UNDERSTANDING CONSUMER BROWSING PATTERNS: A SEQUENCE ANALYSIS APPROACH

ABSTRACT

Despite the huge number of website visits, the visit-to-purchase conversion rates have rarely exceeded 2.5% across online retailers. To address the low conversion rate problem, we argue that a critical first step is to categorize different types of website visits and understand the motivation of each visit based on consumers' browsing behavior. By applying the shopping goals theory, we extend the existing typology of online browsing pattern based on two dimensions: type of shopping goal (i.e., concrete product goal, general product goal, and no goal) and consumers' browsing path to achieve the goal (i.e., planned shallow visit, planned deep visit, and adjusted visit). Moreover, since a shopping process is usually composed of multiple visits, we also investigate how consumers' browsing patterns evolve across different store visits and how the evolvement is related to consumers' purchase decision making.

The empirical analyses are conducted on a large-scale clickstream dataset obtained from a big e-commerce website in China. At the store visit level, we conduct sequence analysis and cluster analysis, and obtain eight browsing patterns: Quick Keyword Search, Consecutive Keyword Search, Quick Category Search, Consecutive Category Search, External Triggered Visit, Recommendation Adjusted Visit, Quick Promotion Check, and Consecutive Promotion Exploration. At the individual level, our results reveal the effectiveness of marketing stimuli (e.g., personalized recommendation and discount promotion) in increasing purchases greatly depends on customers previous browsing patterns. Specifically, personalized recommendation is ineffective in encouraging purchases for consumers who have already defined a concrete product goal; surprisingly, it even reduces the purchase probability for consumers who have deeply explored the interested product category directed by a general product goal. On the contrary, discount promotion is generally effective in encouraging purchases.

Key words: browsing patterns, clickstream data, sequence analysis, evolvement path, personalized recommendation, discount promotion

INTRODUCTION

Despite the prevalence of e-commerce, the visit-to-purchase conversion rates have rarely exceeded 2.5% across online retailers (Smart Insights 2017). To address this low conversion rate issue, a viable approach adopted by many companies is to provide customized services that match consumers' preferences to encourage purchases (Simonson 2005). However, existing customization algorithms are mainly based on a consumer's browsing history of product page content (Ricci et al. 2015), potentially leading to inadequate understanding of his/her motivations of each store visit. That is, even for two consumers with same browsing history of product pages, their motivations to visit the store can still be greatly different (e.g., one searching for a specific product vs. one exploring a general category), resulting in their distinct desired customized services (Lee and Ariely 2006). Therefore, the inadequate understanding of consumers' visitation motivations may produce discrepancies between consumers' desired customization demand and provided services, which would, instead, impede firms' profit making (Fogliatto and Da Silveira 2010).

In traditional brick-and-mortar stores, experienced sales people usually first identify consumers' visitation motivations based on their in-store browsing behavior, and then provide customized services to persuade them to buy (Moe 2003). For instance, for two customers taking a look at a MacBook in a department store, the one who directly heads for the MacBook probably has a concrete shopping goal before entering the store, whereas the one who wanders around the laptop area and then stops at the MacBook tends to only have a general purchase idea. Based on these two distinct browsing paths, an experienced sales assist would introduce more advantageous features of MacBook for the first customer to increase his/her purchase likelihood, while asking the second customer about his/her preferences to recommend other models. Thus, in addition to browsing history, consumers' navigational paths provide additional insights into their visitation motivations, based on which sales people can make timely adjustments in customized services.

Learning from offline sales people, we "observe" online shoppers' browsing behavior based on their sequential page-to-page views, and then categorize browsing patterns to recognize their visitation motivations and provide effective customization. Extant typology categorizes online shoppers' browsing patterns into goal-directed search and exploratory search, depending on whether they have a shopping goal in mind (Moe 2003). However, there are several limitations of this typology. For instance, according to the shopping goals theory, a consumer's shopping goal can be further differentiated in terms of concreteness levels (e.g., from general goals to concrete goals), which significantly determine his/her sensitivity towards different marketing stimuli (Lee and Ariely 2006). Moreover, even for consumers with a similar goal, their browsing path towards goal attainment may still vary depending on the information amount they need. Hence, this study proposes a more comprehensive typology to characterize different browsing patterns, based on which consumers' visitation motivations can be better identified.

While the existing typology focuses on a consumer's browsing behavior within a given store visit, it is worth noting that his/her browsing patterns tend to evolve dynamically towards purchase decision making across multiple store visits. As indicated in a recent survey, more than 60% of sales in e-commerce websites are driven by individuals who make multiple store visits (Forrester Research 2015). With accumulative product information, a consumer is likely to have various visitation motivations and his/her interest in marketing stimuli also changes along the shopping journey (Lee and Ariely 2006). Recognizing this, an increasing number of companies begin to implement customized marketing stimuli based on consumers' historical shopping behavior, aiming to increase their purchase probability (e.g., Hauser et al. 2009). Nevertheless, little empirical analysis has compared the effectiveness of these marketing stimuli for consumers with different evolvement paths. Therefore, beyond a single store visit, we also investigate how the evolvement of an individual's browsing patterns, especially the evolvement into patterns that are influenced by different marketing stimuli (e.g., recommendation vs. promotion), is related to consumers' purchase decision making.

To achieve these two research objectives, we conduct an empirical study by leveraging a large-scale clickstream dataset. Based on consumers' page-to-page navigation behavior at each store visit, we employ sequence analysis and cluster analysis, borrowed from bioinformatics, to categorize consumers' browsing patterns. By leveraging the shopping goals theory, we lay a theoretical foundation for the browsing patterns emerged from the analysis. Beyond a single visit, we treat a series of an individual's visits as a whole and investigate how a customer's browsing patterns evolve into those influenced by marketing stimuli. By employing econometric models, we quantify the effect of such evolvements on the customer's purchase decisions and provide insights into the effectiveness of different marketing stimuli (e.g., recommendation vs. promotion) when the customer's evolvement paths originate from different browsing patterns.

THEORETICAL BACKGROUND

Online Browsing Behavior and Purchases

Extant literature has demonstrated that consumers' browsing behavior influences their likelihood of purchasing (Bucklin et al. 2002; Yadav and Pavlou 2014). For example, some studies used aggregate clickstream data to predict each consumer's purchase decision (Moe et al. 2002; Moe and Fader 2004; Park and Fader 2004). However, due to the common weakness of aggregate data, most sequential information is lacking in the data analyses of these studies. In order to take full advantage of the richness of clickstream data, Montgomery et al. (2004) used page-level clickstream data, together with a dynamic multinomial probit model, to predict purchase conversion by analyzing the path consumers choose to navigate through a website. Likewise, Sismeiro and Bucklin (2004) also used clickstream browsing data to predict the conversion rates by linking the completion of each sequential nominal user task to what consumers do and to what they are exposed to at the website. Moreover, Padmanabhan et al. (2006) utilized user-centric data, which recorded all sites a user visited in a session, to predict her purchase probability. By employing a data mining approach, the authors demonstrated the magnitudes of gains achieved from user-centric data in the prediction tasks. Although these studies have significantly improved the accuracy of purchase prediction, due to the lack of proper summary of consumers' browsing patterns, they provide limited insights into customers' visitation motivations.

While the above studies mainly focus on a single store visit, more attention has been drawn to individuals' browsing behavior across different store visits recently (e.g., Park and Park 2016; Zhang et al. 2014). Specifically, researchers have observed customers' online store visit patterns tend to be clustered (Park and Park 2016; Zhang et al. 2014). A clustered store visit pattern exhibited concentrations of visit events in close temporal proximity, with clusters separated by empty (or less dense) visit events. Notably, within each visit cluster, purchase rates were higher at later visits, compared to earlier visits (Park and Park 2016). In this sense, Zhang et al. (2014) added visit clumpiness (i.e., a measure of the degree to which an individual's store visits were clustered) to the traditional RFM (recency/frequency/monetary value) framework, and significantly improved the predictive power of the framework for purchases. These findings hinted that an individual's store visits could be described as an evolving series of interrelated choices, where subsequent visit events depended on the outcome of earlier visits, and purchases occurred after gathering adequate information in earlier visits. However, these studies merely characterizes a customer's store visit history (i.e., a series of visit events to an e-commerce website) as a temporal point process, which neglects his/her browsing behavior that take place within each "point" (Park and Park 2016). Without such information, it is difficult to characterize the evolvement path of the customer's browsing behavior over time, and to quantify the effect of such evolvements on his/her purchase decisions.

Typology of Online Browsing Patterns

Existing Typology of Browsing Patterns

In general, there is a wide spectrum of browsing behavior observed at e-commerce websites, which reflects shoppers' different motivations to visit the store. Janiszewski (1998) dichotomized browsing behavior into two general types: goal-directed search and exploratory search. Goal-directed search referred to the browsing pattern when the consumer had a shopping goal in mind. The search pattern was thus relatively focused towards collecting relevant information to attain the goal. Exploratory search pattern, on the contrary, tended to be undirected and less focused when consumers did not have such a goal.

Moe (2003) extended this dichotomy by adding a new dimension of purchasing horizon and developed a typology of online consumers' browsing behavior. Based on the analysis of page content consumers browsed, the author theoretically categorized browsing behavior into four patterns: directed buying, search and deliberation, hedonic browsing, and knowledge building. Direct buying visits were made by consumers who had concrete purchase goals and were not lacking any substantial information before making decisions. The in-store behavior was thus very focused towards a specific and immediate purchase. Search and deliberation visits were also goal-directed; but different from direct buying pattern, consumers in this pattern were motivated by a future purchase and collected relevant information to construct consideration sets. Hedonic browsing pattern was motivated less by utilitarian purposes of making a purchase, but more by hedonic utility derived from shopping experience. Browsing process tended to be stimulus-driven and impulse shopping might occur when certain stimuli were encountered. Knowledge building visits were made by users who had no purchase intention and came to the site for the purpose of increasing product or market expertise. Through the empirical analysis, besides these four theoretically categorized browsing patterns, the majority of store visits (75.83% of total visits) emerging from the data were shallow visits. In these shallow visits, visitors came to the site and viewed only one to two pages with little time spending.

Although this typology provided more insights into shoppers' visitation motivations, there were still several limitations. First, this typology does not consider that consumers probably have different shopping goals, which determine their interest in different types of marketing stimuli. Second, the typology is only based on consumers' initial motivation and search strategies. It neglects the potential adjustment in their search strategies along the way, which is prevalent especially in the online shopping environment with more contextual cues. Lastly, many details in browsing paths, especially the sequential information, are lacking in this typology since the author merely uses aggregate clickstream data.

New Typology of Browsing Patterns: Theoretical Framework

Few behaviors of human beings are as purposeful as shopping. Research in cognitive psychology defines a goal as a hybrid of mental representation of the goal and path to achieve it (Pervin 1982). In the shopping context, customers' shopping goal is considered as both motivators and organizing forces in information searching. As a motivator, it provides a general guide in determining the relevant and useful properties for creating meaning in each store visit (Murphy and Medin 1985). With such guiding force derived from the goal, the consumer organizes the information search path to achieve the goal by focusing attention on relevant information cues and excluding other cues as irrelevant (Puccinelli et al. 2009). Nevertheless, with cumulated experience with a goal, consumers' goal-derived guide can span a continuum from loose, poorly organized property categories to tightly integrated, well-defined property categories (Barsalou 1991). In this study, we extend the existing typology based on customers' shopping goals and their browsing paths towards goal achievement.

Concreteness of Shopping Goals

According to Lee and Ariely (2006)'s shopping goals theory, consumers' shopping goals are always unlikely to be highly specified. Rather, their goal concreteness can range from relatively abstract (e.g., general categories) to very specific and precise (e.g., specific products). Specifically, in some cases, when consumers are uncertain about exact products they want, they would think about their goals in superordinate and more abstract terms. To define a desired product for purchasing, consumers tend to be in a deliberative mind-set to gather information of preferred product features (Gollwitzer 1999). In other cases, when consumers are looking for a specific product with well-defined desirable features, they conceive of their shopping goals in subordinate and more concrete terms. Under such circumstances, consumers are usually in an implemental mind-set where they facilitate the attainment of the specific goals by well-defined solutions (Gollwitzer 1999). Importantly, goal concreteness determines consumers' sensitivity towards external cues in online stores (Gollwitzer et al. 1990). Specifically, individuals with general goals are usually more receptive and open-minded to external information cues (e.g., promotion and recommendation), compared to those with concrete goals.

Path Towards Shopping Goal Attainment

Extant literature suggests that even for consumers with the same type of shopping goals, the length of their browsing paths may still vary depending on the quantity of information they need (Moe 2003; Moe 2006). In most cases, online consumers visit the e-commerce site mainly for particular information related to their goals, and their browsing paths are relatively short (Moe 2003). These shallow visits are usually made due to the following reasons: 1) consumers are non-serious buyers who search the information just for reference or for information updates; 2) consumers collect basic information for the preparation of subsequent browsing sessions; 3) consumers have obtained substantial information and only need to make a final confirmation before the purchase (Moe 2003).

In contrast, other customers may travel longer paths to collect a complete set of information to achieve their goals. Typically, when consumers are considering purchases but lack sufficient information to make a choice, they usually first evaluate a large product set and reduce it to a smaller, manageable choice set. After that, they compare alternatives in the choice set to make purchase decisions (Moe 2006). The process is quite costly due to the large number of products being evaluated, which results in long browsing paths.

Notably, consumers' browsing paths are not always fixed. According to the shopping goals theory, people do not always stay with a fixed search path based on their initial goals; instead, they will probably dynamically modify their strategies to maximize the rate of gaining valuable information (Lee and Ariely 2006). That is, as long as consumers enter the e-commerce website, they will go through a constructive search path during which their preferences and strategies are constructed and adjusted depending on the information environment (David et al. 2007; Payne et al. 1992). Therefore, when we categorize different browsing patterns, it is important for us to take into account all the cues in the information environment, such as the abundant customized recommendations and regularly updating promotions at e-commerce websites. These information cues may match consumers' information needs, so that they are willing to follow these cues to get to the promising information (Pirolli 1997; Pirolli and Card 1999). Hence, it is possible that consumers may adjust their initial browsing paths as they encounter various information cues at the e-commerce website.

Evolvement of Browsing Patterns: Hypothesis Development

Based on the theoretical division of online consumers' browsing patterns in a single visit, we further investigate how consumers' browsing patterns dynamically evolving across store visits will affect their final purchase decisions. In particular, we focus on the evolvement into browsing patterns that are adjusted by marketing stimuli, such as personalized recommendation and discount promotion that are prevalent in online stores. Although a substantial number of studies have investigated the effectiveness of personalized recommendation and discount promotion in increasing sales in many settings (e.g., Chandon et al. 2000; Kumar and Benbasat 2006; Winterich and Barone 2011; Xu et al. 2014), very few of them consider customers' browsing patterns in previous store visits. The evolvement of browsing patterns provides hints into consumers' dynamically changing goals and cumulated information they have gathered, which may affect the effectiveness of marketing stimuli.

Evolving into Recommendation Adjusted Pattern

Personalized recommendations, as a type of prevalently adopted marketing stimuli, provide online consumers with recommendations on what products for purchasing based on their individual needs (Komiak and Benbasat 2006). A central function of the personalized recommendation system is to elicit a customer's preferences of product attributes, and then to provide product recommendations which satisfy these personal preferences (Xiao and Benbasat 2007; Xu et al. 2014). To achieve this function, a recommendation system usually first collects each customer's information to elicit his/her preferences of product attributes either explicitly (e.g., based on the customer's rating on products) or implicitly (e.g., based on the customer's purchase history and navigational pattern). Based on these preferences, the system then assigns weights on product attributes in the underlying algorithm and generates personalized recommendations for customers. Notably, most of extant research on recommendation systems focuses on developing and evaluating the underlying algorithms of generating recommendations (e.g., Herlocker et al. 2004; Sarwar et al. 2000).

However, for consumers who have different goals and accumulate different levels of product knowledge in previous store visits, even one same algorithm may generate recommendations that are properly designed for one type of consumers but are poorly designed for another type of consumers. Considering the nature of personalized recommendations, this type of marketing stimuli seems to be particularly helpful for consumers who have general product goals in the decision-making process. Specifically, for these customers with incomplete and ill-defined preferences of product attributes, once personalized recommendations are congruent with their desired products, consumers would like to rely on these stimuli to extend the preference set and be more determined to make a purchase decision (Lee and Ariely 2006). However, this effect may be contingent on the quantity of product information a customer has gathered in previous visits.

For consumers who have a general product goal and have traveled a shallow visit path previously, they have ill-defined preferences and probably have insufficient knowledge about the product category due to the limited search effort made in previous visits. When their browsing pattern evolves into recommendation adjusted pattern, given the relatively small set of desired product attributes, there is a high probability of recommendations meeting consumers' requirements. Moreover, personalized recommendations enable consumers to quickly obtain information of a certain product category and help them manage the overwhelming product information in the online shopping environment. They largely lower consumers' information search cost by directly guiding them to products that have the potential to fit their needs (Häubl and Trifts 2000; Kumar and Benbasat 2006). In addition to these utilitarian values, recommendations also increase the social presence of the website (Kumar and Benbasat 2006). Specifically, personalized recommendation creates an impression of engaging customers in a one-to-one dialogue, which gives consumers a sense of social connection with the website. Such a social connection creates affection which arouses purchasing, especially for those who may not seriously intend to buy before entering the website (Huang 2016). Hence, we posit:

H1a: Compared to consumers who have general goals over time, those whose browsing pattern evolves from general goal-directed shallow visit pattern into recommendation adjusted pattern are more likely to purchase.

For consumers who have a general goal but have traveled a deep visit path previously, although their preferences are not concretely defined, they probably have gained a certain level of product knowledge by comparing many alternatives but do not choose to purchase in previous visits. Since these consumers do not have a concrete product goal to guide their search routine, their browsing path tends to be less focused towards attaining certain product attributes (Janiszewski 1998). Thus, options customers have browsed are probably non-alignable (i.e., options varying along unique dimensions, instead of comparable dimensions) (Griffin and Broniarczyk 2010). Feature learning in browsing these options may increase customers' desires as they integrate all attractive product attributes into parts of an ideal product, which serves as a basis of comparison but is difficult to attain (Griffin and Broniarczyk 2010). As a result, choosing from these options requires trade-offs among customers' desired attributes, which probably cause disappointment and lead to the non-purchase decision in the previous store visits (Diehl and Poynor 2010; Griffin and Broniarczyk 2010).

When customers' browsing pattern evolves into recommendation adjusted pattern, personalized recommendations seem to make such "trade-offs" automatically for consumers. Nevertheless, given consumers have browsed a relatively comprehensive set of products in the focal category, these recommendations, generated based on their browsing history, have probably been considered by consumers in previous visits. Therefore, such "trade-offs" made by the system provide an official confirmation that there are no potential choices matching consumers' all desired product attributes. In this sense, there is a minimal chance for these consumers to make a purchase. The purchase probability is even lower than that for customers who continue to explore product alternatives guided by a general product goal, because they may occasionally succumb to an impulse purchase in the product exploration. Hence, we have:

H1b: Compared to consumers who have general goals over time, those whose browsing pattern evolves from general goal-directed deep visit pattern into recommendation adjusted pattern are less likely to purchase.

However, for consumers who have concrete product goals, personalized recommendations may have very subtle influence in their decision-making process. First, given the central function of personalized recommendation is to elicit consumers' preferences of product attributes, consumers who have concretely defined their preference are less sensitive to this type of marketing stimuli because they do not match customers' needs (Lee and Ariely 2006). Second, for consumers whose preferences are well defined, when their browsing pattern evolves into recommendation adjusted pattern, it indicates these recommendations fit consumers' preferences of product attributes. They simplify the search process guided by the concrete product goal, and lead to similar search results. Hence, we do not hypothesize the effect of evolvement into recommendation adjusted pattern on purchases for consumers with concrete product goals.

Evolving into Promotion Adjusted Pattern

Different from personalized recommendation, as a traditional marketing means, discount promotion provides consumers with a diverse set of benefits (Chandon et al. 2000; Stilley et al. 2010; Winterich and Barone 2011). Discount promotion mainly provides utilitarian benefits, such as the monetary savings, higher product quality (i.e., enabling customers to upgrade to higher-quality products at affordable prices), and improved shopping convenience (i.e., reducing customers' search and decision costs by advertising available promotional items) (Chandon et al. 2000). In addition to utilitarian benefits, discount promotion also generates hedonic values in enhancing consumers' self-perception of being smart, fulfilling their needs of information exploration (Chandon et al. 2000), and allowing them to hunt unexpected “treasures” and realizing fantasies at affordable prices (Bardhi and Arnould 2005).

Given these benefits, there is ample evidence that discount promotion is effective in encouraging purchases (Chandon et al. 2000; Hardesty and Bearden 2003; Stilley et al. 2010). Therefore, compared to consumers whose search paths are directed by their goals (i.e., both general and concrete product goals) as planned, those whose browsing pattern evolving into promotion adjusted pattern can gain additional benefits provided by discount promotion, and therefore are more likely to purchase. Hence, we posit:

H2: Compared to consumers who have general goals over time, those whose browsing pattern evolves from general goal-directed pattern into promotion adjusted pattern are more likely to purchase.

H3: Compared to consumers who have concrete goals over time, those whose browsing pattern evolves from concrete goal-directed pattern into promotion adjusted pattern are more likely to purchase.

DATA

In the study, we leverage a large-scale clickstream data set to categorize consumer browsing patterns and test the proposed hypotheses. Nevertheless, as a type of path data, the structure of clickstream data is very complicated since each record is a multivariate sequence that represents a consumer’s continuous page viewings over time (Hui et al. 2009). To take full advantage of the richness of clickstream data, we leverage a bottom-up approach to analyze consumers’ navigation paths at different levels (i.e., session- and individual-level). Here, a session (used interchangeably with “store visit”) is defined as “a period of sustained Web browsing or a sequence of page viewings" (Montgomery et al. 2004, p.581). If a consumer has no page viewing in 20 minutes, the session is assumed to be ended, and the next page viewing starts a new session (Montgomery et al. 2004). An individual-level browsing path refers to a sequence of the individual's browsing sessions in a given time period.

We obtain the clickstream data set through collaborating with one of the biggest e-commerce websites in China. The website supplies a diverse set of products, including apparel, kitchenware, electric appliance, digital products, food, accessories, etc. Our data set covers the time period from June 1st to June 30th in 2012, including consumers' clickstream data and session-level order records. In each session, a consumer's browsing behavior was recorded as a sequence of URLs with timestamps and click actions. Therefore, we can obtain the full text and HTML content of each page through recapturing the page by URL (Montgomery et al. 2004). We can also know what links the consumer clicked within each page. At the end of the session, the consumer's purchase decision (i.e., purchase vs. not purchase) was recorded. In total, there are189,909 sessions made by 112,781 unique consumers during June 2012.

METHOD

Typology Development

Information Seeking Sub-Sequence

As clickstream data encapsulate many details about each individual's page viewing history, it is difficult for practitioners and researchers to consider all the page level information in such large and cumbersome data sets (Moe and Fader 2004). In this case, parsimony and efficiency become two important criteria to choose the appropriate data analysis method (Moe and Fader 2004). Nevertheless, we find that, in previous studies, the decisions about what to use in the data analysis were almost made based on what was available in the data set, rather than what would be most useful to describe the navigation paths. In fact, as Sismeiro and Bucklin (2004) suggested, we are able to identify and model the consumer's shopping progress based on several critical points in the shopping experience. Therefore, in order to make our algorithm more parsimonious and efficient, we simplify the whole browsing sequence by using product pages as well as their navigation sources (i.e., links from which consumers get to the product pages) to describe the information seeking process.

As discussed in the previous sections, online browsing behavior is composed of a series of information seeking processes (David et al. 2007). During each information seeking process, consumers use different strategies (e.g., search, category, etc.) to reach the product page, which contains detailed product information, item description, price information, availability, product reviews, and return policies (Montgomery et al. 2004). In order to distinguish and analyze different information seeking strategies, a consumer's page viewing sequence can be further divided into several sub-sequences, each of which represents an information seeking process and ends up with a product page viewing. A consumer will repeat the information seeking process and view various product pages until her product knowledge is accumulated to a threshold to make a purchase decision (Bloch et al. 1986).

Since navigation source indicates the strategy that a consumer uses to seek the product information, for the purpose of parsimony and efficiency, it is reasonable to use the navigation source instead of the whole information seeking sub-sequence to describe the navigation path to the product page. In a typical e-commerce website, consumers may reach product pages from either internal navigation sources (i.e., sources within e-commerce websites) or external ones

Table 1. Navigation Source Categories			
Navigation Sources	**Code**	**Definition**	**Examples**
Internal			
Search	S	The product links in the within-website search result list.	Product links in the result list by searching "iPhone 8 plus gold" within the e-commerce website
Category	C	The product links in the category page which contains a list of items belonging to the specific category.	Broad product categories: digital products, apparel, electric appliance, etc. Product categories for digital products: phones, cameras, televisions, etc. Product subcategories for phones by brand: iPhone, Samsung, HTC, etc.
Promotion	P	The links of specific products with discounts.	Promotion for special events: anniversary sales, season-end sales, etc. Group-buying initialized by the e-commerce website
Recommendation	I	The links of relevant products that consumers may be interested in based on their browsing histories and preferences.	Recommendation link of "iPhone 8 plus" in the product page of "Samsung Note 8"
External			
Advertisement	A	The advertisement links at external websites such as entertainment websites, surfing portals, email logon pages, etc.	Advertisements at entertainment websites: www.pps.tv Advertisements at surfing portals: www.hao123.com Advertisements at mail logon pages: mail.126.com and mail.163.com
UGC	U	The links provided by users at external websites where consumers discuss certain products.	Online forums: www.tianya.cn Information portals: www.zol.com.cn Product sharing sites: www.etao.com
Search Engine	E	The links in the search result list obtained from a search engine.	www.baidu.com www.google.com.hk

(i.e., sources outside e-commerce websites) (Hoffman and Novak 2000; Montgomery et al. 2004).

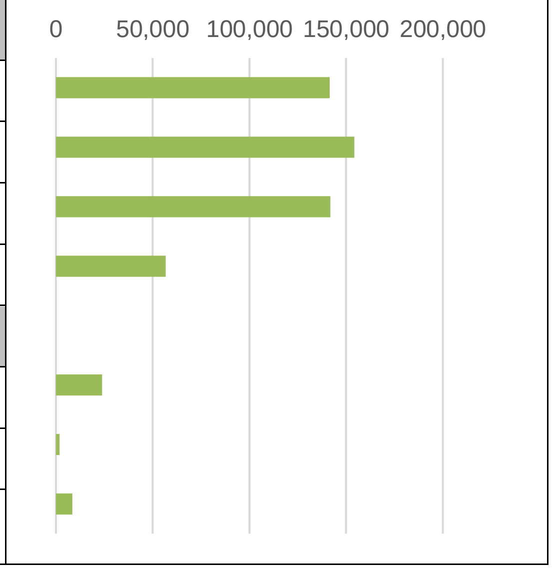
Adapted from Montgomery et al.'s (2004) categorization schemes of e-commerce web pages, we conduct a task analysis of what users do at e-commerce websites and find four categories of internal navigation sources. They are Search, Category, Promotion, and Recommendation. Search refers to the product links in the within-website search result list, which is derived from consumers' keyword search. Category refers to the product links in the category page which contains a list of items belonging to the specific category. These categories are pre-defined by online retailers and are usually general in characterize products. Promotion refers to the links of specific products with discounts. Online retailers usually update promotional items on a regular basis. Recommendation refers to the links of relevant products that consumers may be interested in based on their browsing histories and preferences.

As to external navigation sources, there are three categories: Advertisement, User-Generated Content (UGC) and Search Engine. Advertisement refers to the advertisement links at external websites, such as entertainment websites, surfing portals, and email logon pages. UGC refers to the links provided by users at external websites where consumers discuss certain products, such as online forums, information portals, and product sharing sites. Search Engine refers to the links in the search result list obtained from a search engine. Table 1 presents the definitions and examples of these seven categories of navigation sources.

Table 2 shows the distribution of different navigation sources in our data. Most of the product pages are navigated from the internal sources, which account for 93.51% of the total, while the rest are from the external sources, which account for only 6.49% of all the sub-sequences. Among the internal navigation sources, Category and Search are the top two most frequent navigation sources. As to external sources, advertisement is the main source type leading consumers to the product pages.

Table 2. Distribution of Navigation Sources

Navigation Source	Number	Percentage	Distribution
Internal			
Search	141,475	26.77%	
Category	154,175	29.17%	
Promotion	141,913	26.85%	
Recommendation	56,668	10.72%	
External			
Advertisement	23,863	4.51%	
UGC	1,888	0.36%	
Search Engine	8,576	1.62%	



Sequence Analysis

Due to the complexity of consumer browsing behavior, traditional statistical methods as well as other commonly used multivariate techniques, which ignore the temporal relationship between click actions, are not suitable to analyze browsing patterns (Bucklin et al. 2002; Hui et al. 2009). In order to consider the sequential information in the clickstream data, we use sequence analysis and clustering technique to identify consumers' browsing patterns. Adapted from bioinformatics, sequence analysis is tailored to measure the similarity/dissimilarity of sequences (Abbott 1995). Different from previous approaches in analyzing customer clickstream data, sequence analysis treats each data sequence as a whole unit, and retains sequential information in each unit (Abbott and Tsay 2000). Together with the clustering technique, sequence analysis is able to identify characteristic patterns quantitatively, and separates different patterns efficiently among a large sample of sequences (Abbott and Forrest 1986).

We use the navigation source of product page to represent the information seeking strategy in each sub-sequence. Then we regard each information seeking sub-sequence as a single unit in the session-level sequence (see figure 1). Table 3 presents the distribution of the number of sub-sequences per session. Using session-level sequences as input, we conduct two separate analyses: pairwise sequence alignment to establish the distance matrix for all pairs of sequences in the data set (Abbott and Tsay 2000; Needleman and Wunsch 1970), and hierarchical clustering to cluster homogenous sequences and identify consumers' browsing patterns (Brzinsky-Fay and Kohler 2010; Joseph et al. 2012).

<table>
<tr><th colspan="5">Table 3. Distribution of the Number of Sub-sequences per Session</th></tr>
<tr><th>Number of Sub-sequences per Session</th><th colspan="2">Number of Sessions</th><th colspan="2">Percentage in Total Sessions</th></tr>
<tr><td>1</td><td colspan="2">93,212</td><td colspan="2">49.08%</td></tr>
<tr><td>2</td><td colspan="2">37,389</td><td colspan="2">19.69%</td></tr>
<tr><td>3</td><td colspan="2">19,819</td><td colspan="2">10.44%</td></tr>
<tr><td>4</td><td colspan="2">12,040</td><td colspan="2">6.34%</td></tr>
<tr><td>5</td><td colspan="2">7,545</td><td colspan="2">3.97%</td></tr>
<tr><td>6</td><td colspan="2">5,014</td><td colspan="2">2.64%</td></tr>
<tr><td>7</td><td colspan="2">3,433</td><td colspan="2">1.81%</td></tr>
<tr><td>>=8</td><td colspan="2">11,457</td><td colspan="2">6.03%</td></tr>
<tr><th rowspan="2">Number of Sub-sequences per Session</th><th>Mean</th><th>Std Dev</th><th>Min</th><th>Max</th></tr>
<tr><td>2.78</td><td>4.78</td><td>1</td><td>1015</td></tr>
</table>

Pairwise sequence alignment is one of the most prevalent sequence analysis techniques used in bioinformatics (Mount 2004; Needleman and Wunsch 1970). It is often used to compare the similarity of protein or nucleotide sequences. To quantify similarity, two sequences are first aligned using a specific algorithm, and then their similarity score is computed based on the alignment. In the field of bioinformatics, two types of pairwise sequence alignment techniques have been developed, namely global alignments and local alignments (Polyanovsky et al. 2011). We choose the global alignment technique in our analysis, because we are more interested in the overall similarity of two browsing sequences. A general pairwise global alignment technique utilizes the Needleman–Wunsch algorithm, which is based on dynamic programming (Needleman and Wunsch 1970). The essential foundation in the algorithm is the scoring system. For example, when comparing two sequences, we may decide to give a reward of +2 to a match, a penalty of -1 to a mismatch, and a penalty of -1 to a gap. Thus, for the alignment:

S1: ABCCD

S2: ACD- -

We get a similarity score of 2-1-1-1-1=-2. Similarly, the score for the following alignment is 2-1+2-1-1=1.

S1: ABCCD

S2: A- CD-

As can be seen, there can be many alignment ways and corresponding similarity scores for the same pair of sequences. The purpose of the Needleman–Wunsch algorithm is to obtain the optimal global alignment with the largest similarity score based on the scoring system. We explain the algorithm with the above example in Table 4. We started with a zero initialized in $MAT_{0,0.}$ The first row and column are added during the initialization, whose purpose is to allow gaps in the beginning of sequences when aligning. We calculate the matrix from the top-left to the lower-right, achieving the optimal alignment pathway. Moving through the cells row by row, $MAT_{i,j}$ represents the maximum-match similarity score after aligning $S1_j$ and $S2_i$. It is the highest score calculated from existing scores to the left, top or top-left (i.e., $MAT_{i,j-1}$, $MAT_{i-1,j}$, and $MAT_{i-1,j-1}$, respectively). $MAT_{i,j}$ calculated from the left denotes a gap in S2 (i.e., a "- " in S2, $MAT_{i,j} = MAT_{i,j-1} – 1$); $MAT_{i,j}$ calculated from the top refers to a gap in S1 (i.e., a "- " in S1, $MAT_{i,j} = MAT_{i-1,j} -1$); and $MAT_{i,j}$ calculated from the top-left represents a match (i.e., a match between the letters in the current column ($S1_j$) and row ($S2_i$), $MAT_{i,j} = MAT_{i-1,j-1} +2$), or a mismatch (i.e., a mismatch between the letters in the current column ($S1_j$) and row ($S2_i$), $MAT_{i,j} = MAT_{i-1,j-1} -1$). $MAT_{i,j}$ is chosen as the highest value from the above three ways. Thus, every step is ensured to be optimal. Eventually, $MAT_{3,5}$ is the similarity score of the optimal alignment. Meanwhile, trace matrix (TMAX) records the path choices for each $MAT_{i,j}$. Once the two matrices are filled up, we trace the pathway back from $TMAT_{3,5}$ to $TMAT_{0,0}$. Then we follow the path from the start to the end to construct the maximum-match pathway. There may be multiple equally optimal alignments (e.g. two in our example).

Table 4. Example of Pairwise Sequences Alignment

Sequence Alignment Input:
S1: ABCCD
S2: ACD
Match=+2; mismatch=-1; gap=-1;

Procedure:
Needleman–Wunsch Matrix ($MAT_{i,j}$,i∈[0,3], j∈[0,5])

	S1	A	B	C	C	D
S2	0	-1	-2	-3	-4	-5
A	-1	2	1	0	-1	-2
C	-2	1	0	3	2	1
D	-3	0	-1	2	1	4

Trace Matrix ($TMAT_{i,j}$,i∈[0,3], j∈[0,5])

	S1	A	B	C	C	D
S2	done	left	left	left	left	left
A	top	top-left	left	left	left	left
C	top	top	left/top	top-left	top-left/left	left
D	top	top	left/top	top	left/top	top-left

Optimal Alignment (Similarity Score=4):

S1:	ABCCD;	or	ABCCD
S2:	A- - CD;		A- C- D

One critical setting in the algorithm applications is the scoring system (Aisenbrey and Fasang 2010; Brzinsky-Fay and Kohler 2010). Reward and penalty values are suggested to be set based on theories or application scenarios (Abbott and Tsay 2000; Brzinsky-Fay and Kohler 2010). In the current study, we set a reward of +2 to a match, a penalty of -1 to a mismatch, and a penalty of -0.51 to a gap. The rationale for doing this is that when comparing two browsing sequences, we focus on not only the matches of information seeking strategies, but also the differences of sequence lengths. A reward of +2 to a match with a penalty of -1 to a mismatch highlights the importance of matches. As to the gap penalty, too much penalty (absolute value) makes the length effect salient and weakens the importance of matches, while too small value makes the mismatch penalty fail (e.g., |gap penalty| < |half mismatch penalty|) (Abbott and Tsay 2000). Thus, a gap penalty of -0.51 is a balance of our two major concerns.

We do not follow the typical scoring system for building the distance matrix, which abandons the reward to a match and only assigns costs to mismatches and gaps (Abbott 1995; Abbott and Tsay 2000; Needleman and Wunsch 1970). In our settings, more matches in two long sequences are treated as more valuable, which is of significance when comparing consumers' browsing behaviors. We tested different value sets of match rewards, mismatch and gap penalties in an iterative sensitivity analysis. Our setting values have been proven to be reliable and valid (Bernard 2013; Brzinsky-Fay and Kohler 2010). Moreover, we also verified that the following cluster analysis is not sensitive to the variations in distance matrices derived from different scoring systems.

Cluster Analysis

After analyzing all pairs of sequences, we are able to establish a similarity matrix in a symmetric manner. We first take opposite numbers to convert the matrix to a dissimilarity matrix. By subtracting the minimum negative number, the dissimilarity matrix is ensured to be positive, and is ready to be input as the distance matrix for cluster analysis. We apply the agglomerative hierarchical clustering with Ward's method to group customer browsing sequences into "nature patterns" (Aldenderfer and Blashfield 1984).

We use Silhouette coefficient to determine the best number of clusters. The individual Silhouette shows how well an object lies within its cluster, and separates from the neighbor clusters (Rousseeuw 1987). For a given sequence i, the Silhouette of i is calculated in equation

(1), where *a(i)* is the average distance between *i* and all other members in the same cluster; *b(i)* measures the average distance between *i* and all members in the nearest cluster. The average Silhouette of the whole sample provides an evaluation of clustering quality (Rousseeuw 1987), which is bounded between -1 (for bad clustering) and +1 (for good clustering). Therefore, we calculate and compare average Silhouettes for different numbers of clusters (Joseph et al. 2012), and choose the optimal number of clusters with the highest Silhouette (Rousseeuw 1987).

$$s(i) = \frac{b(i) - a(i)}{\max[a(i), b(i)]} \qquad (1)$$

In summary, N browsing patterns will emerge from the sequence analysis and cluster analysis. Based on these browsing patterns, we are able to further examine the differences of purchase rates among different types of consumers. The sequence analysis and clustering analysis are implemented in C++.

Hypothesis Testing

To test the proposed hypotheses, we first construct an individual's evolvement path by tracing the history of his/her browsing patterns. During this process, we address three major concerns. First, considering a customer may browse different product categories within a single store visit, we identify the dominant category of each visit as the one with more than half of product pages viewed belonging to that category. That is, the customer allocates most search efforts to the dominant category and gains most product information of that category in the session. In the cases where there is no such a dominant category, we posit the customer makes the visit to collect information for all the browsed categories. In our data, 156,989 out of 189,909 sessions (82.67%) have a dominant category, and the rest 32,920 (17.23%) do not. In this way, we trace how a consumer's browsing patterns evolve in searching information of a particular product category. Second, since our data points were captured within one month, consumers may have browsing history before this time period, or they may make a purchase after this time period. To address these, we calculate the average time interval between two store visits, which is 2.4 days (SD=3.8 days). That means, the majority of consumers revisit the store within 6 days since the last visit. Thus, in the analysis, we exclude consumers who have visit activities in the first 6 days, which ensures the rest consumers probably start the shopping journey within the sampled time period. However, we do not exclude consumers who may make a purchase in the next month, because it means the evolvement of browsing patterns until the end of the sampled time period does not lead to a purchase. Third, if there are multiple purchases within the time span, we separate the individual's evolvement path into several sub-paths, so that each evolvement path ends up with a purchase decision. In particular, we focus on an individual's evolvement paths which include two or more store visits because we are interested in consumers' evolving browsing behavior.

After constructing an individual's evolvement path of browsing patterns, we operationalize all the variables at the individual-evolvement level. Specifically, the dependent variable of purchase is a binary variable in which 1 represents the consumer purchases after several store visits. As to the independent variables of an individual's evolvement path, we use a dummy variable to represent one monotonic evolvement path. For example, the evolvement path of recommendation adjusted pattern -> recommendation adjusted pattern -> promotion adjusted pattern is labeled as 1 in the dummy variable representing the monotonic evolvement path of recommendation adjusted pattern -> promotion adjusted pattern. For each hypothesis test, we include a full set of dummy variables to represent all possible evolvement paths, exclude the one which serves as the basis of comparison, and focus on the evolvement path that is of primary interest. Lastly, regarding the control variables, we add "number of visits in the evolvement path", "number of previous visits", "average time interval between two visits", "whether there is a deep visit", and "whether there is a shallow visit".

Since the dependent variable is a binary variable, we employ logit regression models with product category fixed effects, which control for the potential heterogeneity of customer browsing behavior across product categories. The model specifications are as follows.

$$\Pr(Purchase_{i,j}=1\,|\,X)=\Lambda\left(\alpha_i+\beta\cdot GeneralShallow\_Rcm_{i,j}+\varphi\cdot OtherEvolPath_{i,j}+\delta\cdot Controls_{i,j}\right)\quad(1a)$$

$$\Pr(Purchase_{i,j}=1\,|\,X)=\Lambda\left(\alpha_i+\beta\cdot GeneralDeep\_Rcm_{i,j}+\varphi\cdot OtherEvolPath_{i,j}+\delta\cdot Controls_{i,j}\right)\quad(1b)$$

$$\Pr(Purchase_{i,j}=1\,|\,X)=\Lambda\left(\alpha_i+\beta\cdot General\_Prm_{i,j}+\varphi\cdot OtherEvolPath_{i,j}+\delta\cdot Controls_{i,j}\right)\quad(2)$$

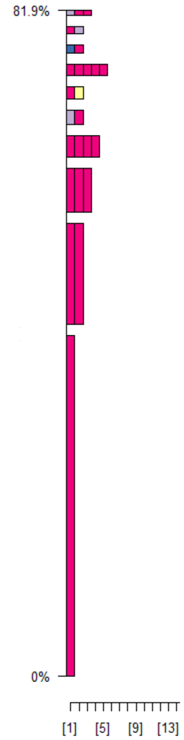
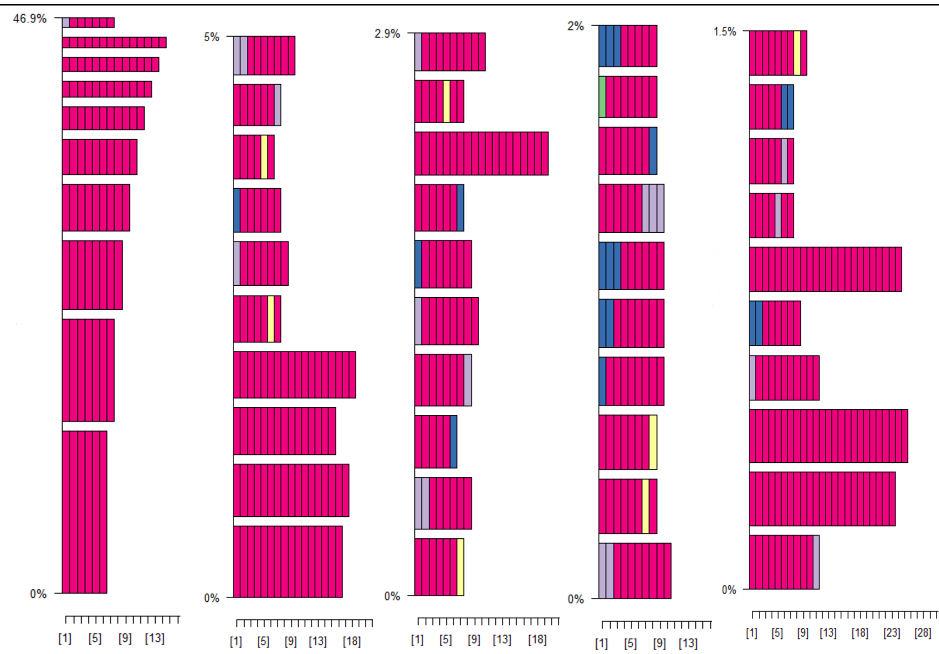
$$\Pr(Purchase_{i,j}=1\,|\,X)=\Lambda\left(\alpha_i+\beta\cdot Concrete\_Prm_{i,j}+\varphi\cdot OtherEvolPath_{i,j}+\delta\cdot Controls_{i,j}\right)\quad(3)$$

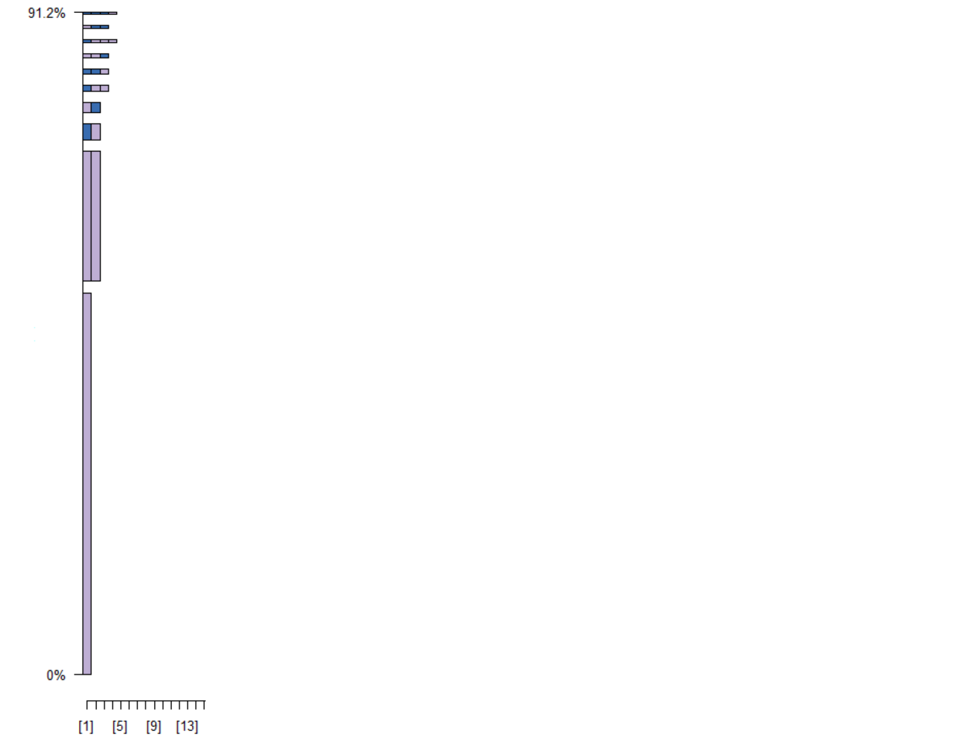
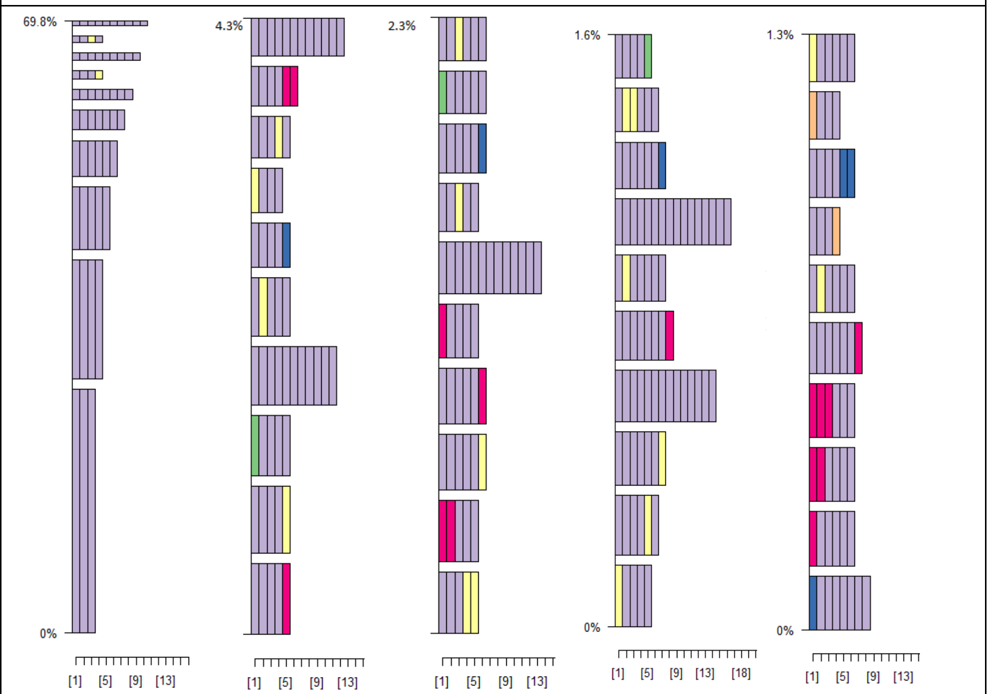
where $\Lambda(x)=e^x/(1+e^x)$, *i* refers to product category *i*, and *j* refers to the j^{th} evolvement path in product category *i*. In model (1a), (1b) and (2), the basis of comparison is the evolvement path of browsing patterns that are all guided by general product goals. In model (3), the basis of comparison is the evolvement path of browsing patterns that are all guided by concrete product goals. We exclude the corresponding dummy variable, which serves as the baseline condition, in the analysis. Across all models, our primary focus is β, which captures the effect of consumers' evolvement of browsing patterns, relative to the baseline condition, on their purchasing behavior.

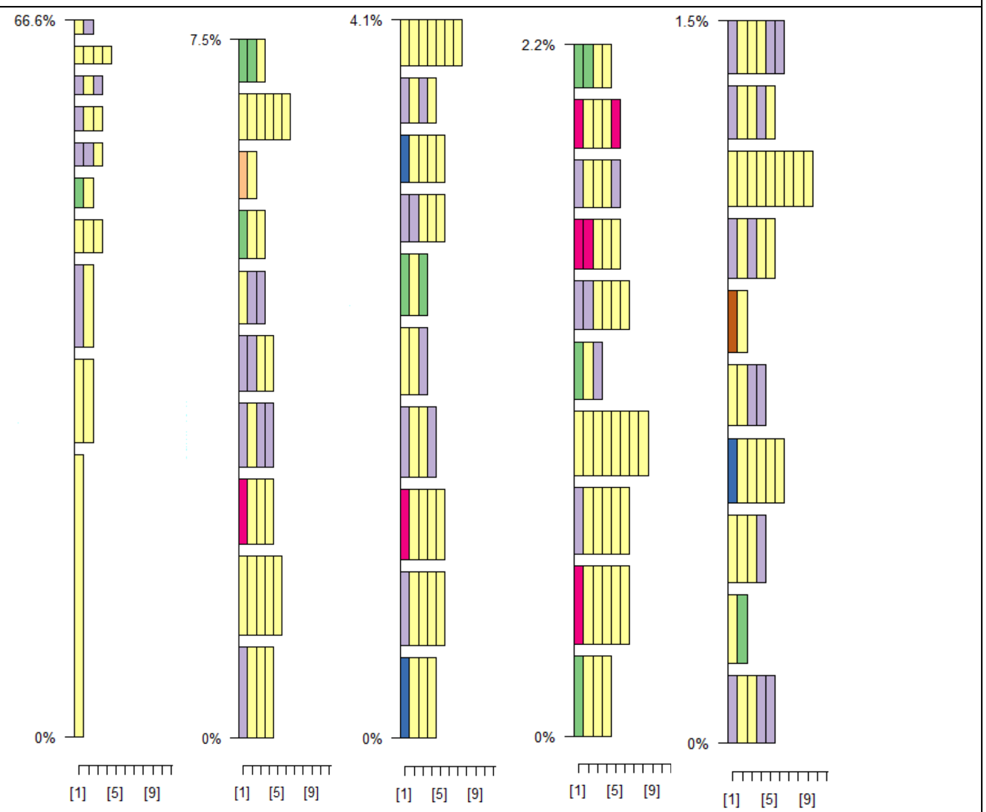
RESULT AND DISCUSSION

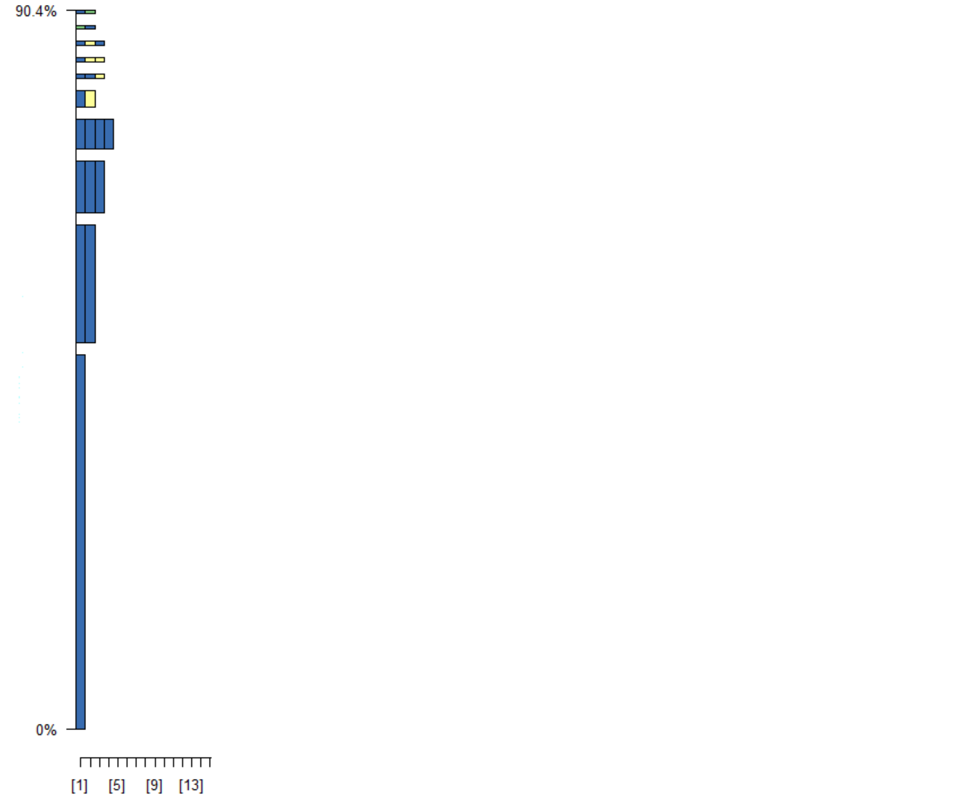
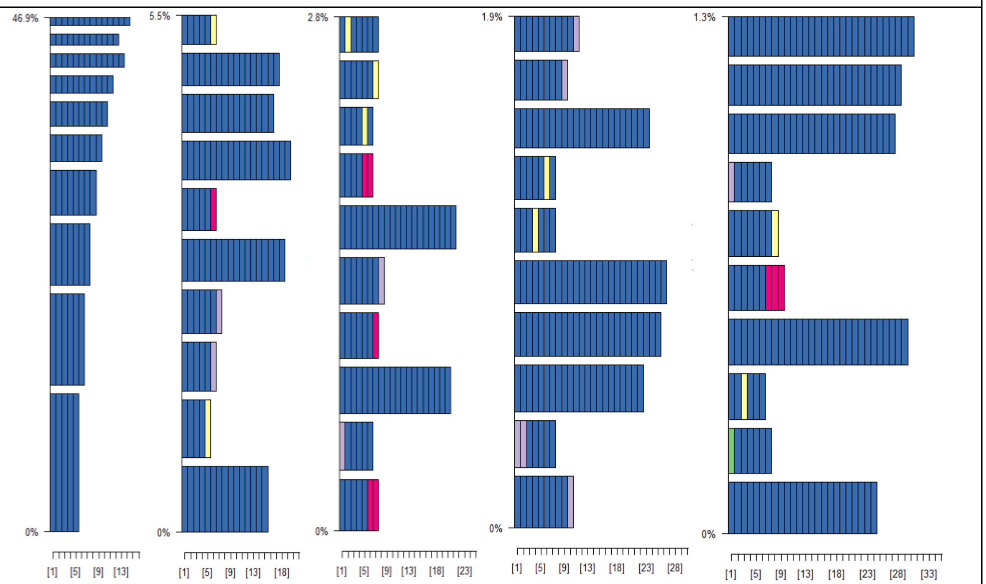
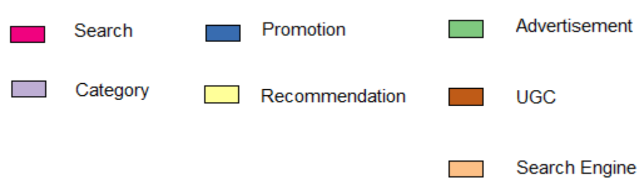
Typology of Online Browsing Patterns

By choosing the highest Silhouette coefficient (0.0010, SD=0.0014) among different numbers of clusters, we categorized the 189,909 sequences into eight clusters. Table 5 presents the most frequent sequences occurring in each cluster. Table 6 presents descriptive statistics of each browsing pattern. The overall visit-to-purchase conversion rate is 1.80% (SD=0.14), and the average sequence length is 2.78 (SD=4.78).

Table 5. Most Frequent Sequences in Each Cluster	
Cluster	**Information Seeking Strategy Sequences**
1 Quick Keyword Search (Obs.=53,423)	
2 Consecutive Keyword Search (Obs.= 3,487)	

3 Quick Category Search (Obs.= 44,264)	
4 Consecutive Category Search (Obs.=12,457)	

5 External Triggered Visit (Obs.= 20,436)	
6 Recommendation Adjusted Visit (Obs.=14,204)	

7 Quick Promotion Check (Obs.= 35,927)	
8 Consecutive Promotion Exploration (Obs.=5,711)	
	

Y axis represents the cumulated percentage of sequences; X axis represents the number of sub-sequences.

We first plot the top ten sequences which represent the majority of sequences in each browsing pattern. If the cumulated percentage of the top ten sequences is less than 80%, we include 40 more frequent sequences to demonstrate a representative sample of that browsing pattern.

Table 6. Descriptive Statistics of Each Browsing Pattern									
	Sample Average	**1. Quick Keyword Search**	**2. Consecutive Keyword Search**	**3. Quick Category Search**	**4. Consecutive Category Search**	**5. External Triggered Visit**	**6. Recommendation Adjusted Visit**	**7. Quick Promotion Check**	**8. Consecutive Promotion Exploration**
Obs. (189909)		53,423	3,487	44,264	12,457	20,436	14,204	35,927	5,711
Purchase Rate	1.80% (0.13)	1.73% (0.13)	8.52% (0.28)	0.94% (0.10)	2.14% (0.14)	0.28% (0.05)	2.16% (0.15)	1.39% (0.12)	11.26% (0.32)
Sequence Length	2.78 (4.78)	2.16 (1.80)	12.83 (21.02)	1.72 (1.44)	5.69 (3.94)	1.54 (1.30)	3.52 (4.04)	1.88 (1.34)	12.65 (12.27)
Number of Previous Visits	14.49 (61.08)	16.61 (73.53)	23.15 (66.27)	11.14 (48.64)	12.20 (63.72)	6.72 (28.65)	10.64 (33.62)	19.16 (71.02)	28.20 (73.37)
Number of Add-to-Cart Actions	0.12 (0.50)	0.07 (0.29)	0.42 (0.96)	0.04 (0.22)	0.18 (0.50)	0.03 (0.17)	0.49 (1.22)	0.07 (0.28)	0.48 (1.11)
Number of Categories	1.63 (1.52)	1.38 (0.81)	3.05 (2.66)	1.36 (0.84)	2.56 (1.87)	1.26 (0.70)	1.66 (1.17)	1.46 (0.84)	5.61 (4.81)
Number of Products in Each Category	1.66 (1.93)	1.59 (1.03)	5.92 (9.76)	1.25 (0.61)	2.76 (1.89)	1.20 (0.62)	2.18 (2.11)	1.30 (0.67)	3.07 (3.28)

The eight browsing patterns emerged from the analysis are consistent with consumers' dynamic browsing strategies as discussed in the theoretical background. From the aspect of shopping goals, consumers' store visits can be motivated by product goals (cluster 1-4), or no shopping goals (cluster 7). Furthermore, as indicated in the shopping goals theory, consumers' product-oriented goals are not always highly specified; they can be either concrete when described as precise key words (cluster 1-2), or abstract when construed as general categories (cluster 3-4). From the aspect of browsing paths, as planned path guided by goals, some consumers travel relatively shallow paths to collect necessary information (cluster 1, 3, 5), while others gather more product information along deep browsing paths (cluster 2, 4). Interestingly, when consumers encounter marketing stimuli, which happen to fulfill their needs (e.g., personalized recommendations and discount promotions), they are likely to adjust their browsing paths (cluster 6-8). Unexpectedly, we also observe customers exhibit both shallow visits (cluster 7) and deep visits (cluster 8) when their browsing paths are influenced by discount promotions. As a summary, Figure 2 depicts the theoretical framework of the eight browsing patterns.

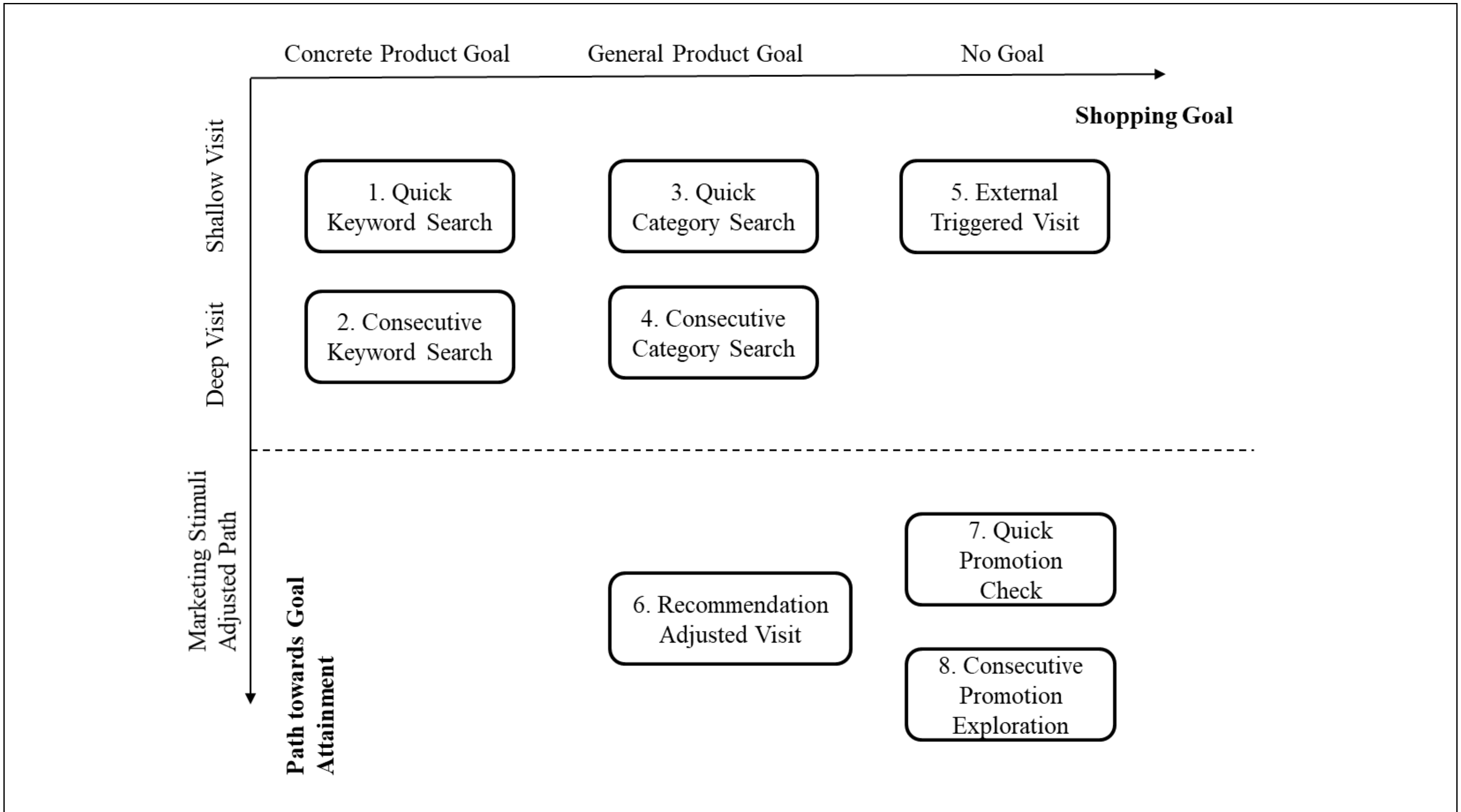


Figure 1.Theoretical Framework of Browsing Patterns

Cluster 1, which we label as “Quick Keyword Search”, contains 53,423 sessions (28.1% of the sample). From Table 5, we find Quick Search is dominated by the navigation source Search, which is occasionally preceded by Category and Promotion. It indicates consumers in this cluster mainly use keyword search to find product pages. Notably, the average number of searching terms is 2.30, implying consumers describe the targeted product with specific features. Thus, this type of consumers probably have a concrete shopping goal before entering the site (Janiszewski 1998). However, the average length of browsing paths in this cluster is only 2.16 (SD=1.80), indicating most consumers in Quick Search browsing pattern exit the website with two search actions. According to previous literature, this type of consumers can be either knowledgeable consumers who need a final confirmation before the purchase, or non-serious consumers who search products just for reference or collecting basic product information (Moe 2003). However, when we look at the purchase rate, only 1.73% of sessions in Cluster 1 end up with purchases, which implies consumers in this cluster are more likely to be non-serious shoppers.

Cluster 2, which we label as “Consecutive Keyword Search”, comprises 3,487 sessions (1.84% of the sample). Similar to Cluster 1, Consecutive Search browsing pattern is characterized by sessions with consistent search behavior, which is sometimes preceded by Category and Promotion. It suggests consumers in this cluster also have specific, precise goals in the browsing process. However, different from the first cluster, the sequence length of this cluster is much longer (12.83, SD=21.02). Moreover, Table 6 shows the purchase rate of Consecutive Search (8.52%) is notably high across all eight clusters. One possible explanation for this high conversion rate is that consumers in this cluster are probably at the relatively late stage of shopping, and are comparing alternatives to identify the optimal product to buy (Lee and

Ariely 2006; Moe 2006). This argument is supported by the highest product-to-category ratio among all patterns (see Table 6). During the process, the initial search result may be close to, but not perfectly meet consumers' expectations. In this case, they will strategically adjust their search key words based on the previous results and search again. Since consumers in the Consecutive Search pattern are constantly approaching their purchase target, it is highly likely that they will find the product they want and purchase it finally.

Cluster 3, labeled as "Quick Category Search", includes 44,264 sessions (23.31% of the sample). As illustrated in Table 5, this pattern is dominated by the navigational source Category. Different from keyword search, Category page lists a complete set of products and provides a broad level of information. As indicated in the extant literature, consumers with nonspecific goals usually seek such general information for quick knowledge acquisition (Vollmeyer et al. 1996). Nevertheless, consumers in this cluster only browse an average of 1.72 product pages (SD=1.44). The statistics imply they are likely to be in a very initial stage of shopping, where they collect basic information of the product category for the preparation of choice set construction. This hypothesis is supported by the low purchase rate of this cluster (0.94%, SD=0.10). Interestingly, consumers in this cluster tend to use Promotion as a complementary navigational source to Category. That is, consumers may first browse serval promotional items, and turn to category pages for a complete list of products; or alternatively, they may gain general knowledge from category pages, and then check promotional items in that category. These browsing strategies are consistent with the shopping goals theory which posits that consumers with abstract goals are more likely to be influenced by information cues, such as promotions (Lee and Ariely 2006).

Cluster 4, which is labeled as “Consecutive Category Search”, encompasses 12,457 sessions (6.56% of the sample). It mainly constitutes the information seeking strategy Category, suggesting consumers in this pattern also have relatively abstract goals during the search process. However, the sequence length of this cluster is significantly longer than that of cluster 3 (5.69, SD=3.94). In addition, Table 6 shows that the purchase rate of this pattern is slightly higher than the average (2.14%, SD=0.14). The evidence indicates that consumers are likely to be in the screening stage, where they screen a universe set of products in a focal category and reduce it to a smaller choice set (Moe 2006). If the alternatives evaluated happen to meet consumers’ needs, they will be likely to purchase in the current session. Notably, different from the previous pattern, the navigational sources complemented to Category are not constrained to Promotion in this cluster (see Table 6). With accumulated knowledge of the focal category, some consumers gradually switch to use keywords for more accurate search. Notably, some of them tend to be more receptive to information cues that are personalized (i.e., recommendations). By comparing the results in cluster 3 and cluster 4, we extend the shopping goals theory by showing that for consumers with abstract shopping goals, those who have accumulated more knowledge tend to be influenced by more specific contextual cues.

Cluster 5, which is labeled as “External Triggered Visit”, contains 20,436 sessions (10.76% of the sample). This browsing pattern, as the name implies, is triggered by external navigation sources such as Advertisement, Search Engine, and UGC. As illustrated in Tale 6, among these external sources, Advertisement is the most frequent session starter, which is often followed by other advertisements or internal navigation sources like Category and Search. That is, most of the consumers in this cluster enter the e-commerce website by clicking external links. For some of them, they may continue browsing other product pages after they enter the website.

However, we find that the conversion rate of this browsing pattern is the lowest across all clusters (0.28%, SD=0.05). This is probably because consumers who are triggered from external websites usually do not have a plan to purchase any products. Therefore, even if they are attracted by the advertisements and enter the website, most probably they will browse the website quickly, and exit without any purchase. This explanation is also supported by the relative short sequence length of this pattern (1.54, SD=1.30).

Cluster 6, which we label as "Recommendation Adjusted Visit", contains 14,204 sessions (7.48% of the sample). The cluster is mainly composed of information seeking strategy Recommendation, which is usually located at the late stage of each session (see Table 5). As can be seen from the table, consumers may start with other navigational sources as planned, such as Category, and switch to Recommendation later. Since recommendations are usually not expected by consumers before they enter the website, if the recommendations happen to satisfy their needs, the recommended products will probably be regarded as serendipities on their shopping journey and modified their planned browsing paths. This is the reason why we label the cluster as "Serendipitous Recommendation". Due to the serendipitous nature of recommendations, the average add-to-cart frequency, which is an indicator for customer revisit and future purchase behavior, is the highest among all patterns (0.49, SD=1.22); and the conversion rate of this cluster is 2.16% (SD=0.15), which outperforms the average conversion rate (1.80%, SD=0.13). Moreover, from Table 6, we can see that the average sequence length of this cluster is 3.52 (SD=4.04), which is also longer than the average sequence length of the whole sample (2.78, SD=4.78). It is inconsistent with the finding in the prior study which shows that recommendations, which provide greater variability, cause consumers to stop searching earlier than the condition of unassisted search (Dellaert and Häubl 2012). One possible explanation for the conflict is that if appropriate recommendations are encountered at the end of browsing, these recommendations will probably let consumers, who intended to exit, continue searching for relevant products.

Cluster 7, labeled as “Quick Promotion Check”, includes 35,927 sessions (18.92% of the sample). This cluster is dominated by the information seeking strategy Promotion, which is sometimes preceded and followed by Recommendation and Advertisement. It is apparent that customer visit paths in this cluster are significantly influenced by promotions. However, as shown in Table 6, most consumers only browse one or two product pages (1.88, SD= 1.34). Additionally, they are found to be regular visitors with an average of 19 (SD=71.02) previous visits to the site. Since promotions are usually updated on a regular basis at the e-commerce site, these regular consumers may be already familiar with the existing promotions and come to check for new updated promotions. Consumers in this cluster are not dedicated to pursuing economic benefits of purchasing promotional products, but rather they generate more hedonic values in hunting unexpected “treasures”. This argument can be confirmed by the relatively low conversion rate of this cluster (1.39%, SD=0.12).

Cluster 8, which we label as “Consecutive Promotion Exploration”, contains 5,711 sessions (3.01% of the sample). We find Utilitarian Search is dominated by the information seeking strategy Promotion, and this browsing pattern has the second longest average sequence length among all patterns (12.65, SD=12.27). Furthermore, it is worth noting that the conversion rate of this pattern is the highest among all clusters (11.26%, SD=0.32), which may be due to the utilitarian nature of this browsing pattern. Different from consumers in cluster 7 who enjoy hedonic values in encountering unexpected promotions, consumers in this cluster are probably price sensitive such that they focus on searching promotions in a persistent manner and aim at saving money. As indicated in Table 6, consumers explore the widest range of product categories among all patterns (5.61, SD=4.81), implying they do not have a concrete goal while browsing the site (Janiszewski 1998). As long as customers find the appropriate promotion, they will probably purchase the product. Moreover, as every promotion has a due date, in order not to miss the promotion, consumers in this pattern are more likely to purchase the product in the current session instead of delaying the purchase decision to the future.

Evolvement of Browsing Patterns

Table 7 and Table 8 present the effects of different evolvement paths of browsing patterns on purchases. Specifically, Column (1) to (3) in Table 7 demonstrate how the evolvement from general goal-oriented patterns into Recommendation Adjusted Visit affects consumers' purchase decisions. Column (1) indicates that for consumers whose browsing pattern evolves from general goal-oriented patterns, the evolvement into Recommendation Adjusted Visit has no overall effect on increasing purchases ($\beta = -0.267$, $p > 0.1$). However, when we further differentiate the general goal-oriented patterns based on the dimension of visit path, Recommendation Adjusted Visit exhibit opposite effects. As shown in Column (2), for consumers who have general goals, the evolvement into Recommendation Adjusted Visit improve customers' purchase probability by 81.5% ($p < 0.05$) when they follow a shallow visit path previously (i.e., evolvement from Quick Category Search). In contrast, when consumers follow a deep visit path previously (i.e., evolvement from Consecutive Category Search), the evolvement into Recommendation Adjusted Visit adversely reduces the purchase probability by 72.9% ($p < 0.05$). Thus, H1a and H1b are supported. In addition, as expected, Column (4) to (6) in Table 7 show that the evolvement from concrete goal-oriented patterns (i.e., evolvement from both Quick Keyword Check and

Consecutive Keyword Check) into Recommendation Adjusted Visit has no salient effects on purchases.

These findings suggest that although personalized recommendation is designed to elicit online consumers' preferences and thus improve sales, its effectiveness greatly depends on consumers' previous browsing behavior. Particularly, it is worth noting that for customers who have concretely defined preferences, recommendation has very limited influence on persuading them to buy. More interestingly, recommendations can even reduce customers' willingness to buy when they have evaluated an exhaustive set of alternatives with the guidance of ill-defined preferences. In such circumstance, recommendation, which is probably an option that has been viewed, prevents customers' purchases by confirming there are no potential choices in the store. These findings, which are undocumented in extant literature, provide valuable insights into the area of personalized recommendation (Kumar and Benbasat 2006; Xiao and Benbasat 2007).

In Table 8, Column (1) to (3) consistently reveal a positive impact of the evolvement from general goal directed pattern into promotion adjusted pattern on consumers' purchases. Specifically, the result in Column (1) indicates the evolvement into promotion adjusted pattern increase consumers' purchase probability by 101.2% ($p < 0.01$), which confirms H2. In a similar vein, Column (4) to (6) consistently show a positive effect of the evolvement from concrete goal directed pattern into promotion adjusted pattern on purchases. The result in Column (4) suggests the evolvement into promotion adjusted patterns improve a customer's purchase propensity by 107.5% ($p < 0.01$). Thus, H3 is also supported. These results imply the significant impact of monetary savings in customer decision making (Chandon et al. 2000).

Table 7. The Effect of Evolvement into Recommendation Adjusted Pattern on Purchases						
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: Purchase	Baseline: GeneralGoal	Baseline: GeneralGoal	Baseline: GeneralGoal	Baseline: ConcreteGoal	Baseline: ConcreteGoal	Baseline: ConcreteGoal
General_Rcm	-0.267 (0.243)					
GeneralShallow_Rcm		0.596** (0.271)				
GeneralDeep_Rcm			-1.306** (0.600)			
Concrete_Rcm				0.037 (0.196)		
ConcreteShallow_Rcm					-0.681 (0.614)	
ConcreteDeep_Rcm						0.122 (0.210)
NumVisit	0.059*** (0.012)	0.057*** (0.012)	0.058*** (0.012)	0.062*** (0.012)	0.061*** (0.012)	0.061*** (0.012)
NumPrevVisit	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)
AvgTimeInterval	-0.005*** (0.000)	-0.00502*** (0.000)	-0.005*** (0.000)	-0.005*** (0.000)	-0.005*** (0.000)	-0.005*** (0.000)
OtherEvolPath	Added	Added	Added	Added	Added	Added
Pseudo R-squared	0.057	0.055	0.057	0.054	0.054	0.054
No. Obs.	20,653	20,653	20,653	20,653	20,653	20,653
No. Categories	210	210	210	210	210	210

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 8. The Effect of Evolvement into Promotion Adjusted Pattern on Purchases

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: Purchase	Baseline: GeneralGoal	Baseline: GeneralGoal	Baseline: GeneralGoal	Baseline: ConcreteGoal	Baseline: ConcreteGoal	Baseline: ConcreteGoal
General_Prm	0.699*** (0.139)					
GeneralShallow_Prm		0.648*** (0.152)				
GeneralDeep_Prm			0.855*** (0.210)			
Concrete_Prm				0.730*** (0.121)		
ConcreteShallow_Prm					0.598*** (0.138)	
ConcreteDeep_Prm						1.086*** (0.212)
NumVisit	0.057*** (0.012)	0.058*** (0.012)	0.057*** (0.012)	0.058*** (0.012)	0.060*** (0.012)	0.061*** (0.012)
NumPrevVisit	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)
AvgTimeInterval	-0.005*** (0.000)	-0.005*** (0.000)	-0.005*** (0.000)	-0.005*** (0.000)	-0.005*** (0.000)	-0.005*** (0.000)
OtherEvolPath	Added	Added	Added	Added	Added	Added
Pseudo R-squared	0.055	0.055	0.055	0.055	0.054	0.054
No. Obs.	20,653	20,653	20,653	20,653	20,653	20,653
No. Categories	210	210	210	210	210	210

Standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$

CONTRIBUTIONS

Theoretical Contributions

This study makes valuable theoretical contributions to the IS and marketing literature. First, we provide a new typology of consumers' online browsing patterns. Due to the development of new technologies in e-commerce and the evolution of online shopping patterns, we find that the old typology of online shopping behaviors (Moe 2003) cannot meet the needs of distinguishing all types of browsing patterns. Therefore, this study extends the prior typology and builds the theoretical foundation for the emergence of new types of browsing patterns from the following three perspectives. First, we extend the general typology of browsing patterns, which divides consumer browsing behavior into goal-directed search and exploratory search, by identifying different types of consumers' goals at a more granular level (Lee and Ariely 2006). With the refined customers' goals, we demonstrate that even for exploratory search, it is still possible to infer the motivation of consumers' store visits based on the stimuli they follow (Janiszewski 1998). Second, different from Moe's (2003) typology of browsing behavior, which further divides goal-directed search and exploratory search by the time horizon of purchase, we focus on the nature of browsing behavior and afford another approach to categorize browsing behavior by navigation path towards goal attainment. We highlight that due to the low "transportation costs" for online store visits, instead of persistent search, many consumers choose to make multiple visits with short navigation paths for product knowledge acquisition, which is consistent with the effect of low-cost strategies in other domains (e.g., Paperny and Hedberg 1999). This new categorization builds the theoretical foundation for the emergence of Quick Keyword Search and Consecutive Keyword Search, Quick Category Search and Consecutive Category Search, Quick Promotion Check and Consecutive Promotion Exploration, where each pair of browsing patterns is guided by a same type of goals or information stimuli. Furthermore, we extend the previous literature by considering marketing stimuli in online stores. Based on the shopping goals theory, consumers' online navigation paths do not only depend on their planned strategies, but also rely on environmental cues, such as marketing stimuli, encountered that fulfill their needs (Lee and Ariely 2006).

Second, in addition to the theoretical division of browsing behavior, we also uncover eight previously undocumented browsing patterns from our multi-hierarchy analysis framework. Categorizing visits into different patterns helps researchers understand the unique characteristics of each type of browsing behavior. Therefore, another key contribution of this study is the new typology of browsing patterns, which provides a parsimonious framework for describing complex online consumer behavior (Doty and Glick 1994).

Third, we extend the literature of individual-level browsing behavior across store visits. Specifically, different from previous literature which characterizes a consumer's visiting history as a temporal point process (Park and Park 2016; Zhang et al. 2014), we leverage the browsing patterns to indicate the browsing activities that take place in each "point". Therefore, we are able to reveal the heterogenous effects of different evolvement paths, especially the evolvement into patterns influenced by marketing stimuli, on consumers' subsequent purchases. By considering online shoppers' evolvement of browsing patterns, our results also contribute to the literature on personalized recommendation (Komiak and Benbasat 2006; Kumar and Benbasat 2006; Xiao and Benbasat 2007) and discount promotion (Chandon et al. 2000; Stilley et al. 2010).

Finally, in our framework, in order to make the trade-off between the richness of clickstream data and the efficiency of analysis method, we contribute to the literature by using the critical points instead of the whole sub-sequences to represent consumers' page-viewing process. Moreover, we introduce the advanced data analysis techniques such as sequence analysis (Abbott and Tsay 2000) and cluster analysis (Aldenderfer and Blashfield 1984), which are rarely used in IS research (Joseph et al. 2012), to analyze the sequential information in clickstream data. Therefore, the novel methods lay the foundation for a comprehensive data analysis of browsing patterns upon which future research can continue to build and extend.

Practical Contributions

Our study also has several important practical implications on how to improve customization and increase the visit-to-purchase conversion rate at e-commerce websites. First, through the session-level analysis, we obtain the optimal eight browsing patterns, which provide a foundation for e-commerce companies to categorize each consumer visit into one of these eight patterns in a real-time manner. By doing so, online retailers are able to infer consumers' visiting motivations, and implement different strategies for different groups of customers. Moreover, the recognition of browsing patterns provides a great opportunity for websites to adjust content "on the fly" and provide customized services for a given user based on her ongoing browsing patterns (Bucklin et al. 2002). In this sense, our study takes a major step towards understanding online consumers' browsing behaviors as well as increasing their conversion rates.

Second, although Internet navigation is an evolving series of interrelated choices, where both marketers and consumers have an influence on shaping the context of subsequent choice events (Bucklin et al. 2002), most of previous studies analyzing clickstream data pay limited attention to the influence of marketing interventions, such as recommendations and promotions, on consumers' browsing behavior. In our study, as we consider all the possible information seeking strategies to reach the product page, the information sub-sequence analysis reveals that recommendation systems play an important role in helping consumers narrowing down their consideration sets and stimulating their purchase intention. Moreover, the sequence analyses suggest that consumers following Serendipitous Recommendation pattern have more chances to make a purchase at the website and tend to allow consumers to make a purchase decision within fewer store visits. Thus, a more accurate algorithm, which adjusts recommendations timely based on customers' accumulated browsing history, would be an effective approach to increase customer visit-to-purchase rates.

On the other hand, as to another frequently-used marketing strategy, we find although consumers following Consecutive Promotion Exploration have undoubtedly high purchase rates due to the utilitarian nature of visiting motivations, those exhibiting Quick Promotion Check turns out to have disappointingly low purchase rates. One possible explanation for the low purchase rates may be due to the lack of personalization techniques used in promotion information. In this case, for consumers with hedonic visiting motivations, the mismatch between promotions and their needs will lead to a low conversion rate. Thus, applying personalization techniques in promotion offers an opportunity for the improvement in conversion rate. In addition, sessions triggered by advertisement links or search result links at external websites have the lowest conversion rate among all the browsing patterns. It indicates consumers in this cluster are highly likely to be non-serious buyers.

Third, the analysis of the evolvement of browsing patterns suggests the provision of personalized recommendation or discount promotion should depend on customers' previous browsing patterns. Specifically, given the large expenditure of discount promotions for many companies, promotion may be substitutable in some cases. For instance, our results suggest for consumers whose browsing pattern evolves from Quick Category Search, recommendation can effectively help them narrow down choice set and stimulate their purchase intention. However, when recommendation has no effect (i.e., for consumers whose browsing pattern evolves from Quick Keyword Search or Consecutive Keyword Search) or even negative effect (i.e., for customers whose pattern evolves from Consecutive Category Search), promotion is a better approach to increasing sales.

LIMITATIONS

Although our study has both theoretical and practical implications, it also has some limitations, which provide opportunities for future research. First, different from traditional causal-explanatory statistical modeling studies in main-stream empirical IS research (Shmueli and Koppius 2011), our study, which focuses on designing new methodologies and creating empirical predictions, does not test any causal relationships between browsing patterns and purchase behavior. However, in spite of this limitation, our study is still useful for improving existing theories of browsing behaviors, developing new theories and algorithms of categorizing browsing patterns and assessing the predictability of purchase behavior. In order to examine the causal relationship between browsing patterns with marketing interventions (e.g., promotion) and purchase behavior, we may design a field experiment, in which a coupon is delivered to the consumers in the treatment group, in future research.

Another limitation of the study is that we only use the site-centric data rather than more powerful user-centric data to predict purchase rates in this study. According to Padmanabhan's (2006) study, this may lead to the inaccuracy of purchase rate prediction due to the incomplete picture of consumers' browsing activities. However, since we are only interested in consumers' session-level browsing patterns within the website and we are also able to know the navigation source even it is from external websites, the influence of cross-site visitation within a session on consumers' purchase decision is trivial. Therefore, it is not necessary to conduct our analysis based on the user-centric data. Moreover, as user-centric data are very expensive and difficult to obtain, a solution to efficiently utilize the available site-centric data seems to be more practical and valuable for most e-commerce merchants.

CONCLUSION

In conclusion, in order to improve customization and address the low visit-to-purchase conversion rate problem, we provide a new typology of online consumer browsing patterns. Based on the shopping goals theory, we lay a theoretical framework, from both dimensions of goal concreteness and path towards goal attainment, for the new typology. By leveraging cutting-edge techniques (e.g., sequence analysis from bioinformatics and clustering analysis from computer science), we obtain eight theoretically distinct browsing patterns from a large-scale clickstream dataset. In addition, we further study the evolvement of browsing patterns across an individual's multiple store visits. Our results reveal the effectiveness of marketing stimuli (e.g., personalized recommendation and discount promotion) in increasing purchases greatly depends on customers previous browsing patterns.

REFERENCES

Abbott, A. 1995. "Sequence Analysis: New Methods for Old Ideas," *Annual Review of Sociology* (21), pp. 93-113.

Abbott, A., and Forrest, J. 1986. "Optimal Matching Methods for Historical Sequences," *Journal of Interdisciplinary History* (16:3), pp. 471-494.

Abbott, A., and Tsay, A. 2000. "Sequence Analysis and Optimal Matching Methods in Sociology Review and Prospect," *Sociological Methods & Research* (29:1), pp. 3-33.

Aisenbrey, S., and Fasang, A.E. 2010. "New Life for Old Ideas: The" Second Wave" of Sequence Analysis Bringing the" Course" Back into the Life Course," *Sociological Methods & Research* (38:3), pp. 420-462.

Aldenderfer, M.S., and Blashfield, R. 1984. *Cluster Analysis*. Thousand Oaks, CA: Sage Publications.

Bardhi, F., and Arnould, E.J. 2005. "Thrift Shopping: Combining Utilitarian Thrift and Hedonic Treat Benefits," *Journal of Consumer Behaviour* (4:4), pp. 223-233.

Barsalou, L.W. 1991. "Deriving Categories to Achieve Goals," *The Psychology of Learning and Motivation: Advances in Research and Theory* (27), pp. 1-64.

Bernard, H.R. 2013. *Social Research Methods: Qualitative and Quantitative Approaches*. Thousand Oaks, CA: Sage.

Bloch, P.H., Sherrell, D.L., and Ridgway, N.M. 1986. "Consumer Search: An Extended Framework," *Journal of Consumer Research* (13:1), pp. 119-126.

Brzinsky-Fay, C., and Kohler, U. 2010. "New Developments in Sequence Analysis," *Sociological Methods & Research* (38:3), pp. 359-364.

Bucklin, R.E., Lattin, J.M., Ansari, A., Gupta, S., Bell, D., Coupey, E., Little, J.D., Mela, C., Montgomery, A., and Steckel, J. 2002. "Choice and the Internet: From Clickstream to Research Stream," *Marketing Letters* (13:3), pp. 245-258.

Chandon, P., Wansink, B., and Laurent, G. 2000. "A Benefit Congruency Framework of Sales Promotion Effectiveness," *Journal of Marketing* (64:4), pp. 65-81.

David, P., Song, M., Hayes, A., and Fredin, E.S. 2007. "A Cyclic Model of Information Seeking in Hyperlinked Environments: The Role of Goals, Self-Efficacy, and Intrinsic Motivation," *International Journal of Human-Computer Studies* (65:2), pp. 170-182.

Dellaert, B.G.C., and Häubl, G. 2012. "Searching in Choice Mode: Consumer Decision Processes in Product Search with Recommendations," *Journal of Marketing Research* (49:2), pp. 277-288.

Diehl, K., and Poynor, C. 2010. "Great Expectations?! Assortment Size, Expectations, and Satisfaction," *Journal of Marketing Research* (47:2), pp. 312-322.

Doty, D.H., and Glick, W.H. 1994. "Typologies as a Unique Form of Theory Building: Toward Improved Understanding and Modeling," *Academy of Management Review* (19:2), pp. 230-251.

Fogliatto, F.S., and Da Silveira, G.J. 2010. *Mass Customization: Engineering and Managing Global Operations*. Springer Science & Business Media.

Forrester Research. 2015. "The State of Retailing Online 2015: Key Metrics, Business Objectives and Mobile Benchmarks," Cambridge, MA.

Gollwitzer, P.M. 1999. "Implementation Intentions: Strong Effects of Simple Plans," *American Psychologist* (54:7), p. 493.

Gollwitzer, P.M., Heckhausen, H., and Steller, B. 1990. "Deliberative and Implemental Mind-Sets: Cognitive Tuning toward Congruous Thoughts and Information," *Journal of Personality and Social Psychology* (59:6), p. 1119.

Griffin, J.G., and Broniarczyk, S.M. 2010. "The Slippery Slope: The Impact of Feature Alignability on Search and Satisfaction," *Journal of Marketing Research* (47:2), pp. 323-334.

Hardesty, D.M., and Bearden, W.O. 2003. "Consumer Evaluations of Different Promotion Types and Price Presentations: The Moderating Role of Promotional Benefit Level," *Journal of Retailing* (79:1), 2003/01/01/, pp. 17-25.
Häubl, G., and Trifts, V. 2000. "Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids," *Marketing Science* (19:1), pp. 4-21.
Hauser, J.R., Urban, G.L., Liberali, G., and Braun, M. 2009. "Website Morphing," *Marketing Science* (28:2), pp. 202-223.
Herlocker, J.L., Konstan, J.A., Terveen, L.G., and Riedl, J.T. 2004. "Evaluating Collaborative Filtering Recommender Systems," *ACM Transactions on Information Systems (TOIS)* (22:1), pp. 5-53.
Hoffman, D.L., and Novak, T.P. 2000. "How to Acquire Customers on the Web," *Harvard Business Review* (78:3), pp. 179-188.
Huang, L.-T. 2016. "Flow and Social Capital Theory in Online Impulse Buying," *Journal of Business Research* (69:6), 2016/06/01/, pp. 2277-2283.
Hui, S.K., Fader, P.S., and Bradlow, E.T. 2009. "Path Data in Marketing: An Integrative Framework and Prospectus for Model Building," *Marketing Science* (28:2), pp. 320-335.
Janiszewski, C. 1998. "The Influence of Display Characteristics on Visual Exploratory Search Behavior," *Journal of Consumer Research* (25:3), pp. 290-301.
Joseph, D., Boh, W.F., Ang, S., and Slaughter, S. 2012. "The Career Paths Less (or More) Traveled: A Sequence Analysis of It Career Histories, Mobility Patterns, and Career Success," *MIS Quarterly* (36:2), pp. 427-452.
Komiak, S.Y.X., and Benbasat, I. 2006. "The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents," *MIS Quarterly* (30:4), pp. 941-960.
Kumar, N., and Benbasat, I. 2006. "Research Note: The Influence of Recommendations and Consumer Reviews on Evaluations of Websites," *Information Systems Research* (17:4), pp. 425-439.
Lee, L., and Ariely, D. 2006. "Shopping Goals, Goal Concreteness, and Conditional Promotions," *Journal of Consumer Research* (33:1), pp. 60-70.
Moe, W.W. 2003. "Buying, Searching, or Browsing: Differentiating between Online Shoppers Using in-Store Navigational Clickstream," *Journal of Consumer Psychology* (13:1), pp. 29-39.
Moe, W.W. 2006. "An Empirical Two-Stage Choice Model with Varying Decision Rules Applied to Internet Clickstream Data," *Journal of Marketing Research* (43:4), pp. 680-692.
Moe, W.W., Chipman, H., George, E.I., and McCulloch, R.E. 2002. "A Bayesian Treed Model of Online Purchasing Behavior Using in-Store Navigational Clickstream," Working paper, University of Texas at Austin, Austin.
Moe, W.W., and Fader, P.S. 2004. "Dynamic Conversion Behavior at E-Commerce Sites," *Management Science* (50:3), pp. 326-335.
Montgomery, A.L., Li, S., Srinivasan, K., and Liechty, J.C. 2004. "Modeling Online Browsing and Path Analysis Using Clickstream Data," *Marketing Science* (23:4), pp. 579-595.
Mount, D.W. 2004. *Bioinformatics: Sequence and Genome Analysis*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
Murphy, G.L., and Medin, D.L. 1985. "The Role of Theories in Conceptual Coherence," *Psychological Review* (92:3), p. 289.
Needleman, S.B., and Wunsch, C.D. 1970. "A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins," *Journal of Molecular Biology* (48:3), pp. 443-453.
Padmanabhan, B., Zheng, Z., and Kimbrough, S.O. 2006. "An Empirical Analysis of the Value of Complete Information for Ecrm Models," *MIS Quarterly* (30:2), pp. 247-267.
Paperny, D.N., and Hedberg, V.A. 1999. "Computer-Assisted Health Counselor Visits: A Low-Cost Model for Comprehensive Adolescent Preventive Services," *Archives of Pediatrics & Adolescent Medicine* (153:1), pp. 63-67.

Park, C.H., and Park, Y.-H. 2016. "Investigating Purchase Conversion by Uncovering Online Visit Patterns," *Marketing Science* (35:6), pp. 894-914.
Park, Y.-H., and Fader, P.S. 2004. "Modeling Browsing Behavior at Multiple Websites," *Marketing Science* (23:3), pp. 280-303.
Payne, J.W., Bettman, J.R., and Johnson, E.J. 1992. "Behavioral Decision Research: A Constructive Processing Perspective," *Annual Review of Psychology* (43:1), pp. 87-131.
Pervin, L.A. 1982. "The Stasis and Flow of Behavior: Toward a Theory of Goals," *Nebraska Symposium on Motivation*: University of Nebraska Press.
Pirolli, P. 1997. "Computational Models of Information Scent-Following in a Very Large Browsable Text Collection," *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*: ACM, pp. 3-10.
Pirolli, P., and Card, S. 1999. "Information Foraging," *Psychological Review* (106:4), p. 643.
Polyanovsky, V., Roytberg, M.A., and Tumanyan, V.G. 2011. "Comparative Analysis of the Quality of a Global Algorithm and a Local Algorithm for Alignment of Two Sequences," *Algorithms for Molecular Biology* (6:25), pp. 1-12.
Puccinelli, N.M., Goodstein, R.C., Grewal, D., Price, R., Raghubir, P., and Stewart, D. 2009. "Customer Experience Management in Retailing: Understanding the Buying Process," *Journal of Retailing* (85:1), 2009/03/01/, pp. 15-30.
Ricci, F., Rokach, L., and Shapira, B. 2015. "Recommender Systems: Introduction and Challenges," in *Recommender Systems Handbook,* F. Ricci, L. Rokach and B. Shapira (eds.). Boston, MA: Springer US, pp. 1-34.
Rousseeuw, P.J. 1987. "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *Journal of Computational and Applied Mathematics* (20), pp. 53-65.
Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. 2000. "Application of Dimensionality Reduction in Recommender System-a Case Study," Minnesota Univ Minneapolis Dept of Computer Science.
Shmueli, G., and Koppius, O.R. 2011. "Predictive Analytics in Information Systems Research," *MIS Quarterly* (35:3), pp. 553-572.
Simonson, I. 2005. "Determinants of Customers' Responses to Customized Offers: Conceptual Framework and Research Propositions," *Journal of Marketing* (69:1), pp. 32-45.
Sismeiro, C., and Bucklin, R.E. 2004. "Modeling Purchase Behavior at an E-Commerce Web Site: A Task-Completion Approach," *Journal of Marketing Research* (41:3), pp. 306-323.
Smart Insights. 2017. "Ecommerce Conversion Rates." Retrieved Jan 7th, 2018, from https://www.smartinsights.com/ecommerce/ecommerce-analytics/ecommerce-conversion-rates/
Stilley, K.M., Inman, J.J., and Wakefield, K.L. 2010. "Spending on the Fly: Mental Budgets, Promotions, and Spending Behavior," *Journal of Marketing* (74:3), pp. 34-47.
Vollmeyer, R., Burns, B.D., and Holyoak, K.J. 1996. "The Impact of Goal Specificity on Strategy Use and the Acquisition of Problem Structure," *Cognitive Science* (20:1), pp. 75-100.
Winterich, K.P., and Barone, M.J. 2011. "Warm Glow or Cold, Hard Cash? Social Identity Effects on Consumer Choice for Donation Versus Discount Promotions," *Journal of Marketing Research* (48:5), pp. 855-868.
Xiao, B., and Benbasat, I. 2007. "E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact," *MIS Quarterly* (31:1), pp. 137-209.
Xu, J.D., Benbasat, I., and Cenfetelli, R.T. 2014. "The Nature and Consequences of Trade-Off Transparency in the Context of Recommendation Agents1," *MIS Quarterly* (38:2).
Yadav, M.S., and Pavlou, P.A. 2014. "Marketing in Computer-Mediated Environments: Research Synthesis and New Directions," *Journal of Marketing* (78:1), pp. 20-40.

Zhang, Y., Bradlow, E.T., and Small, D.S. 2014. "Predicting Customer Value Using Clumpiness: From Rfm to Rfmc," *Marketing Science* (34:2), pp. 195-208.