

---

# Uncovering Patterns in Cybershopping

Wendy W. Moe  
Peter S. Fader

**A**cademics and practitioners alike have been arguing about whether the Internet brings a revolutionary change in the fundamental way we do business or if it simply offers a new distribution channel and communication medium. Regardless of the answer to that debate, one thing is for sure: the Internet provides managers with an enormous amount of customer information that was previously unavailable. Thus, the new struggle has been to manage this information and to accurately and efficiently use it to somehow measure customers, trends, and performance.

Of particular interest is the role of Internet clickstream data in customer relationship management by providing a rich source of behavioral information. Until recently, the bricks-and-mortar marketers have focused their measurement and modeling efforts primarily on easily available purchasing data. Marketing research vendors in various industries (e.g., Information Resources, Inc., for customer packaged goods; SoundScan for the music industry; and IMS for pharmaceuticals) have been able to provide manufacturers and retailers with rich information regarding market share, brand switching, and new product performance. However, the spotlight has generally been centered on ultimate purchasing activity—when, what, and how much people buy. Customer activities such as comparison shopping and information gathering have largely been ignored for lack of data. These behaviors may be less visible but still exert strong influences on purchasing. As managers attempt to alter customer behavior using marketing tactics such as advertising and web site design, it is critical to reach and persuade customers at these earlier stages of the decision process.

In contrast, e-commerce managers are often faced with the opportunity of having more data than they know what to do with—from their own transaction servers, customer log files, and a wide array of other new types of primary

and secondary data sources. These “data warehouses” offer unique information on customer shopping behavior beyond just purchasing. However, there can be substantial costs to acquiring and managing this additional data; for a variety of reasons, many companies are unwilling or unable to proceed aggressively in this direction and therefore discard or ignore much of their available data.

Additionally, there has been a drastic decline in the e-commerce climate from the “irrational exuberance” experienced in the late 1990s. The current and more pessimistic view has led managers to severely cut “unnecessary” costs. The target of much of this cost cutting has been marketing—specifically, customer analysis, acquisition, and retention costs. However, in such environments, it is more important than ever to understand customers and focus on their needs in an effort to build stronger relationships that will endure. To this end, e-commerce managers should utilize their customer data to develop more efficient marketing strategies.

Hidden in the clickstream data available to many e-commerce sites is precious information that can provide better diagnostics and more accurate forecasts, allowing e-commerce sites to more profitably target and market to their customers. The key lies in the ability to better understand the data using the correct measures and models. The aim is to make the process as simple as possible, while fully acknowledging the complexities (and inherent randomness) that characterize shopping and purchasing behaviors.

There is a sharp distinction between this approach and so-called “data-mining” procedures. In data-mining methods, analysts merely “feed” the computer with dozens, perhaps hundreds, of different variables about buyers and their situational characteristics. The computer looks for patterns in the raw data with little regard for the validity (or lack) of any behavioral “stories” that might explain the associations they uncover.

In our approach, we first try to tell some appropriate stories starting at the level of the individual customer and then explicitly allow for differences across customers. Only after we have built up to an aggregate model in this fashion do we bring in the data to flesh out (and test) the behavioral assumptions we have made. If our “story” is not borne out well by the actual data, we reconsider certain assumptions and begin the process again. Our experiences in applying these methods to online visiting and purchasing patterns have been quite successful to date. This article reviews some of the patterns and methods that have emerged as dominant themes in our recent work.

## **Separating Purchasing Behaviors**

The shopping process consists of two separate behaviors. First, the customer must decide to visit the store, be it to gather information, evaluate options, or to look for something specific. Once in the store, the same customer must decide whether or not to buy during that store visit. Customer online

**FIGURE 1.** Purchasing Process

behavior can be seen as a process starting with the decision to visit a store and ideally ending with a purchase (Figure 1). Consequently, fully understanding purchasing behavior entails the analysis of several intermediate steps and is more complicated than analyzing just the purchasing events.

Over time, a customer's behavior at a store can be characterized by a pattern of visits and a pattern of purchases. However, since not all visits end in a purchase, understanding purchasing patterns is not at all equivalent to understanding visit patterns. Instead, two separate behaviors underlie purchasing: visiting and "conversion" from being a mere browser to an actual buyer. Therefore, two different models are needed to accurately capture the customer's shopping process. It is very useful and instructive to view these modeling exercises as being separate. Each sub-model has a unique set of behavioral and statistical issues associated with it and, in turn, each one provides a unique set of managerial diagnostics that can be used to properly gauge the effectiveness of marketing actions.

Consider the struggles that e-commerce managers have had in attempting to sort through the many metrics available to them in an effort to find the single best yardstick on which to focus.

Even after the enormous shakeout that began in the spring of 2000, many sites still stress visitor traffic as the primary indicator of success. Accordingly, many sites have dedicated considerable resources to increase visitor traffic, either by attracting new users with mass media advertising campaigns or by cultivating repeat visitor traffic with e-mail campaigns.

However, with many popular sites still not turning a profit despite their high visitor traffic, other managers have begun to value the conversion rate metric (percent of visits that end in a purchase). Sites that evaluate their success based on their conversion rate tend to invest in "in-store" promotions and price discounts to induce the visitor to buy. If e-commerce managers need to allocate their marketing resources between these strategies, they must be able to accurately evaluate their visitor traffic and purchase conversion rates rather than just having a vague notion of overall site performance. Treating store visiting and purchase conversion as fundamentally separate behaviors more accurately forecasts overall purchasing than when the behaviors are combined within a single, more complex, model.

Merely separating purchasing into visits and conversion is not enough to provide the most accurate forecasts. Each model needs to consider the dynamics

**TABLE 1.** Summary of Visit Data over Time

	Bookstore		CD Store	
	Months 1-4	Months 5-8	Months 1-4	Months 5-8
Total Number of Visits	5402	5899	1729	1890
Number of Unique Visitors	2693	2717	988	920
Visits/Visitor	2.01	2.17	1.75	2.05

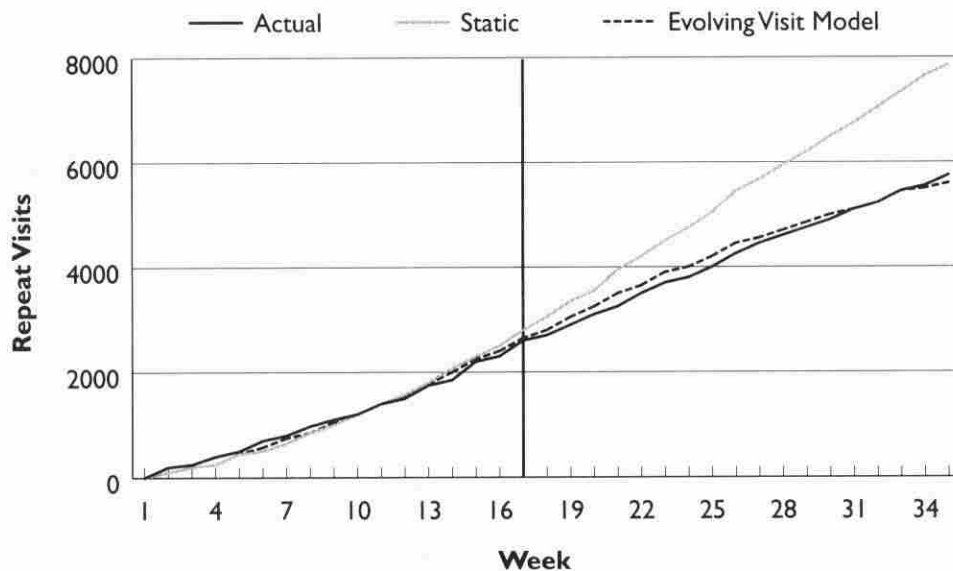
that take place as customers gain experience with a particular site. Their visits may tend to speed up or slow down, and their conversion rates will likely be changing over time as well. As a result, their behaviors, in terms of store visiting rates and purchase conversion, are systematically shifting over time. Combining this evolutionary behavior in a model that separates the two aspects of purchasing more effectively describes and forecasts customer behavior.

### Understanding Visitor Traffic at Your Site<sup>1</sup>

As illustrated in Figure 1, the first component behavior driving ultimate purchasing patterns is store visiting. One of the first steps in building a retail store, both online and offline, is to bring people to the store. Therefore, it is very important for e-commerce managers to be able to accurately measure, monitor, and forecast visitor traffic at their site. Many available metrics refer to the number of visits and visitors to a given site in a specified period of time. To monitor changes, both improvements and declines, these numbers can be compared from one period to another.

As an illustration, imagine that overall visitor traffic has been increasing from one quarter to the next, as seen in Table 1. These figures suggest that not only are the numbers of unique visitors to the site increasing, but they are also visiting the store more frequently, as indicated by the increasing visits per visitor rates. This measure suggest that over time, shoppers begin adapting to the site and visit more frequently, thereby creating more buying opportunities—a very positive development for the site. However, this is a deceptively optimistic view of visitor traffic at the site.

One problem with examining aggregate numbers of visits and visitors, such as those presented in Table 1, is the issue of a changing customer base. In the case of e-commerce, as well as other new and developing markets, new customers are constantly entering—customers who may be fundamentally different in their rates of visit than those more experienced individuals (who may be exiting the market). As a result, such aggregate summary measure could simply be reflecting the inflow of new customers and the resulting change in customer base rather than any behavior evolution among the existing customer base.

**FIGURE 2.** Model Forecasting Plot

So, how should visitor traffic trends be modeled if these metrics are misleading and biased? Our approach employs very flexible but simple probability mixture models to allow for consumer heterogeneity as well as individual behavioral evolution. That is, we can model each individual shopper as having an inherent rate of visiting a site. This rate, from which we can derive inter-visit times, varies across individuals to accommodate differences within the population. However, as consumers visit a site repeatedly, their behavior may systematically change as a result, i.e., there may be *behavioral evolution* in visiting. Therefore, to accurately monitor changes in visiting behavior, we model each consumer's rate of visit for each visit as the rate governing the timing of the previous visit multiplied by a probabilistic updating multiplier ( $RATE^t = RATE^{(t-1)} \times UPDATE^t$ ). The distribution of this updating multiplier provides a description of the trend in visit frequencies. For the same e-commerce site used in the examples above, an individual-level probability model reveals that visits are actually declining in frequency—contrary to the implications offered by the aggregate summary metrics offered in the above table.

Figure 2 offers a plot comparing the forecasts derived from a static model of visiting (i.e., not allowing for behavioral changes) versus one that does allow for behavioral evolution (including attrition). When behavior evolution is not accounted for, the model significantly over-estimates future visitor traffic as it does not allow individuals to change their behavior. Instead, it incorrectly

assumes that historical behavior is an appropriate indicator of future behavior. The evolving visit model, which does accommodate behavioral evolution, is capable of providing extremely accurate forecasts of visitor traffic.

In addition to being able to discern trends in visiting behavior, characterizing individuals' visiting behavior in terms of just frequency and trends offers strong *indications of purchasing*. In general, those shoppers who make more frequent store visits are more likely to purchase in any given visit—this is true both online and offline. However, evolution in visiting frequency also provides a strong indicator of a shopper's purchasing propensity. Those who are visiting a store at an increasingly frequent rate also have higher conversion rates than those who are showing a slow down in their visit frequencies. Furthermore, those shoppers who are considered frequent visitors *and* are increasing their rates of visit reflect an interaction effect between frequency and evolution and are particularly attractive, in terms of conversion rates, to the e-commerce site (see Table 2 for an example).

The differentiation between a shopper's visiting behavior and purchase conversion highlights the distinct roles of various marketing tools. Mass advertisements, targeted promotions, and brand-building efforts affect different stages in the consumer decision process. As such, they should be used differentially depending on the state of the site and the objectives of the e-commerce manager. For example, mass advertisements may generate awareness and drive traffic to the site. However, it does not necessarily have any impact on purchasing conversion. If the site's objective is indeed only to build traffic, the evolving visits model described above provides an effective method to evaluate the results of such a campaign. However, beyond just bringing in visitors, e-commerce sites must also consider the shopper's conversion decision once in the store. Likewise, tools designed to increase purchasing conversion and improve customer retention are quite distinct from tools designed to generate visitor traffic and may not have any effect on building traffic. The attention of marketing efforts at this stage of the game must shift from building awareness to building preference, purchase intentions, and—later on—customer retention. Hence, the marketing tactics available include building the brand, promoting to induce trial, and developing customer loyalty programs.

**TABLE 2.** Visiting Patterns as Indicators of Purchasing Conversion Rates (CR)

	Decreasing Frequency	Increasing Frequency
<b>Bookstore</b>		
Infrequent Visitors	CR = 10.9%	CR = 11.3%
Frequent Visitors	CR = 14.6%	CR = 18.6%
<b>CD Store</b>		
Infrequent Visitors	CR = 3.8%	CR = 5.7%
Frequent Visitors	CR = 4.0%	CR = 7.6%

## Conversion Behavior<sup>2</sup>

Before deciding which purchase conversion campaign to pursue, we must first fully understand the shoppers' conversion behavior at the site. The Internet has made store visits very "cheap" for shoppers. As it is virtually costless for consumers to visit online stores, consumers are not as committed to buy when in an online store as when they travel to visit a bricks-and-mortar store. This makes store traffic management a more important issue online than offline. Though high store traffic provides more buying opportunities, there is a tradeoff—heavy traffic is also expensive for e-commerce sites to accommodate. Therefore, it is important in any cost reduction efforts for an e-commerce site to differentiate between their shoppers and prioritize them for targeting purposes.

As noted, one way to differentiate visitors is by their visit frequency and stage of evolution. However, another very appealing and more direct method of differentiating between visitors is by their likelihood of buying, or conversion probability. High-probability buyers should be identified, and e-commerce managers need to ensure that their experience at the site is positive—for example, by channeling them to a faster server or offering service "perks." Low-probability buyers should also be identified as potential targets for promotions. Though these shoppers are not necessarily more promotion sensitive, it is more cost efficient to promote to unlikely buyers and persuade them to buy than it is to promote to shoppers who are likely to buy without any added incentive.

The challenge is to assess each visitor's likelihood of buying as they enter the site. Sales personnel perform this task at bricks-and-mortar stores, to varying success. In the virtual environment, one available measure is to examine each individual's historical conversion rate (the number of purchases made divided by the number of visits) and assume that this rate governs purchasing probabilities for all future store visits. However, there are several weaknesses to this approach. First, many visitors have very limited histories, i.e., they are relatively new visitors to the store. Without any information on that individual's visiting and purchasing history, there can be little faith in the ability of historical rates to predict future purchasing probabilities. For example, someone who has visited only once in the past and purchased during that visit is unlikely to exhibit a 100% conversion rate in the future. It is more likely that the purchase made was a chance result of a much lower conversion probability. Second, an individual's purchasing probabilities are not necessarily unchanging from visit to visit. A shopper may have buying cycles, e.g., they may visit an online store fairly regularly but may only purchase every other visit. In this case, their probability of buying changes fairly systematically from one visit to the next. In addition to cyclical changes in buying probabilities, there may also be fundamental changes that result from the shopper evolving with every visit and/or purchase. For example, as shoppers gain experience from repeated purchases, does this fundamentally change their latent propensity to buy in future visits?



Again, to solve this problem, we turn to individual-level probability models. In this case, we can model each customer as having an inherent conversion rate (or probability of buying at any given visit). Again, each individual's conversion probability is different and must be modeled as a heterogeneous trait. Furthermore, an individual's conversion probability is affected by his/her history of visits and purchases. The result is a model of conversion behavior that allows a shopper's conversion probability to change from visit to visit, be it as a result of evolution or a systematic purchasing cycle.

There are two sources of such *non-stationarity* (or changes in conversion probability): *cyclical changes* and *behavioral evolution*. First, cyclical changes may result from shoppers making a number of visits before each purchase, perhaps for the purpose of accumulating information prior to the purchase. In these cases, there is a systematic purchasing cycle in terms of visits, i.e., the shopper buys something once every  $x$  visits. Second, purchasing may also be cyclical in terms of calendar time. For example, some shoppers may visit the store irrespective of whether they intend to buy. Instead, purchases are made regularly every  $x$  days. In these illustrations, conversion probabilities may systematically change as the shopper makes repeat visits or as time passes.

In addition to the cyclical changes in purchase conversion, there may also be evolution in buying tendencies brought on by repeat visits and/or purchases causing more permanent changes. For example, as a shopper visits a site repeatedly, he or she may become accustomed to the site, and the effectiveness of each store visit in inducing a purchase may increase or decrease. Additionally, purchases may also fundamentally change a shopper's conversion behavior. That is, as individuals make repeated purchases, they may change their attitudes toward purchasing from the site. Many shoppers are reluctant to buy from an unfamiliar store site the first time, but as they gain purchasing experience, they become more comfortable with the site and may be more likely to buy. Curiously, research has shown that repeated purchases tend to decrease a customer's tendencies to buy from that site in the future, suggesting that there is an initial novelty effect that slowly wears away.

Therefore, when examining conversion rates, the e-commerce manager must realize that purchasing conversion is more complicated than just a single summary measure of purchases to visits. Instead, individuals have very different purchasing tendencies from one another. The ability to differentiate individual visitors by their purchasing probabilities provide e-commerce managers with more valuable information regarding each shopper's value as a customer as well as how to best serve them. Additionally, one must recognize that purchasing probabilities not only vary across individuals, but also change within individuals from visit to visit. Understanding how these probabilities change from visit to visit, as a result of visiting and/or purchasing experiences, provides information useful in forming marketing strategy. For example, are shoppers slow to convert? If this is the site's source of weakness, marketing resources should focus



**TABLE 3.** Taxonomy of Shopping Strategies

Timing	Motivation	
	Planned Purchase	No Planned Purchase
Immediate Purchase	Directed-Purchase Visit	Hedonic Browsing Visit
Future Purchase	Search/Deliberation Visit	Knowledge-Building Visit

on inducing trial and converting the non-buyers (or not-yet-buyers). On the other hand, if customer attrition is the source of problems, the site would be better served not by inducing trial or even building traffic, but by investing in customer retention and loyalty.

### Detecting Shopping Strategies<sup>3</sup>

A unique aspect of Internet clickstream data is its ability to provide information on a variety of different consumer shopping strategies. For example, some consumers are going to the store with a very specific purchase in mind. In those cases, their store visit is very directed and will almost certainly end in a purchase. In other cases, the shopper is interested purely in browsing, and the visit may or may not end in a purchase, depending on what the individual encounters during the visit. The question is, are there store visit characteristics, that can be found in navigational clickstream data, indicating the predominant shopping strategy?

Table 3 offers a taxonomy of different shopping strategies that vary along two dimensions: the *motivation* behind the shopping trip and the *timing* of any resulting purchase. Motivation refers to the underlying benefit (or utility) each store visit provides. Some store visits are made with a specific purchase in mind, and the decision to visit the store is simply a means to that end (*planned purchase*). Other visits, however, are not made for the purpose of a purchasing decision. Instead, the shopper derives pleasure from the simple act of visiting and exploring the store (*no planned purchase*). Both motivations can result in a purchase, either immediately or some time in the future.

The framework offered in Table 3 characterizes store visits by these different motivations and purchase timing into four shopping strategies: *directed-purchase* visits, *search/deliberation* visits, *hedonic browsing* visits, and *knowledge-building* visits. Directed purchase visits refer to those situations where the shopper enters the store with a clearly defined purpose of walking out with a specific purchase. Examples of this type of behavior are prevalent in grocery store visits. Search/ deliberation visits describe those visits where a future purchase is being considered, and the store visit is designed to gather relevant information needed to make a purchasing decision (e.g., pre-purchase visits to a car dealership). Hedonic browsing visits are store visits where the shopper enjoys

the process and is not necessarily looking for anything specific. Instead, these visits are inherently valuable to the shopper as browsing experiences. However, an immediate purchase may still occur as an “impulse” depending on what the individual encounters during the visit (e.g., clothes shopping). Knowledge-building visits are also visits where the shopper enjoys the process itself. The difference lies in the timing of any potential purchase. Shoppers engage in knowledge-building visits to gather useful information regarding the market environment that may be used for a future purchasing decision (e.g., attending trade shows).

So, what differentiates each of these shopping strategies in clickstream data? Directed purchased visits tend to be very focused sessions. In other words, the shopper has a plan of attack and will not likely deviate from it. The resulting store visit is one that can be described as having limited variety in the number of different categories and products being viewed. Furthermore, as directed purchasers are close to making a purchasing decision, they are likely to deliberate on one specific product. This is often manifested by repeat product page viewings. Therefore, when e-commerce sites observe sessions where the shopper views few categories and repeatedly views a product page, one can speculate that the shopper is purchase-directed. In these cases, it is not cost effective to promote to these shoppers as they are fairly goal-directed in their behavior and are likely to buy without any added incentive of promotions.

Search/deliberation visitors are also fairly goal-driven. The difference is how close to buying the shopper is. Search/deliberation visitors are in the market for a specific type of product but are still seeking information to help make their decision. Therefore, these sessions tend to focus on a limited number of product categories but can involve viewing multiple products in the category. In other words, these sessions have a high product-to-category ratio. These sessions will also involve deeper drill-down of product information and longer viewing time per page to process the information. When visits are identified as search/deliberation visits, according to their product-to-category ratio and level of product drill-down, a *utilitarian* market message can better affect the shopper's purchasing decision. That is, since these shoppers are actively considering and evaluating product information, they may be more responsive to promotions that reduce price or highlight positive product attributes.

Hedonic browsing visits are very different from both purchase-directed and search/deliberation visits. Hedonic browsing sessions will tend to be much shallower in nature. That is, the time spent per page is short and product drill-down is minimal. Furthermore, as hedonic browsers are not shopping with any particular product in mind, they might view a number of different categories leading to a low product-to-category ratio. Most purchases resulting from these shopping sessions are impulse purchases stimulated by attention-grabbing attributes of the environment. Therefore, to best promote to these shoppers and increase their purchase conversion, e-commerce marketers should target the

hedonic browsers with flashy and attractive promotions rather than with messages focused on objective product attributes.

Finally, knowledge-building visits are driven by the shopper's desire to learn about the product category and environment. As such, the session can be characterized by long page-viewing duration and a large percentage of the session spent on information related pages, such as advice columns, community areas, and company information pages. These shoppers are primarily interested in learning and not buying, at least not yet. Therefore, the e-commerce site is wise not to promote to them (at the risk of "spamming" their shoppers), as these shoppers are unlikely to respond.

The large numbers of customers entering an e-commerce site vary widely in their behavioral patterns and motivations. The ability to characterize a shopper's in-store behavior and infer the purpose of the store visit allows e-commerce marketers to better segment their customers and target them with the appropriate messages. In other words, a customer's experience at a site can be customized by analyzing the individual's observed navigational clickstream data. The e-commerce marketer can offer targeted promotions with messages customized for a specific type of shopper or provide enhanced customer service when appropriate.

## Discussion

If e-commerce managers wish to effectively allocate their marketing resources across activities, they must carefully consider the costs and benefits of pursuing each campaign. Whether the marketer is considering an advertising schedule to increase visitor traffic, targeted promotions to improve conversion rates, or loyalty programs to build customer retention, one must carefully separate purchasing into its component parts (as seen in Figure 1) in order to assess the potential effects of these marketing activities. For example, if an advertising campaign attracts a number of new visitors, how likely are they to purchase? For how long will they remain active customers? Models that carefully address the many aspects underlying customer purchasing can answer these questions.

Furthermore, the Internet is an ideal medium for e-commerce marketers to experiment with minor adjustments in their store environments and promotions in an effort to identify the most effective marketing mix. What is the ideal layout of the products? Which advertisements and promotions are the most effective and on which group of customers? Which items are more effectively cross-sold? The ease with which an e-commerce retailer can experiment with various attributes of their selling environment, coupled with the richness of data that the Internet is able to offer, provides marketers with an opportunity. With the right metrics and models, marketers can learn more about their customers and what they respond to better than ever before.

## Notes

1. Wendy W. Moe and Peter S. Fader, "Capturing Evolving Visit Behavior in Click-stream Data," Wharton Marketing Department Working Paper, 2000.
2. Wendy W. Moe and Peter S. Fader, "Which Visits Lead to Purchases?: Dynamic Conversion Behavior at E-Commerce Sites," Wharton Marketing Department Working Paper, 2000.
3. Wendy W. Moe, "Buying, Search, or Browsing: Using In-Store Navigational Click-stream to Identify a Typology of Store Visits," Working Paper, 2001.