

Summer 6-15-2016

CLICKSTREAM DATA AS A SOURCE TO UNCOVER CON-SUMER SHOPPING TYPES IN A LARGE-SCALE ONLINE SETTING

Daniel Schellong

RWTH Aachen University, schellong@time.rwth-aachen.de

Jan Kemper

RWTH Aachen University, kemper@time.rwth-aachen.de

Malte Brettel

RWTH Aachen University, brettel@time.rwth-aachen.de

Follow this and additional works at: http://aisel.aisnet.org/ecis2016_rp

Recommended Citation

Schellong, Daniel; Kemper, Jan; and Brettel, Malte, "CLICKSTREAM DATA AS A SOURCE TO UNCOVER CON-SUMER SHOPPING TYPES IN A LARGE-SCALE ONLINE SETTING" (2016). *Research Papers*. 1.

http://aisel.aisnet.org/ecis2016_rp/1

This material is brought to you by the ECIS 2016 Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in Research Papers by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

CLICKSTREAM DATA AS A SOURCE TO UNCOVER CONSUMER SHOPPING TYPES IN A LARGE-SCALE ONLINE SETTING

Research Paper

Schellong, Daniel, RWTH Aachen University, Aachen, Germany,
schellong@time.rwth-aachen.de

Kemper, Jan, RWTH Aachen University, Aachen, Germany, kemper@time.rwth-aachen.de

Brettel, Malte, RWTH Aachen University, Aachen, Germany, brettel@time.rwth-aachen.de

Abstract

Today's technological advancements enable marketers to track consumer touch points in great detail. We analyze the users' sequence of visits to a website (off-site clickstreams) as it gives insights into the overall decision-making process of consumers on their path to purchase or non-purchase. We show that search patterns based on the online advertising channel choice discloses specific shopping types. Operationalizing search behavior as navigational or informational based on information retrieval research and the level of consumer involvement; we use k-means clustering to categorize search patterns as Buying, Searching, Browsing or Bouncing. Our typology is based on a unique and comprehensive dataset from a leading European fashion e-commerce company and includes a total of almost 30 million clickstream journeys based on over 80 million lines of clicks from 11 advertising channels. This paper is the first to link the off-site online channel journey of consumers with their underlying search patterns to establish a typology of search types in a large-scale setting. Understanding the nature and characteristics of online consumer search and its implications for consideration and choice is also of high importance from a managerial perspective as our results offer e-marketers the ability to make inferences for automated marketing system decision-making.

Keywords: online consumer journey, clickstream analysis, advertising channel choice, big data

1 Introduction

Today, technological advancements enable marketers to track consumer choices along each individual's shopping journey in great detail, which is fundamentally different to the offline world (Kauffman et al., 2012; Nottorf and Funk, 2013; Winer, 2009). Clickstream path data plays a more and more important role in helping marketers, practitioners as well as researchers, to uncover consumer behavior online in large-scale settings. The term "clickstream" denotes the electronic record of a user's activity through one or more web sites and reflects a series of choices made in navigating the internet (Bucklin and Sismeiro, 2009; Bucklin et al., 2002; Kalczynski et al., 2006).

At the same time clickstream data tracks the exposure and user effects of internet advertising activities, such as clicks on advertising campaigns and subsequent conversions (Bucklin and Sismeiro, 2009; Chatterjee et al., 2003; Nottorf and Funk, 2013). For our study we use this type of clickstream data which we define as off-site clickstream data. In contrast to within-site data, off-site clickstream data is based on the users' sequence of visits to a website by off-site advertising channel choice over

time (Li and Kannan, 2014). Using off-site clickstream data offers several distinct advantages compared to within-site data: a) we look at the journey of visits per consumer and not only at single visits at a time, and b) consumer goal information is available in real-time the moment the consumer enters the website. Off-site clickstream data can thus provide insights into the choice process prior purchase as online shoppers form consideration sets (Bucklin et al., 2002; Li and Kannan, 2014).

The aim of our study is to detect and understand shopping goals of consumers by operationalizing search behavior as navigational or informational based on information retrieval research and the level of consumer involvement using a distinct set of clickstream measures. We use a k-means clustering approach to categorize the consumer search patterns.

For e-commerce firms it is more than ever success-critical to understand consumers' shopping behavior to improve marketing decision making. This paper is the first to link the off-site online channel journey of consumers with their underlying search patterns to establish a typology of search types in a large-scale online setting.

2 Related Literature

Our work is located at the intersection of several research streams: clickstream research, advertising effectiveness research and consumer behavior research. In the following we provide an overview of the most important studies per research stream relevant for our work.

Clickstream data can be categorized into within-site and across-site data. Across-site or user-centric clickstream data has been used to better understand information search behavior and switching costs across websites (Johnson et al., 2004; De los Santos et al., 2012; Park and Fader, 2004). Existing research on within-site or intra-store research focused mainly on search behavior (Moe and Yang, 2009; Zhou et al., 2007) and to model and predict conversion behavior on the website under consideration (Moe and Fader, 2004a, 2004b; Montgomery et al., 2004; Van den Poel and Buckinx, 2005; Sismeiro and Bucklin, 2004).

There has obviously been quite some research using advertising channel data in advertising research. Most of the studies using individual-level off-site clickstream data focus on single-channel effectiveness (Drèze and Hussherr, 2003; Lambrecht and Tucker, 2013; Manchanda et al., 2006; Rutz and Bucklin, 2012), on multi-channel effects using aggregated channel data (Breuer et al., 2011), on synergy effects such as carryover and spillover (Kireyev et al., 2013; Naik and Peters, 2009), or the attribution of conversion credit to each channel (Abhishek et al., 2012; Haan et al., 2013; Kireyev et al., 2013; Li and Kannan, 2014; Xu et al., 2014).

According to consumer behavior research, consumer's search and shopping behavior is heterogeneous depending on various personal, environmental, and social factors (Ganesh et al., 2010; Moon, 2004; Rohm and Swaminathan, 2004). At the same time online consumer behavior is different from the well-studied traditional behavior (Bucklin et al., 2002; Van den Poel and Buckinx, 2005). Extant research on online shopping typologies has clustered consumers based on their underlying reasons why people engage in shopping (e.g., see meta-studies by Rohm & Swaminathan 2004 or Ganesh et al. 2010 who provide a good overview on existing studies on consumer shopping types), however, all of these studies have been based on qualitative interviews or survey data. The only study using empirical clickstream data to come up with a typology of shopping behavior was conducted by Moe (2003). She used onsite data from 7,143 visit sessions from an online retailer of nutritional products and developed site usage metrics to cluster individual browsing patterns. Following his methodological approach, we attempt to identify shopping types based on a comprehensive set of off-site clickstream data from a leading European fashion retailer.

The remainder of the paper is structured as follows: in the next chapter we introduce our online consumer shopping type framework and the expected clickstream pattern characteristics per type based on

the respective theoretical grounding. After that, we describe our dataset including the measures used for the clustering procedure, before discussing the results and implications.

3 Conceptual Framework and Typology Development

Consumers' motivation to engage in shopping activities can have various reasons. The most obvious is the need for a specific product or service, but can also be purely for personal and social needs such as entertainment, adventure, or recreation (Arnold and Reynolds, 2003; Puccinelli et al., 2009; Tauber, 1998). Shopping behavior as part of the overall consumer decision-making process is also affected by the characteristics of the specific decision problem, including environment, context, and product characteristics (Bettman et al., 1991; Moon, 2004). Therefore, in order to uncover shopping goals we choose a single-firm empirical research setting to make sure that all consumers engage in the same environmental conditions during a defined time period.

The question on how shopping goals can be conceptualized using real data has not been researched much apart from a few exceptions (Moe, 2003; Puccinelli et al., 2009). Most of the existing works on consumer shopping goals use experimental or survey data to extract an assessment of consumers' general orientation and specific goals. One of the key elements of this paper is how do we uncover the types of search and shopping behavior. We choose off-site clickstream data as it forms a new and innovative approach taking into account the entire consumer shopping journey to offer new insights into online consumer behavior.

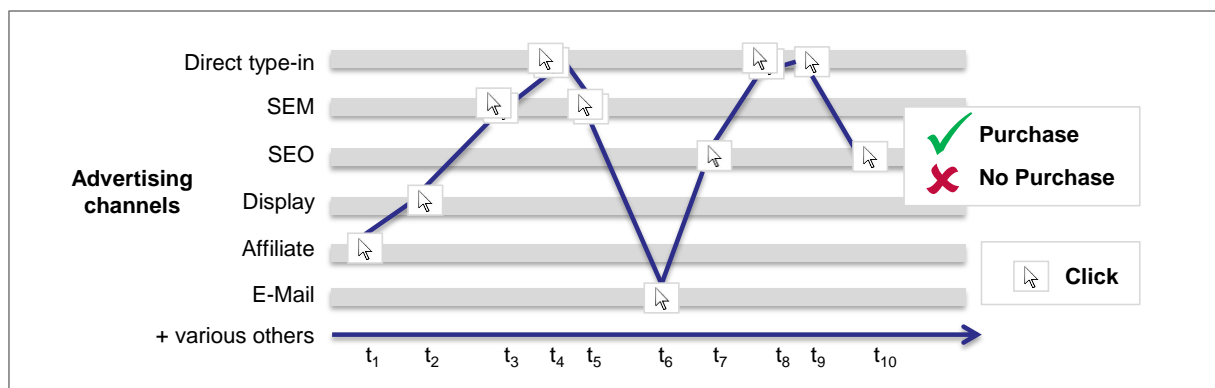


Figure 1. Example of an off-site clickstream

For better illustration purposes, Figure 1 provides an example of an off-site clickstream user journey. In this example, the first touch point was a click on an affiliate link at t_1 , which forwarded the user to the advertiser's landing page or product detail page. At t_2 the user clicked on a display advertisement and on t_3 the user searched for a relevant keyword on a search engine and clicked the respective sponsored advertisement the e-commerce firm was bidding on. The user continues his shopping process on various other channels until he finally converts, or does not convert (in this example at t_{10}).

We follow the general research framework from Moe (2003), who segments customers based on their search behavior and visit duration based on onsite clickstream data, however, adapt it in several ways: using off-site journey data rather than single-visit onsite data and replacing the visit horizon by the degree of involvement. We base our typology on two dimensions: consumer involvement and search behavior. From this theoretical framework we can classify shopping types as Buying, Searching, Browsing or Bouncing. We explicitly do not segment people, but rather shopping search modes because a consumer can have multiple shopping modes or types over time - for instance, a consumer can be a Browsing type in one month and a Buying type in the next month.

3.1 Typology Dimensions

Based on early theoretical work by Howard and Sheth (1969) and Engel et al. (1968) on consumer buying behavior, every consumer runs through multiple decision stages before making a purchase (also called purchase funnel). When consumers have different goals, it may point to different stages of their purchase decision-making process, or more precisely, varying degrees in how advanced consumers are as part of their process of consideration (or evoked) sets. In order to move from one stage to another in the purchase funnel consumers engage in various information acquisition activities (Rohm and Swaminathan, 2004). Consumer involvement is the degree to which the user is involved or engaged with the website helping to form his consideration set, pursuing information acquisition activities about a brand or product while considering alternatives before making a purchase or non-purchase (Puccinelli et al., 2009; Zaichkowsky, 1985). Hence, the involvement level is determined by the amount of information needed to satisfy the purchasing or information goal.

Based on Janiszewski (1998), we differentiate between two forms of search behavior: goal-directed and exploratory search. Goal-directed search is characterized as gathering relevant information regarding a specific or planned purchase in mind. Forming a consideration set, deliberating the options with a purchase or product type in mind to “get the job done”. Exploratory search on the other hand is less-focused, more browsing, undirected and rather stimulus driven with voluntary exposure to a varied set of product and product category considerations (Janiszewski, 1998). An exploratory shopper derives hedonic utility from the shopping process and is not motivated by any decision-making need (Arnold and Reynolds, 2003).

Table 1 summarizes the framework of online consumer shopping types.

Consumer Involvement	Search Behavior	
	Directed	Exploratory
High Involvement	<i>Buying</i>	<i>Browsing</i>
Low Involvement	<i>Searching</i>	<i>Bouncing</i>

Table 1. Online Consumer Shopping Types

3.2 Shopping Types

The “Buying Type” (Directed Search/High Involvement) has a specific shopping goal in mind when visiting the e-commerce store, to gather information regarding a product or the need to purchase a product. Visits to the store have a certain goal-directed strategy and the selection of the channels are mainly navigational in the way that the shopper is certain regarding his website choice where to find the relevant shopping information. For example, the user might simply type in the websites URL in the browsers address field (direct type-in), or click on an e-mail newsletter. In most cases the consumer converts at the website, not necessarily during his first visit, but eventually during one of the following visits. A further distinctive characteristic of being a “Buying” shopping type means regular frequent engagement within a rather short time horizon, collecting most relevant shopping information to satisfy the need to make a purchase or non-purchase decision.

Similar to the Buying Type, the “Searching Type” (Directed Search/Low Involvement) is also goal-directed. Not as focused as the Buying Type, the Searching Type is in the process of forming a consideration set in order to satisfy his information or shopping need, with the difference in the level of involvement. Whereas the Buying Type has a high engagement frequency during a rather short horizon, the Searching Type is not as engaged with a medium frequency during a mid-horizon.

The “Browsing Type” (Exploratory Search/High Involvement) differs to the Buying and Searching Type in the way that Browsing is characterized by an exploratory search behavior. Meaning the search

behavior is rather unplanned, without a specific utilitarian goal to be met. A Browser tends to be more focused on informational channels, as they offer more inspiration from a broad set of websites.

The “Bouncing Type” (Exploratory Search/Low Involvement) is the search type that is the most exploratory in the way that if engaged in a shopping process it is very much likely that the channels used are unfocused and that his involvement level is low. Bouncing translates into a single-visit with no further interaction or return to the e-commerce website.

Summing up, the framework of shopping types presented in Table 1 is characterized by involvement level (engagement frequency and horizon) and search behavior in the form of type of channels used (navigational or informational), resulting in 4 specific shopping types: Buying, Searching, Browsing and Bouncing.

In regards to clickstream conversion rate we would expect the Buying Type to have the highest conversion rate because the shoppers have a purchase goal in mind when engaging with the website followed by the Searching type as it is also goal-directed. Third the Browsing Type and lastly with a close to zero conversion rate the Bouncing Type.

Table 2 provides an overview on the expected shopping patterns per shopping/search type.

	Consumer Involvement		Search Behavior	Conversion Rate
	Channel Engagement Frequency	Channel Engagement Horizon	Browsing Goal	
	High/Mid/Low	Short/Mid/Long	Navigational/ Informational	
Buying Type	<i>High</i>	<i>Short</i>	<i>Navigational</i>	<i>High</i>
Searching Type	<i>Mid</i>	<i>Mid</i>	<i>Navigational</i>	<i>Mid</i>
Browsing Type	<i>Mid</i>	<i>Long</i>	<i>Informational</i>	<i>Mid</i>
Bouncing Type	<i>Low</i>	<i>Short</i>	<i>Informational</i>	<i>Low</i>

Table 2. Expected Shopping Pattern

For better illustration purposes of the shopping types we further illustrate the expected patterns in form of a timeline in Figure 2.

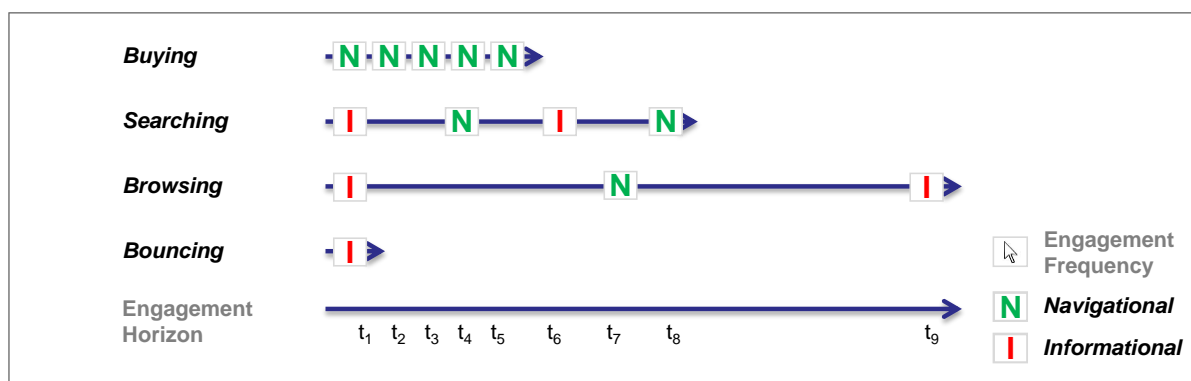


Figure 2. Example of an expected shopping type clickstream pattern

4 Data and Methodology

4.1 Research Partner Firm and Data Specifications

We were able to obtain a large and unique set of off-site clickstream data from a leading online-only fashion company across the European market. Our research partner uses a broad range of available online channels for advertising and engagement purposes, such as display reach, display retargeting, search engine marketing (SEM), search engine optimization (SEO), comparison, affiliate, social networks, and e-mail campaign links.

The unit of analysis in our study are off-site clickstreams depicting the individual's event-specific click behavior rather than the individual itself in the form of a consumer segment or persona due to the fact that consumers can have multiple motivations and hence behaviors over time (Puccinelli et al., 2009). We define off-site clickstreams the following way. If the consumer has not yet placed an order within a 30 day timeframe, every subsequent visit within this timeframe is an extension of the initiated shopping process. If the user journey leads to a purchase it forms a converting clickstream. Our sample holds 5.69% converting clickstreams (not to be mistaken for the visit conversion of an e-commerce site, which only focuses on the conversion percentage of single visits and is usually lower). If no conversion occurs within the 30 day period, all clicks made within the time period account to a non-converting clickstream.

The observation period covers the months of March and April of 2014 and the dataset consists of a total of 81,412,696 rows of clicks resulting in 29,939,213 off-site clickstream journeys. As our research partner firm is active in most European countries, this data is based on consumer clickstreams from Germany, Italy, Poland, Sweden, and the United Kingdom.

Each and every clickstream differs in terms of measures that describe it. Clickstreams are typically captured in semi-structured website log files. These website log files contain data elements such as a date and timestamp of each request, a visitor ID based on cookie information, the origin URL, and detailed information about the channel and advertising type clicked per visit. This data is transformed via a tracking solution that converts the unstructured log files into large tables of semi-structured data. However, companies struggle to use these vast amounts of available consumer path data to systematically create actionable marketing insights (Kauffman et al., 2012) – for applications such as the distribution of advertising budgets, more granular website personalization, or targeted advertising activities. Our research tries to close this gap by focusing on individual consumer search behavior during the consumer decision journey - from the initial need recognition, along the search and consideration process up to the purchase decision.

4.2 Measures to Operationalize Online Shopping Patterns

Based on real clickstream data, we define measures describing these patterns to identify search patterns of consumers. To operationalize shopping goals for the clustering procedure, we choose several variables in order to fully represent both framework dimensions: level of involvement and goal-direction of consumers based on the characteristics of the off-site clickstream dataset.

Table 3 provides a description of the measures used.

Measures	Unit	Description of Measures
<i>Engagement Frequency measure</i>		
CLICKS	<i>in # clicks</i>	<i>Total number of channel clicks to the e-commerce website</i>
<i>Engagement Horizon measures</i>		
TOTDURATION	<i>in days</i>	<i>Time between first and last channel click of customer clickstream journey</i>
CLICKGAP	<i>in days</i>	<i>Average time of gaps in between each visit of clickstream journey</i>
<i>Channel Focus measures</i>		
NAVISHARE	<i>in %</i>	<i>Share of navigational channel clicks as part of clickstream journey</i>
INFOSHARE	<i>in %</i>	<i>Share of informational channel clicks as part of clickstream journey</i>
<i>Channel Variety measure</i>		
UNIQUECH	<i>in #</i>	<i>Number of different channels used during clickstream journey</i>
<i>Conversion measure</i>		
CONVERSION	<i>in %</i>	<i>Indicator variable: "1" if CS ends with a purchase, "0" otherwise</i>

Table 3. Summary of clickstream measures

We operationalize involvement via the engagement measures frequency and horizon. We provide insights in the level of involvement in terms of amount of clicks, length, and intra-visit times of each clickstream, because they give the best insights into the degree of involvement of the individual with the e-commerce website. CLICKS stands for the amount of advertising channel clicks (or visits) per clickstream, independent of the type of channel used. TOTDURATION is the overall length or horizon of the clickstream, ranging up to maximum 30 days in duration. The third engagement measure CLICKS represents the average time in between each click, also called intra-visit times (TOTDURATION divided by the amount of CLICKS). Short average click gaps describe strong efforts in information acquisition. Taken together, engagement measures reveal the level of involvement and information acquisition motivation of the consumer. The High/Mid/Low involvement scale in Table 2 refers to how frequently consumers interact and the overall timeframe devoted to the shopping process, from a lot to a little.

Next to the level of involvement, we depict search behavior using the users browsing goal via the respective type of channel choice to identify directed compared to exploratory shoppers. We argue that the selection of channels entails information regarding the underlying shopping goal. Our research partner uses 11 different advertising channels (including direct type-in); however, using 11 variables would overfit the clustering model, make calculations overly complex and the number of unique clickstreams would grow exponentially. There are existing channel-categorizations which have been used in multi-channel advertising research. Most notably contact origin, separating between customer- and firm-initiated channels (Haan et al., 2013; Li and Kannan, 2014; Wiesel et al., 2011), or branded compared to generic channels (Jansen et al., 2011; Rutz and Bucklin, 2012). But to the best of our knowledge there is no existing categorization identifying user goal orientation. Therefore, we rely on a taxonomy developed in information retrieval research on user intention on the Internet and purchase decision-making theory. With the aim to understand the underlying goal of search, Broder (2002) proposes a categorization to classify web search queries into navigational and informational queries. In other words, a user's goal is deduced from the channels he or she uses. Based on these insights from information retrieval research we categorize online advertising channels based on the user's assumed browsing goal (Broder, 2002; Rose and Levinson, 2004). A user's goal in web keyword search is of navigational nature if the user wants to access "a specific website in mind". In contrast, the user's search is informational if he or she wants to learn something by simply reading or viewing web pages (Broder, 2002). We transfer these interdisciplinary findings to classify the broad array of online channels under investigation: Direct type-in, branded search (SEM Branded and SEO Branded), display retargeting and e-mail as navigational channels, and non-branded search (SEM Non-Branded and SEO Non-Branded), comparison, display and affiliate as informational.

Table 4 gives an overview of online marketing channels and respective goal categorization.

Channel	Description of Channel	Goal Categorization
		Navigational vs. Informational
Affiliate	Partner websites placing advertising material or links on behalf of the e-commerce firm in order to get compensated when a visitors clicks the link or purchases a product (e.g. fashion blog)	Informational
Comparison	Price comparison website where the e-commerce firm pays to be listed and provides a direct-link to the product detail page (e.g. pricegrabber.com)	Informational
CRM (E-Mail, etc.)	Any type of E-Mail marketing campaign (e.g. newsletter for certain promotion, new arrivals, product recommendations, etc.)	Navigational
Display - Reach	Placement of standard, non-personalized display banners on selected websites through the use of a display advertising network (including brand campaigns)	Informational
Display - Retarget	Placement of personalized display banners on selected websites based on previous browsing behavior - can be generic (targets users who have visited the website) or dynamic (targets users with the exact product the consumer has viewed before)	Navigational
DTI (Direct Type-In)	Consumers who directly access the e-commerce website by entering the URL in the browsers address box (or by bookmark, favorite or shortcut that leads directly to the predefined website)	Navigational
SEM Branded	Search Engine Advertising or Paid Search when a consumer enters keyword(s) in a search engine (mainly Google), and the keyword string includes the brand name of the e-commerce website under investigation	Navigational
SEM Non-Branded	Search Engine Advertising or Paid Search when a consumer enters keyword(s) in a search engine (mainly Google), and the keyword string does not include the brand name of the e-commerce website under investigation	Informational
SEO Branded	Search Engine Optimization or Organic Search when a consumer enters keyword(s) in a search engine (mainly Google), and the keyword string includes the brand name of the e-commerce website under investigation	Navigational
SEO Non-Branded	Search Engine Optimization or Organic Search when a consumer enters keyword(s) in a search engine (mainly Google), and the keyword string does not include the brand name of the e-commerce website under investigation	Informational
Social Media Owned	Social Media Channels (e.g. facebook, instagram, etc.) or any other extension of the e-commerce firms website	Navigational

Note: No Offline Channels (such as TV or Print) included

Table 4. Goal categorization of online marketing channels

The measures NAVISHARE for navigational clicks and INFOSHARE for informational clicks entail the share of visits that are navigational or informational per clickstream. This describes the degree of goal-direction in terms of website choice per clickstream. If the consumer is certain about his website choice as part of his shopping process NAVISHARE should be higher, if the consumer does not have a clear website goal and a rather broad consideration set in mind, then INFOSHARE should be higher.

In general, goal-directed shoppers tend to respond to channels which directly lead them to the target website that helps them make a better decision, whereas exploratory shoppers tend to choose channels that are rather unfocused and stimuli-driven such as Affiliate websites.

Furthermore, navigational behavior is of particular interest to search engine marketers, since navigational activity indicates prior intent to visit a website, possibly influenced by other advertising investments. Navigational channels (excluding retargeting) are in general more cost efficient than informational channels.

Although channel focus measures represent a good categorization in terms of the channel share regarding the consumers browsing goal, it does not give detail on the variety of different channels used per clickstream. UNIQUECH represents exactly this, the number of unique channels used.

Table 5 shows the descriptive statistics for each selected off-site clickstream measure.

Measures	Unit	Mean	SD	Min	Max
<i>Engagement Frequency measure</i>					
CLICKS	<i>in # clicks</i>	2.770	4.870	1.0	8954.0
<i>Engagement Horizon measures</i>					
TOTDURATION	<i>in days</i>	4.727	8.746	0.0	30.0
CLICKGAP	<i>in days</i>	1.136	2.408	0.0	15.0
<i>Channel Focus measures</i>					
NAVISHARE	<i>in %</i>	0.511	0.469	0.0	1.0
INFOSHARE	<i>in %</i>	0.364	0.449	0.0	1.0
<i>Channel Variety measure</i>					
UNIQUECH	<i>in #</i>	1.488	0.928	1.0	12.0
<i>Conversion measure</i>					
CONVERSION	<i>in %</i>	0.057	0.232	0.0	1.0

Note: Full dataset comprises 29,392,629 clickstreams based on a total of 81,412,696 clicks from March and April of 2014. For the clustering procedure all non-binary measures have been winsorized (5% of outliers on both ends have been replaced with the value of the 95%-quantiles respectively)

Table 5. Descriptive statistics of clickstream measures

4.3 Method

The aim of this study is to detect and understand online shopping goals of consumers by establishing a typology of search behavior. Therefore, the methodology applied to identify the shopping types is cluster analysis. We use a k-means clustering approach for several reasons: a) the technique is frequently used in marketing research, in particular in the field of segmentation and behavioral studies (Wedel and Kamakura, 2000), b) it serves as the main methodological tool in testing for consumer typologies and in revealing user patterns, especially in the field of clickstream research (Bucklin and Sismeiro, 2009; Dias and Vermunt, 2007; Hong and Kim, 2012; Moe, 2003), and c) k-means is a partitioning-based clustering method that works well for large datasets (Wedel and Kamakura, 2000). Furthermore, we build on a proven concept by extending the work of Moe (2003) who has used the same methodology with the same research goal, but for different kind of data (onsite data).

Based on our large dataset (>29 million lines of clickstreams) we use the k-means partitioning-based clustering algorithm developed by Hartigan and Wong (1979). Variables used for clustering are CLICKS, TOTDURATION, CLICKGAP, NAVISHARE, INFOSHARE and UNIQUECH. CONVERSION is not included and serves as profiling variable as it would skew clustering results in converting and non-converting clusters. All values are winsorized to set all outliers to the 95%-percentile in order to correct for both ends of the data but keep the values in the dataset. Furthermore, all non-binary variables are standardized for the clustering procedure using z-scores. We observed no multicollinearity issues (correlation > 0.7).

The k-means clustering algorithm tries to minimize the within cluster sum of squared distance errors between all points within one cluster. We run the clustering algorithm 20 times for all possible number of clusters and plot the values looking for the “elbow” criterion — number of clusters at which the sum of squared distance errors decreases abruptly — indicating the optimal number of clusters (Wedel and Kamakura, 2000). Ultimately, we select the 4-cluster solution as the decrease of sum of squared distance errors is only minimal in adding an additional fifth cluster.

For this study we do not apply a modeling or predictive approach since understanding the shopping type classification as well as framework and measure development is an essential first step and avenue to pursue by itself. Therefore, this study focuses on identifying measures that elicit shopping goals of consumers by establishing a typology of search behavior and the level of involvement.

5 Clustering Results

The optimal four-cluster solution is shown in Table 6. This fits to our proposed framework presented in Table 1. The patterns observed are further consistent with the characteristics of our theoretically derived typology in Table 2; if not, we provide an explanation in the remainder of the section.

Cluster	Unit	1 <i>BUYING</i>	2 <i>SEARCHING</i>	3 <i>BROWSING</i>	4 <i>BOUNCING</i>	Kruskal-Wallis- Test*
N	in CS in %	2,942,244 10.0%	3,936,826 13.4%	3,318,507 11.3%	19,195,052 65.3%	
<i>Engagement Frequency measure</i>						
CLICKS	in # clicks	7.941	3.248	3.176	1.196	***
<i>Engagement Horizon measures</i>						
TOTDURATION	in days	19.851	2.476	19.655	0.160	***
CLICKGAP	in days	2.187	0.746	5.622	0.066	***
<i>Channel Focus measures</i>						
NAVISHARE	in %	0.710	0.596	0.597	0.448	***
INFOSHARE	in %	0.287	0.393	0.362	0.369	***
<i>Channel Variety measures</i>						
UNIQUECH	in #	2.690	2.142	1.813	1.000	***
<i>Conversion measure</i>						
CONVERSION	in %	0.194	0.098	0.067	0.026	***
* = $p < .05$, ** = $p < .01$, *** = $p < .001$						
Note: All measures except CONVERSION were used for the k-means clustering procedures. All non-binary measures have been winsorized (5% of outliers on both ends have been replaced with the value of the 95%-quantiles respectively) as well as scaled (centering variables and creating respective z-scores)						

Table 6. Result of Cluster Solution

In order to test for significance of the cluster results we use the non-parametric statistical Kruskal-Wallis-Test (Hollander and Wolfe, 1973) showing highly significant differences in the cluster solution means across all variables and cluster results.

Cluster 1 “Buying Type” has the highest number of clicks (7.9) and the highest share of navigational channels (71%) as part of their path to purchase. They also use the highest number of channels (2.7), and an average total duration per clickstream of 19.8 days. This result regarding the engagement horizon is not in line with the expected pattern - the Buying Type will show a rather short horizon based on the fact that a goal-oriented shopper wants to satisfy his purchase or information need in a shorter timeframe. A possible explanation for this could be that the typical Buying Type is a loyal fan of the website with frequent visits to stepwise narrow down his consideration set over time to eventually convert on a regular basis (19.4%).

Cluster 2 “Searching Type” uses an average number of channels (3.2) in a medium timeframe (2.4 days). With a utilitarian task in mind gathering information on more navigational (0.6) compared to informational channels.

Cluster 3 “Browsing Type” as exploratory type searches over a longer period of time (19.6) before converting or not-converting on an average number of visits (3.1). The Browsing and Searching Types are quite similar except the duration of the clickstream is significantly different (19.6 Browsing compared to 2.4 days Searching Type). This points to a possible explanation that the classification between the two is more of a continuum rather than a binary decision. This means that if the right information or right inspirational stimuli is encountered during the consumer decision-process, the Browsing Type might transfer to a Searching Type or vice versa.

The Cluster 4 “Bouncing Type” which covers almost two-thirds of all clickstreams of our study (65.3%) has the lowest engagement measures (frequency and duration) compared to the other Types as well as the lowest navigational share (0.4) and the lowest share of converting clickstreams (2.6%). This cluster which shows characteristics as an “Others” cluster entails website visitors that leave the site shortly after and did not find right away what they were browsing for. Another explanation for the occurrence can be consumers who regularly delete cookies which makes it impossible to track their user journey on a consistent basis. Furthermore, consumers who exhibit a brand effect, meaning that the click on the advertising channel leading to the website has no immediate effect, but a future effect in regards to interactions with the e-commerce firm outside of our sample period.

6 Implications and Discussion

The aim of our study is to detect and understand goals and objectives of consumers in the form of shopping types. This paper extends existing clickstream research by combining off-site clickstream data with involvement theory and shopping goal behavior. Off-site clickstream data promises especially enriching insights about consumers, even before consumers enter the website of e-commerce firms.

We can confirm that consumer shopping goals show distinct search characteristics and allow a categorization based on their heterogeneous browsing patterns to a website. These findings build initial evidence that off-site clickstream data can greatly contribute to the understanding of individual consumer behavior online. Findings on distinct search characteristics are also significant enough to assume that off-site clickstream data are an excellent source for real-time user goal identification. We theoretically define and empirically validate four types of online shopping behaviors: Buying, Searching, Browsing and Bouncing. We further find that the higher the goal-direction of consumers in the form of share of navigational channels, the higher the purchasing propensity. Additionally, we further find initial confirming evidence that customers who visit a particular store frequently also tend to buy something during a relatively high proportion of those visits. Meaning the more involved and engaged a consumer is the higher his conversion rate.

Our study enhances the understanding of online consumer shopping behavior in e-commerce and illustrates how online journeys can differ depending on the consumer shopping motivation. An important managerial guideline for e-marketer is the fact that knowledge about consumer goals when entering the website (is he in the mood for Buying or just here for mere Browsing) enables marketers to specifically tailor the onsite experience (e.g., landing page). The different shopping type clusters can be used to develop marketing strategies, for instance, more targeted advertising campaigns such as e-mail offers can be targeted only to consumers who showed a principle willingness to convert at one of their previous visits. This offers e-retailers insights to optimize the customer journey on an individual basis with the goal of higher conversion rates and ultimately more loyal customers. We strongly believe our study can serve as a starting point for online marketing managers and customer management teams to integrate off-site clickstream analytics in consumer insight generation for an overall better shopping experience of consumers.

As with every empirical research study our work also has limitations to be reported. As we focus on single-company data, further research on this topic should investigate the stability of the results by performing similar studies with different companies, industries and customer groups. Overall, the aim for further research studies should be to try to analyze the entire shopping journey of consumers, including across-site information, and post-purchase behavior of consumers; any data element in addition to what we use would add another piece towards better online consumer behavior understanding.

Several avenues for future research exist and questions still remain. Including how do online shopping types differ by customer characteristics, or in different industry or cultural settings. Researchers could also combine shopping goals with customer and transactional data, in order to move from analyzing browsing behavior to better understanding the link with actual consumer purchasing behavior. As a

further step to make this research even more useful for practitioners a field study application setting-up various experiments to test if a personalized onsite experience (e.g., A/B test on landing page design) based on different shopping types will lead to direct effects on, for instance, conversion rate or basket size.

Despite the potential clickstream data analysis holds for research and practice in marketing, these data are almost surely being underutilized (Bucklin and Sismeiro, 2009; Kauffman et al., 2012). We are the first study to link individual off-site clickstreams with shopping goals of consumers. We hope our study motivates further research in the field that contributes to the understanding of consumer behavior in online retailing.

References

- Abhishek, V., Fader, P.S. and Hosanagar, K. (2012), "Media Exposure through the Funnel : A Model of Multi-Stage Attribution", *SSRN eLibrary*.
- Arnold, M.J. and Reynolds, K.E. (2003), "Hedonic shopping motivations", *Journal of Retailing*, Vol. 79 No. 2, pp. 77–95.
- Bettman, J.R., Johnson, E.J. and Payne, J.W. (1991), "Consumer decision making", *Handbook of consumer behavior*.
- Breuer, R., Brettel, M. and Engelen, A. (2011), "Incorporating long-term effects in determining the effectiveness of different types of online advertising", *Marketing Letters*, Vol. 22, pp. 327–340.
- Broder, A. (2002), "A taxonomy of web search", *ACM SIGIR Forum*, Vol. 36 No. 2, p. 3.
- Bucklin, R.E., Lattin, J.M., Ansari, A., Gupta, S., Bell, D., Little, J.D.C., Mela, C., et al. (2002), "From Clickstream Choice and the Internet : to Stream", *Marketing Letters*, Vol. 13, pp. 245–258.
- Bucklin, R.E. and Sismeiro, C. (2009), "Click Here for Internet Insight: Advances in Clickstream Data Analysis in Marketing", *Journal of Interactive Marketing*, Vol. 23 No. 1, pp. 35–48.
- Chatterjee, P., Hoffman, D.L. and Novak, T.P. (2003), "Modeling the Clickstream: Implications for Web-Based Advertising Efforts", *Marketing Science*, Vol. 22 No. 4, pp. 520–541.
- Dias, J.G. and Vermunt, J.K. (2007), "Latent class modeling of website users' search patterns: Implications for online market segmentation", *Journal of Retailing and Consumer Services*, Vol. 14, pp. 359–368.
- Drèze, X. and Hussherr, F.-X. (2003), "Internet advertising: Is anybody watching?", *Journal of Interactive Marketing*, Vol. 17 No. 4, pp. 8–23.
- Engel, J.F., Kollat, D.T. and Blackwell, R.D. (1968), *Consumer behavior*, Holt, Rinehart, and Winston.
- Ganesh, J., Reynolds, K.E., Luckett, M. and Pomirleanu, N. (2010), "Online Shopper Motivations, and e-Store Attributes: An Examination of Online Patronage Behavior and Shopper Typologies", *Journal of Retailing*, New York University, Vol. 86 No. 1, pp. 106–115.
- Haan, E. De, Wiesel, T. and Pauwels, K. (2013), "Which Advertising Forms Make a Difference in Online Path to Purchase ?", *Marketing Science Institute Working Paper Series*, Vol. 13 No. 13.
- Hartigan, J. a. and Wong, M. a. (1979), "A K-Means Clustering Algorithm", *Journal of the Royal Statistical Society*, Vol. 28 No. 1, pp. 100–108.
- Hong, T. and Kim, E. (2012), "Segmenting customers in online stores based on factors that affect the customer's intention to purchase", *Expert Systems with Applications*, Elsevier Ltd, Vol. 39 No. 2, pp. 2127–2131.
- Howard, J.A. and Sheth, J.N. (1969), *The theory of buyer behavior*, Wiley, New York.
- Janiszewski, C. (1998), "The Influence of Display Characteristics on Visual Exploratory Search Behavior", *Journal of Consumer Research*, Vol. 25 No. December, pp. 290–301.
- Jansen, B.J., Sobel, K. and Zhang, M. (2011), *The Brand Effect of Key Phrases and Advertisements in Sponsored Search*, *International Journal of Electronic Commerce*, Vol. 16.
- Johnson, E.J., Moe, W.W., Fader, P.S., Bellman, S. and Lohse, G.L. (2004), "On the Depth and Dynamics of Online Search Behavior", *Management Science*, Vol. 50 No. 3, pp. 299–308.
- Kalczynski, P.J., Senecal, S. and Nantel, J. (2006), "Predicting On-Line Task Completion with Clickstream Complexity Measures: A Graph-Based Approach", *International Journal of Electronic Commerce*, Vol. 10 No. 3, pp. 121–141.
- Kauffman, R.J., Srivastava, J. and Vayghan, J. (2012), "Business and data analytics: New innovations for the management of e-commerce", *Electronic Commerce Research and Applications*, Vol. 11 No. 2, pp. 85–88.
- Kireyev, P., Pauwels, K. and Gupta, S. (2013), "Do Display Ads Influence Search? Attribution and Dynamics in Online Advertising", *Harvard Business School Working Paper*, pp. 1–30.

- Lambrecht, A. and Tucker, C. (2013), "When Does Retargeting Work? Information Specificity in Online Advertising", *Journal of Marketing Research*, American Marketing Association, Vol. 50 No. 5, pp. 561–576.
- Li, H. (Alice) and Kannan, P.K. (2014), "Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment", *Journal of Marketing Research*, Vol. 51 (Februa No. 1, pp. 40–56.
- De los Santos, B., Hortaçsu, A. and Wildenbeest, M.R. (2012), "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior", *American Economic Review*, Vol. 102 No. 6, pp. 2955–2980.
- Manchanda, P., Dubé, J.-P., Goh, K.Y. and Chintagunta, P.K. (2006), "The Effect of Banner on Internet Advertising Purchasing", *Journal of Marketing Research*, Vol. 43 No. 1, pp. 98–108.
- Moe, W.W. (2003), "Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream", *Journal of Consumer Psychology*, Vol. 13 No. 1-2, pp. 29–39.
- Moe, W.W. and Fader, P.S. (2004a), "Capturing evolving visit behavior in clickstream data", *Journal of Interactive Marketing*, Vol. 18 No. 1, pp. 5–19.
- Moe, W.W. and Fader, P.S. (2004b), "Dynamic Conversion Behavior at E-Commerce Sites", *Management Science*, Vol. 50 No. 3, pp. 326–335.
- Moe, W.W. and Yang, S. (2009), "Inertial Disruption: The Impact of a New Competitive Entrant on Online Consumer Search", *Journal of Marketing*, American Marketing Association, Vol. 73 No. 1, pp. 109–121.
- Montgomery, A.L., Li, S., Srinivasan, K. and Liechty, J.C. (2004), "Modeling Online Browsing and Path Analysis Using Clickstream Data", *Marketing Science*, Vol. 23 No. 4, pp. 579–595.
- Moon, B.-J. (2004), "Consumer adoption of the internet as an information search and product purchase channel: some research hypotheses", *International Journal of Internet Marketing and Advertising*, Vol. 1 No. 1, pp. 104–118.
- Naik, P.A. and Peters, K. (2009), "A Hierarchical Marketing Communications Model of Online and Offline Media Synergies", *Journal of Interactive Marketing*, Elsevier Inc., Vol. 23 No. 4, pp. 288–299.
- Nottorf, F. and Funk, B. (2013), "The economic value of clickstream data from an advertiser's perspective", *Proceedings of the 21st European Conference on Information Systems*, pp. 1–12.
- Park, Y.-H. and Fader, P.S. (2004), "Modeling Browsing Behavior at Multiple Websites", *Marketing Science*, Vol. 23 No. 3, pp. 280–303.
- Van den Poel, D. and Buckinx, W. (2005), "Predicting online-purchasing behaviour", *European Journal of Operational Research*, Vol. 166 No. 2, pp. 557–575.
- 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*, Vol. 85 No. 1, pp. 15–30.
- Rohm, A.J. and Swaminathan, V. (2004), "A typology of online shoppers based on shopping motivations", *Journal of Business Research*, Vol. 57 No. 7, pp. 748–757.
- Rose, D. and Levinson, D. (2004), "Understanding user goals in web search", *the 13th international conference on World Wide Web*, pp. 13–19.
- Rutz, O.J. and Bucklin, R.E. (2012), "Does banner advertising affect browsing for brands? Clickstream choice model says yes, for some", *Quantitative Marketing and Economics*, Vol. 10 No. 2, pp. 231–257.
- Sismeiro, C. and Bucklin, R.E. (2004), "Modeling Purchase Behavior at an E-Commerce Web Site: A Task-Completion Approach", *Journal of Marketing Research*, Vol. 41 No. 3, pp. 306–323.
- Tauber, E. (1998), "Why Do people Shop", *Journal of Marketing*, Vol. 33 No. 3, pp. 53–55.
- Wedel, M. and Kamakura, W. (2000), *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publishers, Boston.
- Wiesel, T., Pauwels, K. and Arts, J. (2011), "Practice Prize Paper--Marketing's Profit Impact:

- Quantifying Online and Off-line Funnel Progression”, *Marketing Science*, Vol. 30 No. 4, pp. 604–611.
- Winer, R.S. (2009), “New Communications Approaches in Marketing: Issues and Research Directions”, *Journal of Interactive Marketing*, Direct Marketing Educational Foundation, Inc., Vol. 23 No. 2, pp. 108–117.
- Xu, L., Duan, J.A. and Whinston, A. (2014), “Path to Purchase: A Mutually Exciting Point Process Model for Online Advertising and Conversion”, *Management Science*, INFORMS, Vol. 60 No. 6, pp. 1392–1412.
- Zaichkowsky, J. (1985), “Measuring the Involvement Construct”, *Journal of Consumer Research*.
- Zhou, L., Dai, L. and Zhang, D. (2007), “Online Shopping Acceptance Model: Critical Survey of Consumer Factors in Online Shopping”, *Journal of Electronic Commerce Research*, Vol. 8 No. 1, pp. 41–62.