage of category pages (CATPG=0.19) relative to product pages (PRODPG=0.09) with a moderate degree of variety (UNIQCAT=0.22; UNIQPROD=0.23) in both. Additionally, very few products of the same category are viewed

(PRODCAT=0.20). All this suggests that these shoppers focus their attention primarily on category level summaries of products rather than specific product details. Furthermore, when product details are viewed, they are for products from a variety of different categories implying that the shopper is not focused on any specific product decision. Therefore, we conclude that node 2 visits appear to be exploratory in nature.

Node 3 shoppers stand in sharp contrast to those in node 2. These shoppers view more product pages (PRODPG=0.46) than category pages (CATPG=0.34). Additionally, almost all products viewed are unique (UNIQPROD=0.95) and come from a limited variety of product categories (PRODCAT=2.10). This behavior reflects a shopper that is focused around a particular product decision. That is, they appear to be examining product specific information on a number of different products within a limited set a product categories as if constructing a consideration set. Therefore, we conclude that node 3 shoppers are goal-directed in nature.

We examine the navigational profiles of the remaining nodes in a similar manner and conclude that nodes 4 and 5 consist also of goal-directed shoppers. Shoppers in both nodes seem to focus much of their attention on a high variety of product pages but limited within a small variety of categories. We classify node 6 shoppers as exploratory with the high level of category page views (CATPG=0.95) coupled with a low product-to-category ratio (PRODCAT=0.07).

Comparing the goal-directed shoppers (nodes 3-5) to the exploratory shoppers (nodes 2 and 6), we find that goal-directed shoppers are more likely to buy than exploratory shoppers. Those visits categorized as goal-directed have an overall purchase conversion rate of 10.17% (382 purchases, 3755 visits) compared to 1.72% (180 purchases, 10,492 visits) for the exploratory visits (p-value[5] < 0.001). This result supports **proposition 1**.

In addition to the overall difference in purchase conversion rates between goal-directed and exploratory shoppers, there are also difference within each group. For instance, Figure 1 suggested earlier that each type of shopper goes through stages in the purchasing process. As a result, nodes characterized as goal-directed (or exploratory) will differ from one another depending on the stage of the purchasing process. These differences are reflected both in the navigational profiles of these store visits as well as the model coefficients that describe purchasing influences.

We start by more closely examining goal-directed searchers (nodes 3-5). Among these shoppers, those in node 3 search across the highest variety of products (UNIQPROD=0.95, PRODCAT=2.10). The navigational behavior of shoppers in this node reflect behavior typical of a shopper searching across a variety of products (but within a pre-specified product category) in order to form a consideration set. This behavior suggests that node 3 shoppers are *early stage* goal-directed searchers. This is confirmed by a low level of repeat product viewing (MAXREP=0.17) indicating that the shopper has yet to find a product worth revisiting and deliberating. In contrast, node 5 shoppers exhibit a much higher rate of repeat product viewing (MAXREP=0.75). This reflects increased deliberation of a specific product that may lead to a final purchasing decision. Node 5 shoppers are more likely to be in a later stage of the decision process than those in node 3. This is further illustrated by a lower product-to-category ratio

[5] This and all subsequent reported p-values are conditional on the selected tree structure. Although this ignores selection bias, such p-values can still serve as a rough guide for assessing significance.

indicating that node 5 shoppers are not considering as many different products in the decision category as demonstrated by node 3 shoppers. Therefore, according to **proposition 2**, we should expect node 5 shoppers to exhibit a higher purchasing rate than node 3 shoppers. In fact, we find that node 3 shoppers only buy 4.52% of the time whereas node 5 shoppers buy 12.74% of the time, supporting **proposition 2** (p-value < 0.001).

Purchasing influences also vary between early stage and late stage goal-directed shoppers. **Proposition 3a** argues that early stage shoppers would be positively affected to purchase by any experience that helps them construct a better consideration set. Consistent with **proposition 3a** we find that node 3 shoppers are more likely to buy if they have completed a successful search and see are exposed to a search results page ($\beta^3_{PSEARCH}$=22.19).

Late stage goal-direct shoppers, on the other hand, are in deep deliberation and should be positively affect by repeat viewing product information, according to **proposition 3b**. The purchasing model results seem to support this proposition. In terms of the navigational experience within the visit, node 5 shoppers are positively affected by repeat viewing of both category and product pages ($\beta^5_{PREPCAT}$=1.17, $\beta^5_{PREPPROD}$=3.13). New category and product pages have no significant effect on purchasing.

In addition to those sessions grouped into nodes 3 and 5, store visits in node 4 can also be classified as goal-directed with a high percentage of product page viewings (PRODPG>0.99) and product-to-category ration (PRODCAT=1.26). However, unlike the other two nodes, these shoppers have a very low purchasing conversion rate (0.90%). This seems to contradict our beliefs that goal-directed shoppers are more likely to buy. But upon closer examination of this node, we find that store visits in this node are not typical. Shoppers in nodes 3 and 5 view an average of 20.40 pages for approximately 125 seconds per page. In contrast, node 4 shoppers view an average of 2.97 pages for 372 seconds per page. In addition, these 2.97 pages are focused entirely on product level pages (PRODPG=1.00). This leads us to suspect that these visits are repeat visits where the shopper had previously identified a specific product and returned to it using a bookmark. The large amounts of time spent per page suggests a very difficult deliberation process on the part of the shopper. The fact that these return visits exhibit such low conversion rates contradict prior research suggesting that returning visitors are more likely to buy than first time visitors (CNET News.com 2001). One explanation may involve the nature of the product category sold by the online retailer in our study. Vitamins and nutritional supplements are not necessarily high involvement products that require multiple visits to evaluate. Those visitors in node 4 may have identified the product in a previous visit, deliberated upon it, and left, for whatever reason, without purchasing it. This behavior indicates a lower baseline probability of buying for these shoppers. This is what is reflected in the low conversion rate for node 4. All else being equal, a returning customer may be more likely to buy. However, these node 4 shoppers had their reasons for leaving without buying in their last visit, and these reasons are sure to carry over to the return visit.

Exploratory shoppers also vary depending on their stage in the purchasing process. In general, exploratory shoppers tend to navigate the store site with a focus on brief category level descriptions of products. However, if an exploratory shopper finds an item of interest at this level, the shopper will progress to the next stage and more closely examine the product by viewing product level pages. Node 2 shoppers view slightly more product level pages than node 6 shoppers (PRODPG=0.09 compare to PRODPG=0.01), indicating that node 2 shoppers have progressed to a later stage of the exploratory shopping process. **Proposition 4** argues that late stage exploratory shoppers are more likely to buy than early stage exploratory shoppers. Our data does not support this argument. The difference in conversion rates between nodes 2 (CR=1.72%) and 6 (CR=1.82%) is not significantly different (*p*-value=0.4779). However, the differences in navigational behavior across these two nodes are slight and node 6 contains a very low number of observations. As such, it is not a sufficient test of the proposition.

Proposition 5 suggests that the more exposure an exploratory shopper has to products in the store, the more likely he/she will be to buy. The purchasing model coefficients for node 2 support this proposition. We can see that many of the pages containing product related information have strong positive effects on purchasing ($\beta^2_{PREPCAT} = 17.58$, $\beta^2_{PNEWPROD} = 21.86$, $\beta^2_{PREPPROD} = 27.60$) whereas utilitarian pages that do not offer direct product exposure have negative effects ($\beta^2_{PSEARCH} = -11.07$, $\beta^2_{PINFO} = -25.28$)[6]. **Proposition 6** further argues that late stage exploratory shoppers are particularly influenced to buy when reviewing previously see information as part of a deliberation process. Again, this proposition is supported by a highly significant and positive coefficient for the number of repeat product pages viewed ($\beta^2_{PREPPROD} = 27.60$). In fact, PREPPROD has a stronger effect on purchasing than any other measure used in the logit model for node 2. However, because of the low observation counts in node 6, we cannot compare this effect found among late stage exploratory shoppers to early stage

[6] Though search pages provide a list of items that fit the user's search criteria, it provides only the product name as a line item and does not offer any picture of the product or pricing information, unlike category level pages.

exploratory shoppers. In general, node 6 shoppers are not likely to buy and are therefore unlikely to be swayed by any purchasing influence measure.

Discussion

In this paper, we proposed a framework with which to examine online shopping behavior in terms of in-store navigational patterns and how in-store experiences affect the shopper's purchasing decision. Specifically, online shopping behavior can be described along two dimensions: (1) goal-directed versus exploratory search behavior and (2) stage of decision process. We also introduced a number of new measures that capture patterns in navigational clickstream data. These measures are derived from a data set that characterizes each pageview by type and not by the cryptic URLs that only webmasters deem useful. These measures can provide marketers and marketing researchers with clickstream data that is significantly more interpretable, as we illustrated in this paper. Though this type of data is not commonly available, it is easy to collect, and we encourage researchers and online marketers to do so.

However, the main contribution of this paper is to offer a method with which to model and interpret the diverse range of in-store behavior and their respective purchasing behavior. To this end, we present a Bayesian treed model that segments customers based on their navigational behavior and models the unique purchasing decision within each segment. So in addition to differentiating between store visits, this paper provides a tool with which to examine how the in-store experience, in terms of pages viewed, affects purchasing for a variety of shopping visits.

We find that navigational factors can drive how the model segments visits, thereby supporting the need to differentiate online behavior based on in-store navigation and underlying shopping motives (e.g., goal-directed versus exploratory search). Additionally, the in-store factors that influence purchasing differ dramatically across visits, again depending on the visits' underlying motivations. Therefore, it is imperative in future studies of online shopping behavior to consider and accommodate these varied behaviors.

The fact that store visits differ so significantly in terms of shopping motivations and purchasing response has dramatic implications for how online marketers may wish to communicate with their shoppers. For example, our results show that late-stage goal-directed shoppers are positively influenced to purchase by repeat viewing of product and category information. Consequently, marketers should seek to identify these shoppers, using the navigational measures proposed in this paper, and encourage them to review previously encountered product information. Early-stage goal-directed shoppers seek to construct their consideration set and will be more likely to buy if provided a tool to do so. Thus, marketers would be well served to identify these shoppers and provide help to identify potential products for purchase. Exploratory shoppers, on the other hand, respond to exposure to new information. Therefore, the challenge for online retailers with respect to their exploratory shoppers is to keep them at the site in order to present them with a variety of new products.

In general, goal-directed shoppers tend to respond to utilitarian information that helps them make a better decision whereas exploratory shoppers are more likely to respond to new and interesting stimuli. Any marketer who wishes to target a marketing message to their shoppers must be able to differentiate between them and understand how their purchasing decision is influenced. The treed model presented in this paper does just that. However, though this paper models the purchasing response to in-store experiences in terms of pageviews, it does not address how shoppers will respond to different types of promotional messages. One would expect that the various types of store visits will respond very differently not only to pageviews but also to any promotional message with which they may be presented. Therefore, we encourage future research to experiment with promotional messages that vary for different types of store visits. With this effort, there is the potential to develop customized marketing campaigns that will present each shopper with the most effective message.

References

Babin, Barry J., William R. Darden, and Mitch Griffin (1994), "Work and/or Fun: Measuring Hedonic and Utilitarian Shopping Value," *Journal of Consumer Research*, 20 (4), 644-656.

Barsalou, Lawrence W. (1991), "Deriving Categories to Achieve Goals," in *The Psychology of Learning and Motivation*, vol. 27, ed. Gordon H. Bower, New York: Academic Press, 1-64.

Brownlow, Mark (2001), "Conversion Rate," ibizBasics.com (October 23), http://www.ibizbasics.com/ecom012301.htm.

Brucks, Merrie (1985), "The Effects of Product Class Knowledge on Information Search Behavior," *Journal of Consumer Research,* 12 (1), 1-16.

Bucklin, Randy and Catarina Sismeiro (2000), "A Model of Website Browsing Behavior Estimated on Clickstream Data," *Working Paper presented at the Marketing Science Conference at UCLA.*

Brucks, Merrie (1985), "The Effects of Product Class Knowledge on Information Search Behavior," *Journal of Consumer Research,* 12 (1), 1-16.

Chipman, Hugh, Edward I. George, and Robert E. McCulloch, R. (2002), "Bayesian Treed Models," Machine Learning, 48(1), 299-320.

Currim, Imran S., Robert J. Meyer, and Nhan T. Le (1988), "Disaggregate Tree-Structured Modeling of Consumer Choice Data," *Journal of Marketing Research*, 25 (August), 253-265.

Gelfand, A. E. and Dey, D.K. (1994), "Bayesian Model Choice: Asymptotics and Exact Calculations", *Journal of the Royal Statistical Society B*, 56,3, 501-514.

Hirschman, Elizabeth C. (1984), "Experience Seeking: A Subjectivist Perspective of Consumption," *Journal of Business Research*, 12 (March), 115-136.

Hoffman, Donna L. and Thomas Novak P. (1996), "Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations," *Journal of Marketing*, 60(3), 50-68.

Jarboe, Glen R. and Carl D. McDaniel. (1987). "A Profile of Browsers in Regional Shopping Malls," *Journal of the Academy of Marketing Science*, 15 (Spring): 46-53.

Janiszewski, Chris (1998), "The Influence of Display Characteristics on Visual Exploratory Search Behavior," *Journal of Consumer Research,* 25 (3), 290-301.

Karalic, A. (1992), "Employing Linear Regression in Regression Tree Leaves," in Proceedings of ECAI-92, John Wiley & Sons.

Liu, J. S. (2001), "Monte Carlo Strategies in Scientific Computing", Springer-Verlag, New York.

Mandel, Naomi and Eric Johnson (2000), "Priming, Search, Expertise and the Construction of Preferences: Can Web Pages Change Choice?", *Working Paper, Arizona State University.*

Kardes, Frank R., Gurumurthy Kalyanaram, Murali Chandrashekaran, and Ronald J. Dornoff (1993), “Brand Retrieval, Consideration Set Composition, Consumer Choice, and the Pioneering Advantage,” *Journal of Consumer Research*, 20 (June), 62-75.

Moe, Wendy W. (2002), “Buying, Searching, or Browsing: Differentiating between Online Shoppers Using In-Store Navigational Clickstream," *Journal of Consumer Psychology (forthcoming).*

Moe, Wendy W. and Peter S. Fader (2001), “Which Visits Lead to Purchases?: A Study of Dynamic Conversion Rates,” *Working Paper, Wharton Marketing Department.*

Novak, Thomas P., Donna L. Hoffman, and Yiu-Fai Yung (2000), “Measuring the Customer Experience in Online Environments: A Structural Modeling Approach,” *Marketing Science*, 19 (1), 22-42.

Putsis, William P., Jr. and Narasimhan Srinivasan (1994), "Buying or Just Browsing? The Duration of Purchase Deliberation," *Journal of Marketing Research*, 31 (August), 393-402.

Quinlan, J. Ross (1983), “Learning Efficient Classification Procedures and Their Application to Chess End Games,” in *Machine Learning: An Artificial Intelligence Approach,* R.S. Michalski, J.G. Carbonell, and T.M Mitchell, eds. Palo Alto, CA: Tioga Publishing, 463-482.

Ratneshwar, S., Cornelia Pechmann, and Allan D. Shocker (1996), “Goal-Derived Categories and the Antecedents of Across-Category Consideration,” *Journal of Consumer Research*, 23 (December), 240-250.

Sherry, John F., Jr., Mary Ann McGrath, and Sidney J. Levy (1993), "The Dark Side of the Gift," *Journal of Business Research*, 28 (November), 225-245.

Shocker, Allan, Moshe Ben-Akiva, Bruno Boccaro, and Prakash Nedungadi (1991), “Consideration Set Influences on Consumer Decision-Making and Choice: Issues, Models and Suggestions,” *Marketing Letters*, 2(3), 181-197.

Wilkie, William L. and Peter R. Dickson (1985), "Shopping for Appliances: Consumers' Strategies and Patterns of Information Search," in Perspectives in Consumer Research, ed. Harold H. Kassarjian and Thomas R. Robertson, Englewood Cliffs, NJ: Prentice Hall, 1-26.

Figure 1. *The Purchasing Process*

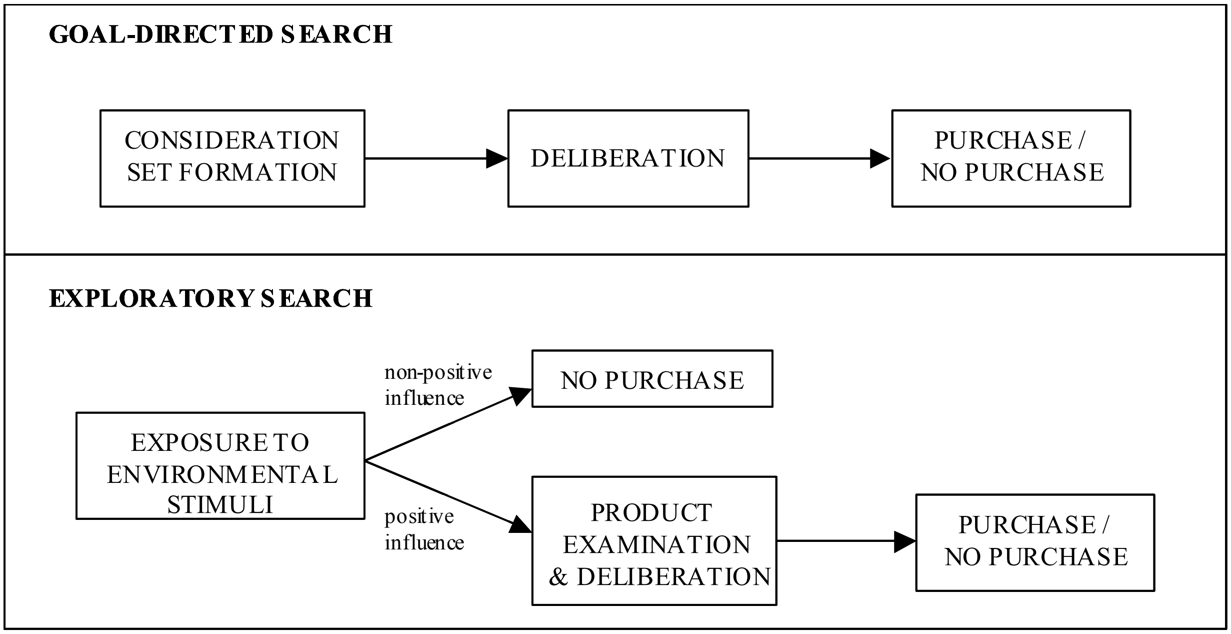

Figure 2. *Visits by Pageview*

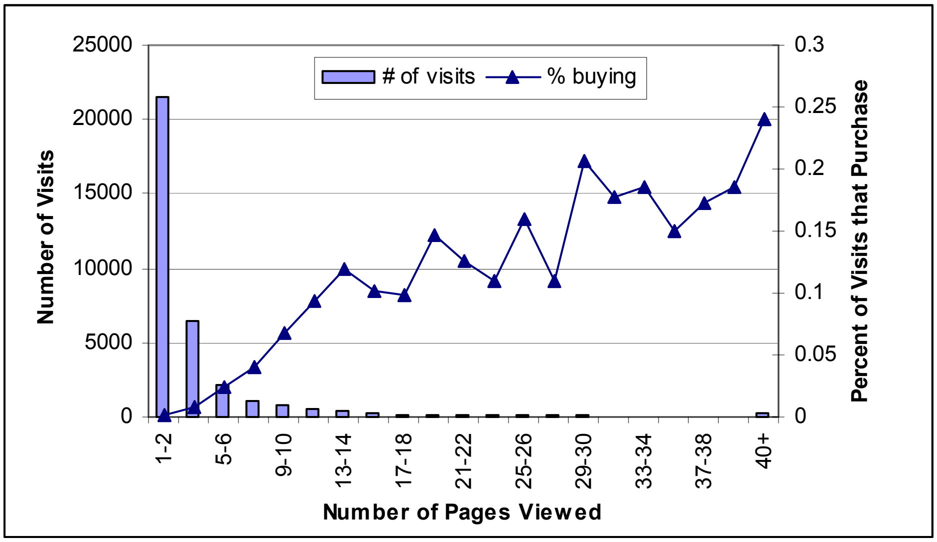

__Figure 3.__ Logit Treed-Model Results

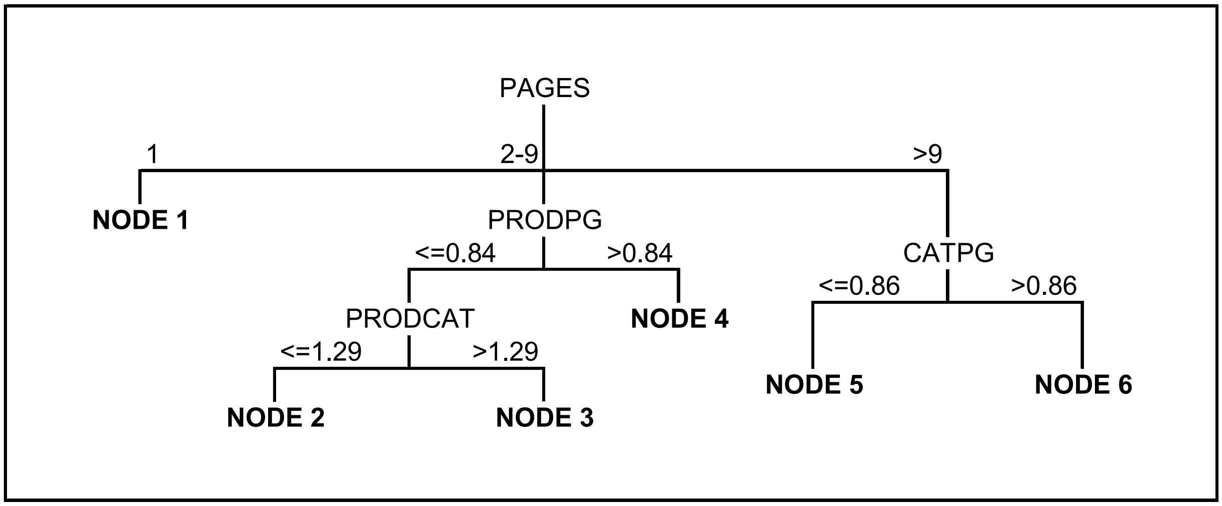

Figure 4. *Lift Chart*

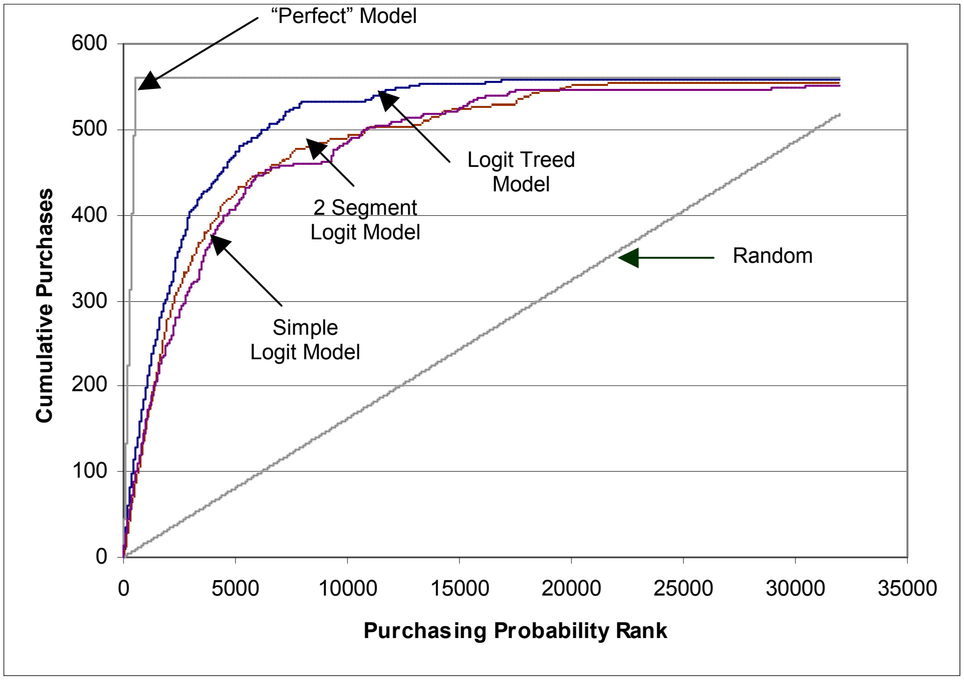

Table 1. *Page Categorizations*

<table>
<tr>
<td rowspan="2"><u>Administrative pages</u>
• Registration pages
• Transaction related pages (including shopping cart pages, shipping address pages, credit card information pages, etc.)</td>
<td><u>Shopping pages</u>: Product -related
• Home page
• Category pages
• Brand pages
• Product pages
• Search result pages</td>
</tr>
<tr>
<td><u>Shopping pages:</u> Information-related
• Company information (including privacy policy, delivery time, shipping and handling costs, etc.)Community area
• Advice columns</td>
</tr>
</table>

Table 2. *Summary of Measures*

NAVIGATIONAL MEASURES *(used as tree splitting variables)*	**PURCHASING INFLUENCE MEASURES** *(used as explanatory variables in logit models)*
General PAGES – total number of pages viewed PGTIME – average time spent viewing each page	**General** PSEARCH – number of search pages viewed PINFO – number of information-related pages viewed
Goal-Directed Vs. Exploratory Focus Measures: CATPG – % of shopping pages viewed that are category level pages PRODPG – % of shopping pages viewed that are product level pages Browsing Variety Measures: PRODCAT – ratio of the number of unique products viewed to the number of unique categories UNIQCAT – % of category level pages viewed that were unique UNIQPROD – % of product level pages viewed that were unique	**Exposure to New Stimuli** PNEWCAT – number of category pages viewed for the first time PNEWPROD – number of product pages viewed for the first time **Deliberating Previously Viewed Stimuli** PREPCAT – number of category pages that are repeat viewed PREPPROD – number of product pages that are repeat viewed
Stage in Purchasing Process MAXREP – maximum number of repeat product page viewings	

Table 3. *Homogenous Logit Model Results*

LOGIT MODEL with PAGE COUNT EFFECTS (LL=-2559.5, BIC=5160.9)		LOGIT MODEL with PAGE-SPECIFIC EFFECTS (LL=-2509.0, BIC=5112.1)	
Variable	Estimate	Variable	Estimate
Constant	-4.52 (0.047)	Constant	-4.619 (0.055)
Past visits	0.002 (0.001)	Past visits	0.002 (0.001)
Past purchases	0.063 (0.012)	Past purchases	0.061 (0.013)
Page count	0.082 (0.004)	Search pages	0.063 (0.023)
		Information pages	-0.101 (0.033)
		New category pages	0.352 (0.035)
		New product pages	0.021 (0.013)
		Repeat category pages	0.102 (0.015)
		Repeat product pages	0.141 (0.026)

* standard errors are provided in parentheses.

Table 4. *Logit Treed-Model Coefficients*[1]

	NODE 1	NODE 2	NODE 3	NODE 4	NODE 5	NODE 6
n	20338	10437	376	553	2826	55
# purchase	7	179	17	5	360	1
% purchase	0.03%	1.72%	4.52%	0.90%	12.74%	1.82%
PAGES	1.00	3.66	6.69	2.97	22.22	18.89
PGTIME	24.60	125.49	100.18	372.08	129.74	92.13
CATPG	0.14	0.19	0.34	0.00	0.30	0.95
PRODPG	0.12	0.09	0.46	1.00	0.25	0.01
PRODCAT	0.12	0.20	2.10	1.26	1.23	0.07
UNIQCAT	0.14	0.22	0.32	0.00	0.28	0.24
UNIQPROD	0.12	0.23	0.95	0.82	0.67	0.11
MAXREP	0.00	0.03	0.17	0.52	0.75	0.07
β_0		-4.16**	-5.54		-2.03**	
$\beta_{PASTVIS}$		1.65**	1.31		-0.53	
$\beta_{PASTPUR}$		2.62**	-193.63		1.32**	
$\beta_{PSEARCH}$		-11.07**	22.19**		-0.36	
β_{PINFO}		-25.28**	-381.17		-7.55**	
$\beta_{PNEWCAT}$		-1.16	0.17		0.15	
$\beta_{PNEWPROD}$		17.58**	-4.83		1.17*	
$\beta_{PREPCAT}$		21.86**	-12.06		-0.53	
$\beta_{PREPPROD}$		27.60**	-9.43		3.13**	

** significant at $p<0.01$

* significant at $p<0.05$

[1] Coefficients for nodes 1, 4, 6 are not reported because the small number of purchasers in these nodes means that estimation of a logit model is unrealistic