JOURNAL OF CONSUMER PSYCHOLOGY, *13*(1&2), 29–39

Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream

Wendy W. Moe
Department of Marketing
University of Texas at Austin

In the bricks-and-mortar environment, stores employ sales people that have learned to distinguish between shoppers based on their in-store behavior. Some shoppers appear to be very focused in looking for a specific product. In those cases, sales people may step in and help the shopper find what they are looking for. In other cases, the shopper is merely "window shopping." The experienced sales person can identify these shoppers and either ignore them and let them continue window shopping, or intercede and try and stimulate a purchase in the appropriate manner. However, in the virtual shopping environment, there is no sales person to perform that role.

Therefore, this article theoretically develops and empirically tests a typology of store visits in which visits vary according to the shoppers' underlying objectives. By using page-to-page clickstream data from a given online store, visits are categorized as a buying, browsing, searching, or knowledge-building visit based on observed in-store navigational patterns, including the general content of the pages viewed. Each type of visit varies in terms of purchasing likelihood. The shoppers, in each case, are also driven by different motivations and therefore would respond differentially to various marketing messages. The ability to categorize visits in such a manner allows the e-commerce marketer to identify likely buyers and design more effective, customized promotional message.

The widespread availability of Internet clickstream data has contributed greatly to marketing research. Specifically, it has allowed marketers, both practitioners and academics, to examine consumer search behavior in a large-scale field setting. The result has been a cornucopia of research that models and tests theories of online shopping behavior. These articles have addressed *interstore* comparisons across multiple online retailers as well as *intrastore* behavior, both on a visit-to-visit basis (examining store visit and purchase decisions over time) and on a page-to-page basis (examining navigation within a single store-session).

Most of the interstore studies have focused almost exclusively on search behavior (Bakos, 1997; Brynjolfsson & Smith, 2000; Johnson, Moe, Fader, Bellman, & Lohse, 2000; Lynch & Ariely, 2000). Studies of consumers' intrastore behavior over time more closely examine and predict purchasing behavior (Fader & Hardie, 2001; Moe & Fader, 2000a, 2000b). A few attempts have also been made to examine search behavior within a given store-session rather than across sessions over time (Bucklin & Sismeiro, 2000; Hoffman & Novak, 1996; Novak, Hoffman, & Yung, 2000). However, the research in this area is limited and does not take into account the content of pages viewed.

Enormous potential exists in studying an individual's behavior as they navigate from page to page. Hoffman and Novak (1996) proposed a concept of *flow* in describing the general customer experience online. Bucklin and Sismeiro (2000) later developed a model of page views in terms of the number of pages viewed and the duration of each page view. However, neither of these studies takes into account the content of the pages viewed or examines the relationship between these navigational decisions and purchasing. Mandel and Johnson (2002) showed that preferences, and hence purchasing decisions, are often constructed online while navi-

Requests for reprints should be sent to Wendy W. Moe, Department of Marketing, University of Texas at Austin, 21st and Speedway, CBA 7.216, Austin, TX 78712. E-mail: wendy.moe@bus.utexas.edu

gating through the store. Therefore, the content of the pages viewed can be very important both in determining the type of shopper involved and in predicting purchases.

The objective of this article is to examine in-store navigational behavior in terms of the pattern and content of pages viewed. The goal is to better understand the objectives of the shopper, thereby providing some insight into purchasing behavior. Although we do not explicitly model the purchasing decision or any consumer response to marketing interventions, understanding the different types of store visits and the metrics that differentiate between them is a key first step toward such ventures. Therefore, this article focus exclusively on identifying the metrics that differentiate between the store visit types and their ability to predict the high purchase conversion visits.

In the next section we develop a typology of shopping strategies and discuss how we expect clickstream patterns to vary depending on the shopping strategy employed. Important to this discussion is the development of relevant session descriptors that we need to characterize the in-store patterns. Therefore, in addition to describing the e-commerce site and the clickstream data we use in this study, we also carefully define a set of session descriptor variables. Using this data, we analyze the observed visit sessions and validate our proposed typology. Finally, we identify the key metrics necessary to differentiate between store visit types and illustrate the value of these metrics by using them to predict future purchasing behavior.

TYPOLOGY OF SHOPPING STRATEGIES

Dimensions of the Typology

Search behavior has been dichotomized into goal-directed versus exploratory search (Janiszewski, 1998). Goal-directed search refers to behavior for which the consumer has a specific or planned purchase in mind. The objective of search in this case is to allow the consumer to gather relevant information for a purchase that is being considered (Brucks, 1985; Wilkie & Dickson, 1985). Search patterns, therefore, are very focused and directed toward the goal of making a purchasing decision. Exploratory search, on the other hand, refers to behavior in which the consumer is less deliberate and focused and perhaps not even considering a purchase. Instead, search tends to be undirected and stimulus-driven rather than goal-driven (Janiszewski, 1998). This type of search is also sometimes referred to as browsing or ongoing search. In many cases, this type of search results when consumers have little stored knowledge with which to process the information. Because ongoing search is not motivated by any specific decision-making need, search patterns are less focused. The consumer derives utility not necessarily from the outcome of search but rather from the visiting/shopping experience itself (Bloch, Sherrell, & Ridgway, 1986).

Both types of search behavior can potentially result in a purchase. For example, goal-directed shoppers may easily walk out of the store with a purchase once they have acquired all the information they were seeking. The nature of the purchase, in terms of involvement level, would determine the amount of information needed to make a decision and therefore, the time horizon of the purchase. Although not driven by the motivation to buy, exploratory search may also incite a purchase. For example, exploratory search tends to be stimulus-driven, and with the right the stimulus, an impulse purchase may occur. In addition, ongoing search is also motivated by the desire to acquire a bank of product knowledge potentially useful in the future (Bloch et al., 1986). As such, the activities in an exploratory search visit may contribute to future purchasing decisions. Therefore, regardless of the type of search driving the store visit (directed or exploratory), purchase may result either immediately or sometime in the future.

Table 1 provides a framework to view the different shopping strategies. Along one dimension is the type of search behavior driving the store visit: directed versus exploratory. The other dimension refers to the potential purchasing horizon. From this framework, we can classify shopping strategies into one of four categories: directed buying, search/deliberation, hedonic browsing, and knowledge building.

Directed Buying

In many instances, the shopper intends to make a purchase and is not lacking any substantial information before making that decision. Store visits, in these cases, are said to be driven by a directed-buying strategy and are likely to result in an immediate purchase. The in-store behavior is very focused and targeted toward a specific and immediate purchase. For example, grocery store visits would be considered directed-buying visits. In the case of a grocery shopping trip, consumers recognize a very specific need and visit the store with the explicit purpose of purchasing the product that meets this need. Very

TABLE 1
Typology of Shopping Strategies

	Search Behavior	
Purchasing Horizon	*Directed*	*Exploratory*
Immediate	Directed buying	Hedonic browsing
Future	Search/deliberation	Knowledge building

little information search outside of simple availability and pricing information is typically required.

Store visit strategies relating to more complex purchasing decisions may also be considered directed-buying strategies if the search process is nearing an end and very little information remains to be gathered. For example, although buying a car is a highly involved and deliberative process that may span multiple store visits, consumers may also engage in direct-buying behavior when visiting a dealership near the end of the process. In these later visits, search becomes very focused and may result in an immediate purchase.

A distinctive characteristic of directed-buying visits is the shopper's tendency to exhibit very focused search patterns, indicative of the goal-directed motivation of the shopper. As a result, more product-level pages rather than category-level pages are viewed, as category pages provide a broader level of information and product pages provide more targeted and detailed information. Specifically, shoppers will likely view pages within a limited number of products and categories. These shopping sessions are not meant to explore the market and expand the consideration set but rather to deliberate and choose one specific product. In addition, because the directed buyers are near purchasing, directed-buying strategies would involve deep deliberations as indicated by high levels of repeat product viewings.

Search and Deliberation

Like directed-buying visits, search/deliberation visits are also goal-directed with a planned purchase in mind. The difference lies in the timing of that purchase. In directed-buying visits, purchases, if any occur, result immediately as the consumer is primarily in the store with the objective of making that purchase. Search/deliberation visits, on the other hand, are motivated by a future purchase. The objective of these visits is to acquire relevant information to help make a more optimal choice. Therefore, the strategy is designed to build the consideration set and evaluate the items in the set. In other words, the searcher is in the market for a particular category of product but is still unsure of which specific product in the category to buy.

Characteristics of this strategy would include a focused search within a product category. However, because the shopper is still building the consideration set and evaluating options, search may expand across a number of different products within the given category. In other words, the product-to-category ratio is high. In contrast to the directed-buying visits, less of the store visit is dedicated to repeat viewing of product information, as the shopper is still seeking out information on the category as a whole as opposed to reviewing the information on any particular product.

Hedonic Browsing

Unlike the very goal-directed behavior seen in the directed-buying and search/deliberation strategies, hedonic browsing is dominated by exploratory search behavior. These store visits are motivate less by the utilitarian motives of making the better purchasing decision and more by the hedonic utility derived from the in-store experience (Babin, Darden, & Griffin, 1994; Hirschman, 1984, Sherry, McGrath, & Levy, 1993). In store behavior, in these cases, tends to be more stimulus-driven and occasionally results in impulse buying, depending on the nature of the stimuli encountered (Janiszewski, 1998; Jarboe & McDaniel, 1987).

In store behavior for hedonic browsers is significantly less focused, and therefore, more of the session is spent viewing the broader category level pages than the product level pages. In addition, because hedonic utility is derived by exploring and encountering new stimuli during these visits, hedonic-browsing sessions should exhibit a lot more variety, both in terms of the products and categories viewed, than the goal-directed strategies of directed buying and search/deliberation. Furthermore, because the hedonic browser is driven by exploration rather than deep product evaluations, we expect to see very little repeat product viewings and limited drill-down[1] on a particular product. Instead, browsing sessions are less focused and may contain more category-level pages that provide a broader array of information.

Knowledge Building

Finally, another motivation of exploratory search is that of acquiring a bank of relevant product information potentially useful in the future. In these situations, search patterns are still exploratory in nature but the utility derived from the experience is utilitarian rather than hedonic. That is, the shopper's objective is to increase product and/or marketplace expertise. The consumer is not necessarily considering any specific purchase, but the information acquired may influence future purchasing decisions.

Because of their desire to acquire general product knowledge, knowledge-building shoppers will tend to focus more on informational pages in an online shopping session. For example, many online retailers provide product-related content on their site through community discussion areas, advice columns, and so forth. Knowledge-building shoppers will be more likely to visit these areas of the site. Because their objective is to learn, knowledge-building shoppers tend to also spend more time processing informational content on the site than those shoppers who are just browsing, for example. Thus, we also expect these shoppers to have longer page-view durations.

Table 2 summarizes the characteristics and patterns of store visits that we expect to see under each strategy. In the next section, we describe the clickstream data we use and the

[1]Product drill-down for the e-commerce site studied in this article is equivalent to clicking through on a product from the category page to obtain more detailed information on it.

TABLE 2
Expected Patterns by Shopping Strategy

	Focus of Session	Category Variety	Product Variety	Repeat Product Viewings
Directed buying	Product pages	Low	Low	High
Search/deliberation	Category and product pages	Low	High	Moderate
Hedonic browsing	Category pages	High	High	Low
Knowledge building	Information pages	Low	Low	Low

measures derived from the data to characterize store visits and strategies.

DATA AND MEASURES

The Store Site

The context of our study is an e-commerce site that sells nutrition products such as vitamins, weight loss aids, body-building supplements, and so forth. 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 home page, 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, and so forth. 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 to transact with the site. Last, but definitely not least, are the shopping pages. From the home page, you can drill down by different product categories (e.g., vitamins, nutrition bars, fat loss, etc.) or by brands (PowerBar, Nature's Best, ProLab, 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 the market research firm, NetConversions, and is employed by the store site. NetConversions uses cookies downloaded onto the visitor's computer when they hit 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, if a shopper is viewing a product and adds that product to their shopping cart, the data record that the shopper added something to their cart on that page. Likewise, when a shopper submits a purchase transaction, that too is recorded. Also recorded, but not used in this study, are the contents of the shopping cart in terms of SKUs and dollar value. Because NetConversions does not use the store site's customer registration information, no demographic or financial data are recorded. The data collected monitor *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 page views as described by these categories. Table 3 provides descriptions of the general page categories used in the data set: administrative, informational, and product related pages. Of particular interest are the product related pages. Online stores are typically organized, by product category or department. In the case of the nutrition

TABLE 3
Page Categorizations

Administrative Pages	Product-Related	Informational
Registration pages	Home page	Company information (including privacy policy, delivery time, shipping and handling costs, etc.)
Transaction related pages (including shopping cart pages, shipping address pages, credit card information pages, etc.)	Category pages	Community area
	Brand pages	Advice columns
	Product pages	
	Search result pages	

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 product page. In addition, each product page is subtyped by the product's category and brand. Stores, including this one, may also organize by brand where a list of all products of a particular brand is provided. This type of page is labeled a brand page and, like the category page, also allows shoppers to click through to a product page. Finally, shoppers can jump directly to a product page as a result of a search. Because the search engine is available on all pages of the site, the resulting page of matches is coded as the search page.

Data Summary

The data used in this study spans 7 weeks from May 18, 2000 to July 5, 2000. In this time, 7,143 visit sessions were made by 5,730 unique visitors. This constituted over 36,000 lines of page-view data. We aggregate this page-to-page data to the session level using measures that we describe in the next section. These measures are designed to describe the nature of the page views within each session. Of these 7,143 sessions, only 89 (78 unique buyers) were accompanied by a purchase. This results in a 1.25% purchase conversion rate that is fairly typical in e-commerce retailing.[2] The average visit contained 4.8 page views lasting for a total of 547 seconds.

CATEGORIZING SHOPPING SESSIONS

Measures

The goal of this article is to categorize shopping sessions and not necessarily to map out the page-to-page decisions of the shopper. Therefore, we use the page-to-page information to develop session-level measures that characterize the content, in terms of behavior, of the store visit. We develop three general categories of variables: *session* measures, *variety* measures, and *before transaction* measures.

Session measures provide a general overview of the visit in terms of how much time was spent at the site, how many pages were viewed, and so forth. Table 4 provides a detailed description of the session measures. We start with a measure of visit duration, PAGES. PAGES is simply the number of nonadministrative pages viewed by the shopper. We omit administrative pages from the analysis for several reasons. First, the number of administrative pages in a session is highly correlated with purchase. If an individual chooses to purchases, they will, by definition, view several transaction/administrative pages to complete the transaction. This correlation will bias any purchase analysis we do. Second, many individuals come to the site simply to check up on an order status or to update their personal profile, nothing more. These visitors are not the type of shoppers in which we are interested. Therefore, by defining PAGES as nonadministrative pages only, we have a way of differentiating these visitors from the shoppers of interest. A second session measure that we use is PGTIME, defined as the average number of seconds spent per page. This variable provides a sense of how deeply the shopper reads each page. Both PAGES and PGTIME reveal the level of involvement the shopper has in a given session. Many visitors are not sincere shoppers and only stay at the site's home page for a brief amount of time, perhaps simply to satisfy their curiosity regarding the site. The PAGES and PGTIME measures allow us to differentiate these sessions from more promising shopping sessions.

Once we describe the overall session in terms of its duration with PAGES and PGTIME, we can then examine more closely the content of pages viewed in each segment. We focus on six page types: home page (HOME), information related pages (INFOREL), search pages (SEARCH), category pages (CATPG), brand pages (BRANDPG), and product pages (PRODPG). These measures reflect the focus of the store visit. Is the shopper interested in scanning across category pages or are the more interested in honing in on the more detailed product-specific information offered on product pages. We define these variables as the percentage of all

TABLE 4
Summary of Measures

General session measures	
PAGES	Total number of shopping pages viewed
PGTIME	Average time spent per page
Session focus measures	
HOME	% of pages that were home pages
INFOREL	% of pages that were information related pages
SEARCH	% of pages that were search result pages
CATPG	# of pages that were category level pages
BRANDPG	# of pages that were brand level pages
PRODPG	# of pages that were product level pages
Variety measures	
DIFFCAT	Category variety measure: % of category pages that were unique
DIFFBRAND	Brand variety measure: % of brand pages that were unique
DIFFPROD	Product variety measure: % of product pages that were unique
PRODCAT	Average number of unique products viewed per category
Repeat product viewing measure	
MAXREP	Maximum number of times any one product page was viewed
Transaction variable	
PURCH	Indicator variable for any completed purchase transaction during the visit session

[2]Forrester reported that over 70% have less than a 2% conversion rate in 1999.

nonadministrative pages (PAGES) spent on each of type of page, regardless of whether the visitor is repeat viewing a particular page. For example, let us say that a visitor's session starts at the home page, continues on to the vitamin category page, and then to vitamin product A. The visitor then returns to vitamin category page to drill down to a different product, vitamin product B. We would characterize this session as having five pages, 20% of which were home pages, 40% of which were category pages, and 40% of which were product pages.

Although the session measures previously mentioned provide a general overview of the session in terms of the number of pages viewed by page type, it does not reflect the content of the page views in terms of the different products and categories being viewed. Therefore, we also calculate a number of variety measures to capture behavior such as repeat viewing of a product, browsing between categories, and so forth. Table 4 also provides a list of the variety measures and how they are defined. The main objective here is to develop measures that reflect browsing across categories, brands, or products as well as repeat viewing of products. Therefore, DIFFCAT, DIFFBRAND, and DIFFPROD represents the percent of all pages of a given page type that are unique. In the example previously mentioned where the shoppers views both product A and product B within the same category, DIFFCAT would be 50% (1 unique category page ÷ 2 category pages) whereas DIFFPROD would be 100% (2 unique products ÷ 2 product pages). In general, the higher the measure, the more variety in category, product, or brand is viewed.

We also have two variables that measure the breadth and depth of search at the site. We mentioned that search/deliberation sessions involve the shopper building the consideration set by viewing multiple different products in the same product category. Browsers, on the other hand, may be jumping from product to product. In addition, directed buyers may be repeat viewing a particular product page multiple times in the final deliberation stage before buying. Therefore, measure these behaviors, we include two more variables: the number of unique products per category (PRODCAT) and the maximum number of repeat product viewings (MAXREP).

Finally, transaction activity is captured by a purchasing indicator variable for each session (PURCH). We use the session and variable measures to assign sessions to the different shopping strategies and then use purchasing activity to validate our typology.

Linking the Session and Variety Measures to Expected Patterns

Because shoppers navigate through the store site with different strategies, each individual's store visit differs in terms of the session and variety measures that describe it. But how does this manifest itself in the session and variety measures we have developed? With respect to the session variables, an important distinction to make is that between category and product page types in terms of the depth and detail available on each. Category pages tend to provide an overview of products in that category without offering any detailed information. This would be of limited usefulness for someone who needs deeper and more focused information to make a purchasing decision, as is the case with the directed buyer. The product page better offers this level of detail. Therefore, directed buyers are likely to devote more of their session to product pages (PRODPG) than category pages (CATPG), especially in comparison to a hedonic browser.

In addition, variety measures also reflect this focused behavior when present. For example, directed buyers have sufficiently narrowed down their decision choices within a category. Therefore, there should be very little variety in the categories (DIFFCAT) and products (DIFFPROD) being viewed. In addition, the product to category ratio (PRODCAT) should also be low, as they are not in search of a product but rather are deeply considering a limited number of finalists. Further, because directed buyers are near the decision process, many are likely to be deliberating on a single item. This behavior can be reflected by repeated viewings of that product page. Therefore, we would expect to see directed buyers with a relatively high MAXREP variable.

Searchers, on the other hand, are less likely to buy in the current visit session. Although they are in the market for a specific category of purchase (e.g., vitamins), they are still gathering the information needed to make the more specific product choice decision. Because searcher are exploring possible options within a category, search sessions are likely to have limited variety in the number of different categories viewed (low DIFFCAT) but more variety in the number of different products pages viewed (high DIFFPROD). This would also be reflected in a high product-to-category ratio (PRODCAT).

Both directed buyers and searchers have fairly focused navigation patterns given their goal-directed motivations. However, browsers will have much less focused behavior as a result of their exploratory nature. Unlike the directed-buyers, more page views in a hedonic-browsing session would be spent on category-level pages (CATPG) than product-level pages (PRODPG). Furthermore, browsers also tend to derive hedonic utility from experiencing new things. Therefore, we would expect the browser to seek out new products and categories to view (high DIFFCAT and high DIFFPROD). The resulting session should be one where the shopper views many category and product pages with a high level of variety within each page type, as each unique page provides more new stimulus than ones that have been previously viewed. Along those lines, we would expect to see sessions with very low levels of repeat product page views (MAXREP).

Finally, the knowledge-building shoppers are similar to the browsers in the sense that they are exploratory in nature, but they are not necessarily browsing for the sake of browsing. Instead, these shoppers' objective is to learn as much as they can about the product and the process, as the knowledge gained is the source of their utility. Therefore, knowledge-building shoppers should seek out pages with high information content, not necessarily information regarding specific products. In the case of this nutrition Web site, knowledge-building shoppers are expected to concentrate their page views on information relation pages regarding nutritional advice, community chat, and so forth (INFOREL). Because these shoppers' objective is to learn information, they should also tend to spend more time per page to process the information-rich pages. This should be reflected by a high PGTIME measure.

Cluster Analysis

We cluster analyze the store visits using the previously specified session and variety measures to uncover the different categories of shopping strategies. We expect to see at least four clusters (directed buying, search/deliberation, browsing, and knowledge-building) emerge with the patterns specified in Table 2. Although cluster analysis has been criticized for its ad hoc nature of classification, it is an empirical classification scheme that allows researchers to test theoretically developed typologies, often thought of as a fundamental role in the development of a discipline (Bunn, 1993; Hunt, 1983; Punj & Stewart, 1983; Wolf, 1926).

Before beginning our cluster analysis, we first omit any outliers that may exist, as they exert undue influence in determining the clusters and skewing the cluster centroids. Of the 7,143 store visits in the data set, 37 were omitted as outliers based mostly on extraordinarily high numbers of pages viewed and time spent per page. In addition, we also omitted 268 sessions where the visitor viewed only transaction pages. These individuals did not come to the site to shop. Their behavior, therefore, is not the target of our study.

K-means cluster analysis was performed on the remaining 6,838 sessions described by standardized session and variety measures. We examined several solutions each with varying numbers of clusters. That is, we started with a two cluster solution and increased the number of clusters until one of the following two conditions occurred: the added cluster contained an insignificant number of observations or the added cluster was virtually identical to one of the existing clusters.

The final solution was one with five clusters (see Table 5). The patterns observed are consistent with our proposed typology and the expected behaviors presented in Table 2. Cluster 1, or knowledge-building sessions, consists of those sessions devoted mostly to viewing information pages. The shopper in these sessions, dedicate a significant amount of time viewing each page in an effort to learn the information provided. Further, very little effort is made to view product related pages, re-emphasizing their desire to acquire knowledge rather than buy.

Cluster 2 shoppers, or the hedonic browsers, spend much of their time on category or product level pages, with slightly more pages dedicated to viewing category level pages than specific products. What is particularly distinctive about this cluster are the variety measures. Of the category and product pages viewed, 72.9% of the category pages and 62.7% of the product pages were unique. These shoppers were seeking out new stimuli to view. In addition, a relatively low product to category ratio (0.62) suggests that the product pages were not focused within a particular category, indicative of a hedonic browser. Further, no product pages were viewed more than once (MAXREP = 0) suggesting that these sessions did not involve heavy deliberation on any particular item.

TABLE 5
Five Cluster Solution

Cluster	*1 Knowledge Building*	*2 Hedonic Browsing*	*3 Directed Buying*	*4 Search/Deliberation*	*5 Shallow*
N	78	1,083	255	237	5,185
Maxdist	14.47	6.20	11.18	14.87	8.56
General session measures					
PAGES	3.95	5.84	18.76	26.24	2.16
PGTIME	1698.01	68.58	116.88	86.86	42.82
Session focus measures					
INFOREL	58.4%	13.8%	11.6%	13.4%	54.5%
SUMHOME	33.1%	11.9%	8.7%	6.2%	38.8%
SUMSEARCH	0.0%	1.5%	2.0%	4.8%	2.5%
SUMCAT	1.9%	38.5%	28.1%	30.6%	0.2%
SUMPROD	1.6%	28.6%	42.1%	32.3%	0.0%
SUMBRAND	3.8%	4.2%	7.7%	11.3%	2.3%
Variety measures					
DIFFCAT	7.7%	72.9%	30.7%	22.7%	0.3%
DIFFPROD	3.8%	62.7%	61.2%	93.0%	0.0%
DIFFBRAND	9.8%	12.1%	10.2%	9.7%	9.4%
PRODCAT	0.04	0.62	1.51	2.83	0.00
Repeat product viewing measure					
MAXREP	0.00	0.00	1.62	0.34	0.11

Cluster 3, or directed-buying sessions, exhibits much more focused behavior with more page views devoted to product-level information and very little variety in the product categories viewed. Furthermore, this cluster is distinctive in the high level of repeat viewing of a product page. This is consistent with the idea that direct-buying sessions are motivated by a particular product purchase and may involve heavy deliberation of a specific item. The product-to-category ratio is slightly higher than that of the hedonic-browsing sessions as shopping is confined to very few categories.

Cluster 4, or search sessions, devotes the most effort, in terms of total number of pages viewed, in to the store visit. Like the directed-buying sessions, these search session seek out very little variety in product categories because the shoppers in these cases are fairly goal-driven to find a particular type of product. But because a specific product in the category has not yet been identified, the objective of searchers is to build their consideration set by examining a number of items in the product category of interest. Hence, these shopping sessions, although exhibiting very little variety, in the number of different categories viewed, seek out a high degree of variety in products to view, most of which are contained within a given product category as indicated by the very high product-to-category ratio. Furthermore, the products are rarely viewed repeatedly (MAXREP = 0.34) as these shoppers are not making their final deliberations on a particular product but merely narrowing down their options.

Finally, a fifth cluster did emerge from the analysis, and that is a cluster of very shallow sessions. These are visitors that came to the site and viewed only two shopping pages (one of which was likely the home page) and spent very little time on each before leaving the site. These are not serious visitors in any sense. With the Internet still evolving and experiencing inflows of new users, many Internet users are simply visiting sites to see what is available rather than to shop. As the environment develops and matures, the size of this cluster should begin to shrink.

From this cluster analysis, there is empirical support that the typology of shoppers offered in section 2 does exist. To test the robustness of these five clusters, we longitudinally divide our data set into two halves. We then repeat the cluster analysis for sessions occurring in the first 24 days of data and then again for the remaining sessions. In both cases, the five-cluster analysis emerged as the optimal solution. Furthermore, the clusters that emerged exhibited the same patterns as those derived from the full data set.

Until now, we have not addressed the expected behavior of the different clusters in terms of the purchasing behavior. We have reserved those measures to validate that the clusters emerging from the analysis accurately reflect the different types of shopping strategies.[3] That is, we would expect the directed-buying sessions to have the highest purchase conversion rate because those are the shoppers who know what they want. In addition, many of these shoppers are ready to buy as a result of previous visits that they have made to the store to evaluate their options. As such, it is likely that they also have left items in their shopping cart from their last visit and have come back to purchase it.

However, because they are less certain of which specific product they want, their conversion rates should be lower than that of the directed buyers. The hedonic browsing sessions should have low conversion rates because they do not have a purchasing goal. However, their conversion should still be non-zero as a few of these browsers will succumb to impulse purchase. Finally, the knowledge-building sessions should have the lowest purchase conversion rate as their objective is to acquire information and not to shop at all. This pattern in purchasing activity can be clearly seen in the actual behavior exhibited in each type of store visit. As expected, the directed-buying sessions have the highest conversion rate (12.94%), followed by the search sessions (8.02%), browsing sessions (2.03%), and knowledge-building sessions (0.00%), respectively.[4] The search sessions will also have somewhat high conversion rates, as these are shoppers that have decided that they are in the market for a particular type of product.

PREDICTIVE METRICS

The objective of developing and defining a typology of store visits is to derive a set of metrics from a visitor's observable in-store clickstream data that can reveal the shopper's store visit objectives and ultimately purchasing tendencies. The cluster analysis in the previous section utilizes a fairly comprehensive set of measures that describe each store visit. In this section, we narrow down this comprehensive list of metrics to a smaller number of key identifying predictors. Specifically, we use just half of the data available to identify the measures that are most important in differentiating between visit types. With the remaining data, we classify each store visit and validate the classification method by comparing transaction related behavior across types.

To begin, we divide the data into two equally sized samples. A common approach would be to divide the data by time period. That is, the estimation sample would comprise those visits occurring in the first half of the period, whereas the predictive sample would comprise those visits occurring in the second half of the period. However, there are several limitations to this approach when applied to online shopping. First, the particular site that we use in this study is a fairly new site. As such, very few purchases occurred in the beginning of the data period as compared to the latter half of the data. Dividing the data based on time period would bias the

[3]We also cluster analyzed the sessions while including the purchasing variable. The PURCH variable exerted too strong an influence in formulating the clusters and effectively provided us with two groups of sessions, one that purchased and one that did not. This was an uninteresting result, and therefore the purchase indicator variable was excluded from the analysis.

[4]All differences are significant at the .05 level.

results in that purchase conversion rates for all visits in the estimation period would be lower than those in the prediction period. Second, studies have shown that in many cases, consumers build up to a purchase (Moe & Fader, 2001; Putsis & Srinivasan, 1994). In other words, consumers will make a series of nonpurchase visits before making a purchase visit. Again, in that case, later store visits will tend to have higher purchase conversions as well as more buying/late-stage search visits that earlier visits in the data period. Therefore, we take a random sample of the visits in the entire data period to use in estimation, and the remaining store visits are held out for predictive validation.

We repeat the cluster analysis on the estimation sample and find that the same clusters emerge in roughly the same proportions. We then use these cluster assignments as the dependent variable in a stepwise multiple discriminant analysis. The objective of which is to identify the key variables that differentiate one cluster of visits from another. The classification function resulting from the discriminant analysis can be written as follows:

$$Z_{js} = W_1{}^s X_1{}^j + W_2{}^s X_2{}^j \dots W_n{}^s X_n{}^j$$

where $W_n{}^s$ is the discriminant weight for variable i when calculating scores for group s, and $X_i{}^j$ describes how visit j rates on variable i. For each visit, we compare the discriminant scores across groups. The group for which the visit has the largest is the group to which the visit most likely to belongs.

To begin the stepwise discriminate analysis, we include the full set of variables used in our cluster analysis (see Table 5). Of the original 13 variables, 4 were deemed unnecessary in differentiating between store visits: SUMHOME, SUMSEARCH, SUMINFO, and DIFFBRAND as their abilities to discriminant between types of visit are insignificant. For the remaining variables, Table 6 provides the parameters of the classification functions associated with each of the five clusters. The statistic, F-to-remove, determines the relative importance of variables included in the model and is also presented in Table 6. Overall, the number of different categories viewed and the repeat viewing behavior of the shopper are most important in discriminating between shoppers (F-to-remove is 2,101.47 and 1,653.59, respectively).

Of particular interest are the differences between the hedonic browsing, the directed buying, and search/deliberation sessions. Sessions with high SUMCAT and DIFFCAT measures have high discriminant scores for the hedonic-browsing cluster. In other words, sessions that focus on a high variety of category level pages are likely to be hedonic-browsing visits. The motivations of these shoppers, and hence their response to the online environment, would vary accordingly. Those visits with lower category variety measures, compared to the hedonic browsers, *and* with high MAXREP measures would rate highly as directed-buying visits because their focus is limited in variety and centered to one product, as indicated by a high MAXREP measure. In contrast, the search/deliberation sessions can be similarly characterized similarly in terms of focus of session and variety, but diverge greatly in terms of MAXREP when compared to the directed-buying visits.

Using the classification functions described in Table 6, we then classified the remaining out-of-sample visits based on each session's calculated Z_{js}. Of the 3,419 visits in the hold out sample, 2,569 were categorized as shallow visits. These were visits that tended to be very brief in terms of both time duration and page views—not the relevant shoppers that should concern the online retailers. Of the remaining visits, we find 58.9% are classified as browsing visits, 28.6% as search visits, 7.1% as buying/late-stage search visits, and 5.6% as knowledge-building visits—mirroring the distribution of visit types seen in the estimation period.

We next examine the purchasing activity for each of the visits categories, allowing us to evaluate the effectiveness of the discriminant functions in the classification of visits and identifying purchasing propensities. We expect virtually no purchasing activity, among the shallow and knowledge-building sessions, as purchasing is not the shopper's objective in these store visits. In fact, the purchase conversion rate among these shoppers is less than 0.1%. Among the browsing visits, the conversion rate is slightly higher at 1.40%, likely the result of impulse purchases. The searchers have a significantly higher conversion rate of

TABLE 6
Classification Functions

Cluster	*F-to-Remove*	*Knowledge Building*	*Hedonic Browsing*	*Directed Buying*	*Search/Deliberation*	*Shallow*
CONSTANT		−50.42	−78.44	−196.59	−71.80	−1.77
SHOPPG	105.03	0.05	0.16	0.43	0.62	0.09
PGTIME	1240.05	0.07	0.01	0.01	0.01	0.00
SUMCAT	1238.55	9.79	93.22	94.31	84.53	0.84
SUMPROD	5.14	6.81	0.53	−9.93	−0.25	1.86
SUMBRAND	7.07	2.03	1.23	6.13	5.78	1.30
DIFFCAT	2101.47	5.28	104.77	84.66	72.06	0.14
DIFFPROD	551.15	1.44	58.34	64.38	57.60	−0.43
PRODCAT	102.82	0.49	4.06	11.35	8.43	−0.17
MAXREP	1653.59	−0.04	18.11	104.65	27.12	−0.58

6.64% ($p < .001$). Finally, the buyers/late-stage visits have the highest purchase conversion rate of 20.00%—a significant increase in purchasing probabilities over the searchers ($p < .001$).

In general, the classification scheme resulting from the cluster analysis is fairly robust across subsamples of the data. In addition, the metrics identified by the stepwise discriminant analysis and their relative importance in the classification functions provide a predictive method of categorizing visits. On examining the purchasing behavior associated with each type of visit, we find evidence validating the predictive abilities of the typology to identify the high conversion visits based only on in-store navigational clickstream.

DISCUSSION

In this article, we have developed a typology of store visits, both theoretically and empirically. We also offer measures found within clickstream data that allow us to effectively differentiate between them. For example, directed-buying sessions are likely to exhibit very focused shopping behavior targeted at a limited number of products that are often viewed repeatedly in the session. Search sessions are limited in terms of category variety, but the individuals may still be viewing a variety of different products. Browsing sessions exhibit very broad search patterns across a high variety of both categories and products. Finally, knowledge-building sessions tend to be in-depth sessions, in terms of time spent, that are concentrated around viewing information-related pages on the site. Each new store visit can be classified using the metrics identified in this article. These metrics also provide a predictive tool with which to identify the high purchasing probability directed-buying visits or the highly variable browsing visits, for example.

Although this method provides a powerful descriptive and predictive tool for online retailers, there are some data limitations to consider. For example, most clickstream data is collected and recorded as very long and mostly meaningless URLs. To employ the typology proposed in this article, online retailers must consider collecting data in such a manner in which the type of content on each page is recorded in a meaningful way. This study illustrates the potential of labeling pages as category pages and/or product pages and developing measures that reflect the focus of the store visit in these terms as well as the variety of pages viewed during a given store visit.

Further, the proportion of visitors that tend to be buyers, browsers, searchers, or knowledge builders may vary across retailers and industries. Therefore, the findings in this article should not be taken as absolute. Rather, the metrics and the methods can be applied to a variety of retailers and industries to assess the behavior of the given customer base.

With the availability of such rich clickstream data, researchers and practitioners have been striving toward a "real-time" model of purchasing that dynamically updates its evaluation, in terms of purchasing probabilities, exit probabilities, and likely promotional response, of the customer as that customer navigates the site. Any such model must address not only general customer heterogeneity, but also heterogeneity in terms of store visit types. Browsers might exhibit a larger promotional response to one type of promotion while searcher may be more affected by another. But before such experiments can be conducted and models can be developed, researchers must first understand the underlying navigational patterns of the consumer, what they mean in terms of customer motivation, and how to identify different types of behavior. This article is a first, but important, step in that direction.

REFERENCES

Babin, Barry J., Darden, William R., & Griffin, Mitch. (1994). Work and/or fun: Measuring hedonic and utilitarian shopping value. *Journal of Consumer Research, 20*(4), 644–656.

Bakos, Yannis J. (1997). Reducing buyer search costs: Implications for electronic marketplaces. *Management Science, 43*(12), 1676–1692.

Bloch, Peter H., Sherrell, Daniel L., & Ridgway, Nancy M. (1986, June). Consumer search: An extended framework. *Journal of Consumer Research, 13*, 119–126.

Brucks, Merrie. (1985). The effects of product class knowledge on information search behavior. *Journal of Consumer Research, 12*(1), 1–16.

Brynjolfsson, Erik, & Smith, Michael D. (2000). Frictionless commerce? A comparison of internet and conventional retailers. *Management Science, 46*(4), 563.

Bucklin, Randy, & Sismeiro, Catarina. (2000). *A model of website browsing behavior estimated on clickstream data.* Working Paper presented at the Marketing Science Conference at UCLA.

Fader, Peter S., & Hardie, Bruce G.S. (2002). Forecasting repeat sales at CDNOW: A case study. *Interfaces, 31,* S94–S107.

Hirschman, Elizabeth C. (1984, March). Experience seeking: A subjectivist perspective of consumption. *Journal of Business Research, 12*, 115–136.

Hoffman, Donna L., & Novak, Thomas P. (1996). Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of Marketing, 60*(3), 50–68.

Janiszewski, Chris. (1998). The influence of display characteristics on visual exploratory search behavior. *Journal of Consumer Research, 25,* 290–301.

Jarboe, Glen R., & McDaniel, Carl D. (1987, Spring). A profile of browsers in regional shopping malls. *Journal of the Academy of Marketing Science, 15,* 46–53.

Johnson, Eric J., Moe, Wendy W., Fader, Peter S., Bellman, Steven, & Lohse, Gerald L. (2001). *On the depths and dynamics of online search behavior.* Wharton Marketing Department Working Paper #00–019.

Punj, William P., Jr., and Srinivasan, Narasimhan. (1994). Buying or just browsing? The duration of purchase deliberation. *Journal of Marketing Research, 31,* 393–402.

Hunt, S. (1983). *Marketing theory: The philosophy of marketing science.* Homewood, IL: Irwin.

Mandel, Naomi, & Johnson, Eric J. (2002). When web pages influence choice: Effects of visual primes on experts and novices. *Journal of Consumer Research, 29,* 235–245.

Moe, Wendy W., & Fader, Peter S. (2000a). *Capturing evolving visit behavior in clickstream data.* Wharton Marketing Department Working Paper #00–003.

Moe, Wendy W., & Fader, Peter S. (2000b). *Which visits lead to purchases? Dynamic conversion behavior at e-commerce sites.* Wharton Marketing Department Working Paper #00023.

Moe, Wendy W., & Fader, Peter S. (2002). *Dynamic conversion behavior at e-commerce sites.* Working Paper.

Novak, Thomas P., Hoffman, Donna L., & Yung, Yiu-Fai. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science, 19*(1), 22–42.

Punj, Girish N. and Stewart, David W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research, 20, 134–148.*

Putsis, William P., Jr., and Srinivasan, Narasimhan. (1994). Buying or just browsing? The duration of purchase deliberation. *Journal of Marketing Research, 31,* 393–402.

Sherry, John F., Jr., McGrath, Mary Ann, & Levy, Sidney J. (1993, November). The dark side of the gift. *Journal of Business Research, 28,* 225–245.

Wilkie, William L., & Dickson, Peter R. (1985). Shopping for appliances: Consumers' strategies and patterns of information search. In Harold H. Kassarjian and Thomas R. Robertson (Eds.), *Perspectives in Consumer Research* (pp. 1–25). Englewood Cliffs, NJ: Prentice Hall.

Wolf, A. (1926). *Essential of scientific method.* New York: Macmillan.

Accepted by Dawn Iacobucci.