LAPPEENRANTA UNIVERSITY OF TECHNOLOGY

LUT School of Business and Management
Master's Degree in Strategy, Innovation, and Sustainability

Michael Lindén

Path Analysis of Online Users Using Clickstream Data: Case Online Magazine Website.

1st Examiner: Sanna-Katriina Asikainen
2nd Examiner: Mikael Collan

ABSTRACT

Author:	Michael Lindén
Title:	Path Analysis of Online Users Using Clickstream Data: Case Online Magazine Website.
Faculty:	LUT School of Business and Management
Major:	Master's in Strategy, Innovation and Sustainability
Year:	2016
Master's Thesis:	Lappeenranta University of Technology
	75 pages, 19 tables, 18 figures
Examiners:	Prof. Sanna-Katriina Asikainen and Prof. Mikael Collan
Keywords:	clickstream, pattern, user behavior, magazine, website, path analysis

This paper explores behavioral patterns of web users on an online magazine website. The goal of the study is to first find and visualize user paths within the data generated during collection, and to identify some generic behavioral typologies of user behavior.

To form a theoretical foundation for processing data and identifying behavioral archetypes, the study relies on established consumer behavior literature to propose typologies of behavior. For data processing, the study utilizes methodologies of applied cluster analysis and sequential path analysis.

Utilizing a dataset of click stream data generated from the real-life clicks of 250 randomly selected website visitors over a period of six weeks. Based on the data collected, an exploratory method is followed in order to find and visualize generally occurring paths of users on the website. Six distinct behavioral typologies were recognized, with the dominant user consuming mainly blog content, as opposed to editorial content. Most importantly, it was observed that approximately 80% of clicks were of the blog content category, meaning that the majority of web traffic occurring in the site takes place in content other than the desired editorial content pages. The outcome of the study is a set of managerial recommendations for each identified behavioral archetype.

Tiivistelmä

Tekijä:	Michael Lindén
Tutkielman nimi:	Verkkokäyttäjien Polkuanalyysi Clickstream Dataa Käyttäen: Case Verkkolehti.
Tiedekunta:	LUT Kauppatieteet ja Tuotantotalous
Pääaine:	Master's in Strategy, Innovation and Sustainability
Vuosi:	2016
Pro-Gradu Tutkielma:	Lappeenrannan Teknillinen Yliopisto
	75 sivua, 19 taulukkoa, 18 kuvaa,
Tarkastajat:	Prof. Sanna-Katriina Asikainen ja Prof. Mikael Collan
Hakusanat:	clickstream; käyttäytymismallit; aikakauslehti; verkkosivu; polkuanalyysi

Tämän pro-gradu tutkielman tarkoituksena oli tutkia verkkokäyttäjien käyttäytymismalleja aikakauslehden verkkoversiossa. Tutkimuksen tarkoituksena on löytää ja visualisoida käyttäpolkuja kerätyn datan perusteella, ja tunnistaa niiden perusteella joitain yleistettäviä käyttäytymis arkkityyppejä.

Teoreettisena pohjana työ käyttää olemassa olevaa kuluttajakäyttäytymiseen perehtyvää kirjallisuutta käyttäytymistypologioiden muodostamiseksi. Datan käsittely puolestaan perustuu sovellettuun klusterianalyysin ja sekvenssianalyysin metodologioihin.

Työ käyttää hyväkseen clickstream datasettiä, joka koostuu 250 satunnaisesti valitun verkkokäyttäjän klikkaustiedosta, jota kerättiin kuuden viikon ajan. Datan perusteella on löydetty ja visualisoitu yleisesti ilmeneviä polkuja verkkosivun sisällä. Kuusi käyttäytymisarkkityyppiä tunnistettiin datan perusteella, joista dominoivin typologia keskittyy blogi-sisältöön, toimituksellisen sisällön sijasta. Tärkeä havainto oli, että jopa 80% klikkauksista kohdistui blogi sisältöön, joka tarkoittaa käytännössä sitä, että suurin osa sivuston liikenteestä ei tapahdu halutulla toimituksellisen sisällön sivuilla. Tutkimuksen tuloksena ehdotetaan käytännön lähestymistapoja kunkin arkkityypin kohdistamiseksi.

ACKNOWLEDGEMENTS

Writing this thesis has been a challenging journey, and finally completing it is thanks to many reasons beyond myself.

First and foremost, I would like to thank my original supervisor Hanna-Kaisa Ellonen for offering the topic to me. Particularly during times of low motivation, Hanna-Kaisa was the primary source of encouragement to bring me back to working on the thesis after prolonged hiatuses. The scope and focus of the study changed many times along the way, but somehow under Hanna-Kaisa's tutelage I managed to keep the essential purpose of the research in mind. The work has been a learning process in new methodologies for both us, and I'm sure I couldn't have finished it without her guidance.

Secondly I would like to thank Sanna-Katriina for taking over Hanna-Kaisa's position as primary supervisor after she left to pursue new opportunities in her professional life. Sanna-Katriina, together with Mikael Collan, provided the needed advice to finally wrap up the work.

I also want to thank my family and girlfriend for remaining supportive throughout the process of completing the thesis. On some days, the support was needed to open up the thesis and continue working, when I really didn't want to.

Lastly I would like to give thanks to the experts at the case company for providing me with access to real user data, without which it would have been impossible to complete hands-on thesis such as this.

Michael Niclas Christopher Lindén

21.3.2016 Lappeenranta

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

The purpose of this study is to examine the behavior of website users of an online magazine Web site. The key question to be explored is whether traditional consumer behavior models correspond with modern day online user profiles. Understanding online user behavior in this context will allow better development of the website content, as well as reaching the right target audience.

Modern websites have come a long way since their early conception. Whereas once websites were more static than interactive, sites of today typically incorporate a multitude of content and information. With an ever expanding amount of content, as well as integration with social media and the Web 2.0, companies are left with an overwhelming tool set with which they can potentially engage and interact with their customers, both existing and potential ones. With the increasing complexity of Web sites, it becomes equally important to understand how the Web site users actually navigate the site, so as to understand which content is found, what is enjoyed, and what keeps bringing users back to the site.

This brings us to the highly topical world of modern data analytics. Modern technology enables companies to identify and track the paths of users on a particular website, which in turn allows for companies to gain a more visual understanding of how their Web site's content is being consumed. Furthermore, following the paths of consumers from the beginning of their sessions to the end enables companies to identify key partners and content co-creators. This is crucial from a business stand-point, as it has been noted companies often fail to properly exploit their vast toolsets for understanding user behavior, and as such cannot contribute to overall business objectives. (Lee, 2002) Further, companies are finding it increasingly difficult to even stay relevant to their user bases, as the online environment has become a hypercompetitive arena where customers have a great deal of power over the companies purely through the variety of choice they have (Porter, 2001; Constantinides, 2004). Indeed, for effective marketing and selling practices to take place online, companies are compelled to identify and exploit existing motivators or create new motivators for transacting and interacting with the company. Therein lies the challenge, as understanding online consumer behavior requires understanding the shopping habits, the complex behaviors involved, and the pre-existing motivators they have built over time. (Kolesar and Galbraith, 2000)

1.1 Background of the Study

Recent years has seen an increasing trend in the importance of data analytics. According to Chen et al. (2012), business intelligence and analytics (BI&A) has its roots in the field of database management. There are many available technologies for database management, along with established pools of talent to handle the work. Hess (2010) identifies the most established player in the database management tools market as MySQL by Oracle, with its' history in database management tools dating back to 1979. Microsoft offers competition with its own line of enterprise database management software, going by the name Microsoft SQL Server. Another in the top three is IBM DB2, offering a cost effective alternative to the other two players.

Data analytics on its own is not a new thing, as data has been collected and recorded and utilized in some form or another for a very long time. However, technological advances in storing information have allowed for continuous collecting of data, whether it is relevant or not. Data storing capabilities are virtually limitless, and as such, the practice of storing all data has become common-place. The challenge that arises from storing absolutely all collectible data, is finding and structuring the data in a way to make the information useful and accessible. Big data attempts to build those connections within the network of information, so as to make it easier for the data to find its way to the user, as opposed to the user having to find a way to the data. (Ruh, 2012; Press, 2013)

A research conducted by the TechAmerica Foundation (2012) characterizes big data by three factors: volume, velocity and variety. A key characteristic of big data is that most of data stored in the world is highly unstructured. By estimation, only fifteen percent is easily accessible in relational databases of spreadsheets. The rest is unstructured in the form of emails, videos, blogs, social media and so forth. Additionally, the report notes that information-producing devices continue to multiply as more and more connected devices appear. Table 1 summarizes the characteristics of Big Data as per the findings in the report.

Table 1: Summary of Big Data Characteristics. (TechAmerica Foundation, 2012)

Characteristic	Description	Attribute	Driver
Volume	The sheer amount of data, which must be ingested, analyzed and managed.	Growing quantities of information: according to IDC Digital University Study, the world is producing information at an exponential rate, projecting 35 Zettabytes in 2020.	Increase in data sources, higher resolution sensors
Velocity	How fast data is being produced and changed, and the speed with which data must be received, understood and processed	• Accessibility: When, where and how the user wants information • Applicable: Relevant, valuable information for an enterprise at a torrential pace becomes a real-time phenomenon • Time value: real-time analysis yields improved data-driven decisions	• Increase in data sources • Improved thru-put connectivity • Enhanced computing power of data generating devices
Variety	New sources of information inflows both inside and outside of the enterprise/organization	• Structured – 15% of data • Unstructured – 85% of data • Semi-structured – The combination of structured and unstructured data is becoming increasingly important • Complexity – Where data sources are moving and residing	• Mobile • Social Media • Videos • Chat • Genomics • Sensors
Veracity	The quality and provenance of received data	The quality of Big Data may be good, bad, or undefined due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations	Data-based decisions require traceability and justification

Van der Zee and Scholten (2014) argue that as increasing amounts of data are collected as a result of technological developments, more and more people are connected to the internet. Bogue (2014) comments further, that the amount of devices capable of producing data is not the only determining factor, as without appropriate sensor technology with capabilities of extracting useful and valuable data, and abilities to process and analyze the data, the increasing amount of data will prove progressively less useful. Such an infrastructure of data collection, where people, things and data are all interconnected, is commonly referred to as the Internet of Things. Borgia (1, 2014) further defines the Internet of Things as an “emerging paradigm consisting of a continuum of uniquely addressable *things* communicating to one another to form a worldwide dynamic network.” The mentioned addressable things can be wearable technology, but for the most part will be embedded into infrastructure. The author presents an estimation of the number of such connected things exceeding seven trillion by 2025, which would be an average of 1000 devices per person.

It is important to remember that Big Data is, for the most part, raw data unsuitable for human “consumption”. As such, it must be filtered into understandable form. Bertolucci (2013) writes, based on the interview with data scientist Michael Wu, that data can be categorized into three rough categories: Descriptive, Predictive, and Prescriptive. (see Table 2) On a basic level, descriptive data can be considered to summarize what has happened, and as such is focused on the past. Descriptive analytics are the easiest to carry out, and as such form the vast majority of all analytics being done. Up to 80%, according to the expert. Predictive analytics gets more complex, utilizing a variety of statistical, modeling, data mining, and machine learning techniques to study and learn from data. As a result, educated predictions can be made of what will happen. Another way to think about it is taking data that is available to predict data that is not. A logical follow-up of predicting is to recommending and administering courses of action. Prescriptive analytics does just that, building on data available to predicting possible outcomes, and offering potential courses of action as a response to a scenario, along with likely outcomes of the decisions required.

Table 2: Categories of Analytics in Data Science. (Bertolucci, 2013)

Descriptive Analytics	Predictive Analytics	Prescriptive Analytics
• Describes what has happened • Forms up to 80% of analytics • Most simple form of analytics	• Utilizes statistical methods such as modeling, data mining, and machine learning to study data • Uses available current data to predict unavailable future data	• Follows predictions formed in previous stage to formulate possible outcomes • Offers courses of action to predicted scenarios • Provides likely outcomes of decisions

The practice of business intelligence is widespread in modern business; according to Bloomberg Businessweek (2011), 97 percent of companies generating revenues in excess of $100 million utilize business analytics. McKinsey Global Institute (Manyika et al., 2011) further states that the expanding usage of analytics in the business environment will cause an ever increasing shortage of qualified people with deep analytical skillsets. This in addition to the lack managers with appropriate capabilities of analyzing big data in their decision-making processes.

One example of such practices is clickstream analysis. Montgomery et al. (2004) define clickstream data as information providing details on the sequence of pages or the path viewed by users as they navigate a website. The authors specify that path data typically is expected to contain information on a user's goals, knowledge, and interests, which can be used to predict consumer behavior.

Clickstream Data in Consumer Behavior

By definition, clickstream data refers to electronic records of visitors' internet usage. The data is typically collected in the form of raw web server log data, but can also be collected in a more refined form through use of third-party services. The process of clickstream analysis involves analyzing the individual page requests made by Internet users, which are sent to the servers of the specific website. Each page request is generated by clicking on any embedded buttons or links on the website. These individual page requests form bigger picture of the users' behavior on the website, and strings of page requests are referred to as "paths", which reflect the actual choices and decisions made by the user on the website. (United States Patent Application Publication 2004; Bucklin & Sismeiro 2009)

Clickstream data can be roughly divided into two separate categories: site-centric data and user-centric data. As the names suggest, site-centric data focuses on all traffic on an individual website, while user-centric data collects browsing behavior of a single web user. Site-centric data allows for individual websites to begin profiling their users at a more detailed level than would otherwise be possible. Individual browsing behavior can be pinpointed through utilizing registration schemes, circumventing the classic delimitation of multiple users using a single device, or a single user utilizing multiple devices. Of course, this only works for users loyal enough to register and log-in actively when using the site. Site-centric data allows for website administrators to answer to much more specific questions. However, individual user behavior on a single website is of little use in building a bigger picture of online browsing habits. User-centric data, typically collected by the Internet Service Providers, allows for larger datasets to be collected on individual users. This allows for the possibility to build a more generally applicable profile on web users, regardless of channels used. However, this method suffers from the limitation of user/device mismatch to a greater extent than site-centric data, as there is no method of knowing who is using which device at a given time. (Bucklin et al., 2009) See Table 3 below for a summary of data categories.

Marketing research has benefitted greatly from the increasing availability of clickstream data. In particular, marketers are able to better observe and examine consumer search behavior on a large-scale setting. This has led to a variety of research, which models and tests theories of shopping behavior. Of particular interest have been interstore comparisons of a vast array of web stores, as well as intrastore behavior both on an individual visit basis, as well as on a page-by-page basis (the former examining visits and buying behavior over time, while the latter examines navigation during a single session within the store). (Moe 2009)

It has long been common place for organizations to rely on the Internet for conducting business, and much work has been done to adapt traditional strategies and techniques into the demands of the ever-more connected world of business. According to Berka (2007), organizations typically generate and collect vast quantities of data from their daily operations, of which the vast majority is automatically generated data (e.g. by

Web servers in the form of server access logs). Combining such raw data ideally leads to insightful information on things like the life-time value of customers, cross marketing strategies across products, as well as effectiveness of promotional campaigns. Further, being able to analyze the server access logs, for instance, allows for companies to better optimize their own websites based on passive customer feedback, in form of paths taken on the website. An example of this data being actively used is by companies selling advertising on the Internet, who can better target ads to specific user groups based on browsing behavior.

Table 3: Summary of Data Types. (Bucklin et al., 2009)

Data Category	Site-centric Data	User-centric Data
Description	data collected from a single website, representing the activities and behaviors of visitors on the website	data collected based on individual users; includes all browsing behavior on all websites; typically collected by Internet Service Provider (ISP)
Strengths	focused data mining; context of the website; user/device matching limitation can be overcome somewhat with registration to website	allows creating a profile of all internet usage across multiple channels
Delimitations	no data on browsing behavior in general; difficult to build a coherent profile of users	not possible to determine who is actively using the computer; an individual can use multiple devices, and many people can share a single device; specific activities within a website are difficult to track due to data filtering;

One purpose of collecting clickstream data is that it offers a unique way to examine consumer online behavior, as well as the effectiveness of marketing actions implemented online, due to its ability to provide information concerning the sequence of pages viewed and actions taken by consumers as they navigate a website. (Ståhlstedt, 2014) The sequence of viewed pages and actions taken are commonly referred to as “paths”, and the clickstream data collected provides valuable insight into how the Website is used by its users. The analysis of clickstream data is important so as to be able to improve the effectiveness of Web Marketing, as well as streamline online sales. (Lee et al., 2001) However, as Clark et al. (2006) note, clickstream data does not have the capability to the full story of recorded user activity, such as the pressing of navigational buttons such as “Back”. Neither does the data reveal the true intentions of the user on the website, or other possible activities that the user is engaged in during the use of the Website.

As per the findings in the literature review section, the conventional online marketing context finds that collecting and analyzing clickstream data helps increase understanding in the following areas:

- internet browsing and site usage behavior
- online channel efficiency as a medium of advertising, persuasion and engagement
- behavior of consumer purchasing and/or transactions on the internet
- identifying different buyer groups and user goals

(Moe 2003; Montgomery et al. 2004; Bucklin & Sismeiro 2009)

The power of clickstream data is evident in the opportunities it presents for improving targeted marketing efforts and sales activities (Montgomery et al. 2004; Bucklin & Sismeiro 2009). Just as with most data, clickstream data leaves much to the interpretation of researcher. However, the clickstream data collection provides a much wider variety of detail than a similar study done in a controlled environment would, and for the most part the data is unbiased and "natural. This allows for the analysis of not only purchasing decision-making, but also analysis of the paths which users take to arrive at the point-of-purchase. The key benefit of clickstream data, despite its delimitations, is that the data is typically collected in an interruption-free environment, removing the risk of interviewer bias, both in interpretation and in influencing the person being observed. (Bucklin & Sismeiro, 2009)

The collection of data and observation of user behavior allows for companies to take action to improve. Lee et al. (2001) suggest that analysis is meaningless without action. Collecting data, particularly in the case of e-commerce, allows the online store manager to observe and analyze many things based on the data collected. The author refers to a cyclical process of incremental improvement in the online store, which starts from collecting data (see Figure 1).

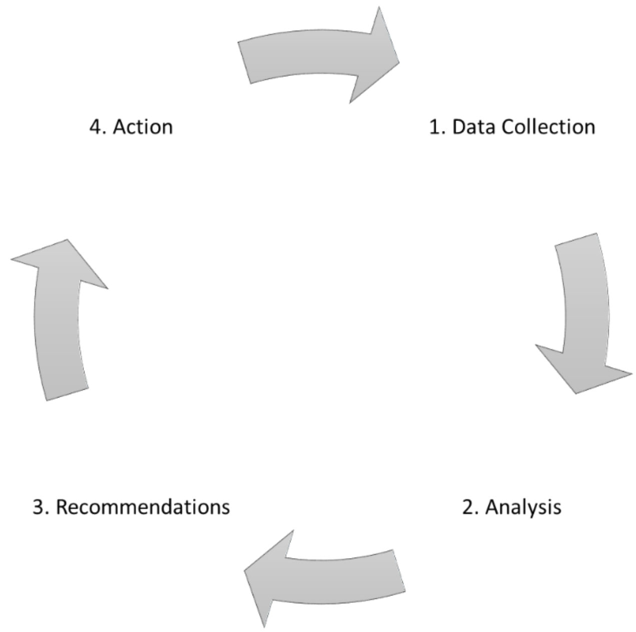

Figure 1: eCommerce KDD process, adapted from Lee et al. 2001.

The analysis of data then provides concrete recommendations for improvement, which the manager then can take action upon. Without the final step however, all the previous is meaningless. With action taken, the next iterative cycle can begin with data collection with consideration for the changes made

1.2 Focus of the Study

This research paper explores methods of incorporating data into understanding Web users. The outcome of the paper has been to gather and analyze data, from which the company can benefit in learning more about how their user base behaves on the website. The focus of the study is limited to the sphere of user behavior study, as opposed to a technical evaluation of big data. Big data and digitalization is rather used to build the conceptual link to the case company’s business. In addition, more in-depth insights may be acquired from the collected data through the use of more robust statistical programs in the analysis process. The theoretical focus herein was based on consumer behavior archetypes (Brown et al. 2003). Based on a categorization scheme derived from established behavioral archetypes, the study aims to distinguish between explorative and goal-oriented behaviors. Additionally, the study seeks to observe and understand behavior from a time perspective, mainly regarding the recurrence of different behavior in the online browsing context.

While academic research in the field of big data analytics is readily available in technically oriented articles and journals, the amount of research linking to fields of study in the business context are scarce. As such, the closely related field of business intelligence and analytics are used to develop empirical methodologies to analyze the data in the present study. Related prior studies by Ellonen, Kuivalainen, and Tarkiainen (2008; 2010) will be used as a conceptual basis for the analytical methodology used in the chosen user behavior framework of this study. In the end, for the purpose of this study, MS Excel is utilized to visualize user paths and analyze descriptive studies based on the clickstream data. The analytical process focuses on a combined methodology of cluster analysis and swim-lane diagrams.

A more advanced clickstream study would possibly include models for calculating probabilities of users ending up in desired areas, as presented by Montgomery et al. (2004) in their study of Barnes and Nobles customer behavior. Predictive analytics such as those are beyond the scope of this study, and would require a significantly better understanding of mathematical modelling.

1.3 Objectives of the Study

This study aims to build a taxonomy based on archetypes identified in traditional consumer behavior literature, and adapt it to a non-commercial online environment. The new taxonomy will thus consist of five distinct archetypes adapted to the online arena, with key characteristics identified for each. Cluster analysis methodology and swim-lane diagrams will be adapted to visualize the behavior of the newly identified archetypes.

This study contributes to the academic discussion around the impact of better understanding individual user behavior in the online arena. Based on existing research, it is clear that great potential exists in the studying of such behavior. As Moe (2003) points out, Hoffman & Novak (1996) originally proposed the concept of flow in the sphere of customer experience online, from which a further analytic model regarding number of pages visited and duration spent on each page has been derived by Bucklin & Sismeiro (2000).

The gap, as identified by Moe (2003), seems to be that the studies often do not include the content of pages visited. To remedy the situation, this study utilizes sequential analysis to categorize user groups and determine which users consume what type of content as per methods introduced by Bakeman & Quera (2011).

1.4 Research Questions

This thesis is an explorative user behavior study in an online non-commercial environment. The study utilizes click stream data generated within the case company website, which is used to build user paths. Based on the previously identified research gaps, the focus will be on the following research questions:

RQ1. How do Web users behave on an online magazine Website?

The main question driving this thesis is how users behave on the Website used in the case. To answer this question, two sub-questions are considered.

SQ1. What behavioral archetypes can be observed in the user base of the Website?

In understanding general behavior on the Website, it is important to be able to categorize types of users with certain parameters. Academic study has identified generally occurring behavioral archetypes in traditional consumer behavior settings. Utilizing the existing archetypes, this thesis builds behavioral archetypes for the online user environment.

SQ2. What commonly occurring “paths” can be identified from the data?

The collected clickstream data set allows for visualization of user paths taken on the Website. This paper analyzes the user data in order to find paths, linking the paths to the previously identified behavioral archetypes.

1.5 Key Definitions

1.5.1 Clickstream Data

An electronic record of internet usage collected by a website's web server log or by third-party services. Clickstream data can be defined as data collected on individual clicks that a visitor makes within a site. (Bucklin et al. 2009; Burby and Brown, 2007)

1.5.2 Customer Paths

The data generated through the data mining process can be processed to form certain visualizations of paths that the customers take within the website. The different segments of the website have been coded through observational methods relying on categorical measurement (Bakeman and Quera, 2011).

1.5.3 Visits/Session

Burby and Brown (2007) define a visit as an interaction by an individual resulting in requests for definable units of content. A session lasts a specified period of time, during which web activity must occur, else the session is considered terminated.

1.5.4 Web Traffic

Web traffic is a measure used to quantify the amount of visits and actions which occur on websites. Burby & Brown (2007)

1.5.5 Behavioral Archetype

Refers to a pattern of behavior that can be observed among consumers.

1.6 Research Methodology

This thesis utilizes theoretical frameworks from consumer behavior study and data science practices to observe, identify and form behavioral categories based on user behavior. Contributions from Brown et al. (2003) are used to formulate the basic behavioral archetypes from traditional consumer behavior study, which are then adapted to an online, non-commercial environment. Behavioral archetypes are further divided into overall categories based on generic browsing habits, as presented by Cutledge (1995). Furthermore, fundamental drivers of consumer behavior will be used to appropriately categorize the online user behavior based on collected data (Constantinides, 2004).

For the purpose of this case study, an exploratory research design was selected. Exploratory research is well suited for a click-stream data analysis study, due in part to the fact that prior consumer behavior oriented academic research in the field is scarce. The ideal outcome is to produce useful insights and familiarity into the topic, as well as generating ideas and assumptions which the company can incorporate into their future development of the website. (Labardee, 2014) To achieve this goal, the case study methodology was chosen for its suitability. According to Yin (2003), the case study methodology is ideal for research where the desired outcome is to understand phenomenon in their natural contexts. While it can be argued that case studies ultimately lead to papers with heavy degree of “story telling” elements (Dyer & Wilkins, 1991), it is none-the-less a powerful tool to emphasize the importance of understanding context through relatable examples and explanations through the real world case examples (Eriksson & Koistinen, 2005). Thus, as per the definitions found in literature, this study should be considered as an explorative single-case study (Yin, 2003).

Vartanian (2011) describes the difference between primary data and secondary data as primary data being collected by the same people who will use it, while secondary data is collected by someone other than the data user. This thesis uses extensive documentary secondary data in order to build a theoretical premise. Primary data on the other hand has been collected in the form of click-stream data from the case company website, using OpenTracker software.

Documentary secondary data further includes theoretical themes such as business intelligence and analytics, and online consumer behavior. Sources for publically available literature on big data have been obtained from the biggest players in the field, such as Oracle, Microsoft, and IBM. In addition, journal articles and relevant online materials were used to complement the data collection.

In addition to the theoretical framework of consumer behavior used as a basis for identifying user patterns, this study applies cluster analysis and swim-lane diagrams to understand and present the collected data. (Cerrato, 2015; Ketchen & Shook, 1996; Natschläger & Geist, 2013; Rummler & Brache, 1995)

1.7 Theoretical Framework

This thesis utilizes two primary theoretical fields: User behavior and path analysis. In the field of user behavior, archetypes of behavior are borrowed from consumer behavior study, and adapted to the online magazine Website case. Building on the identified typologies of behavior that the study aims at finding, the website content is divided into categorical content, to which the clickstream data is applied. As a result, certain paths users take on the website could be identified. Based on the findings, recommendations and user profiles were constructed. (see Figure 2)

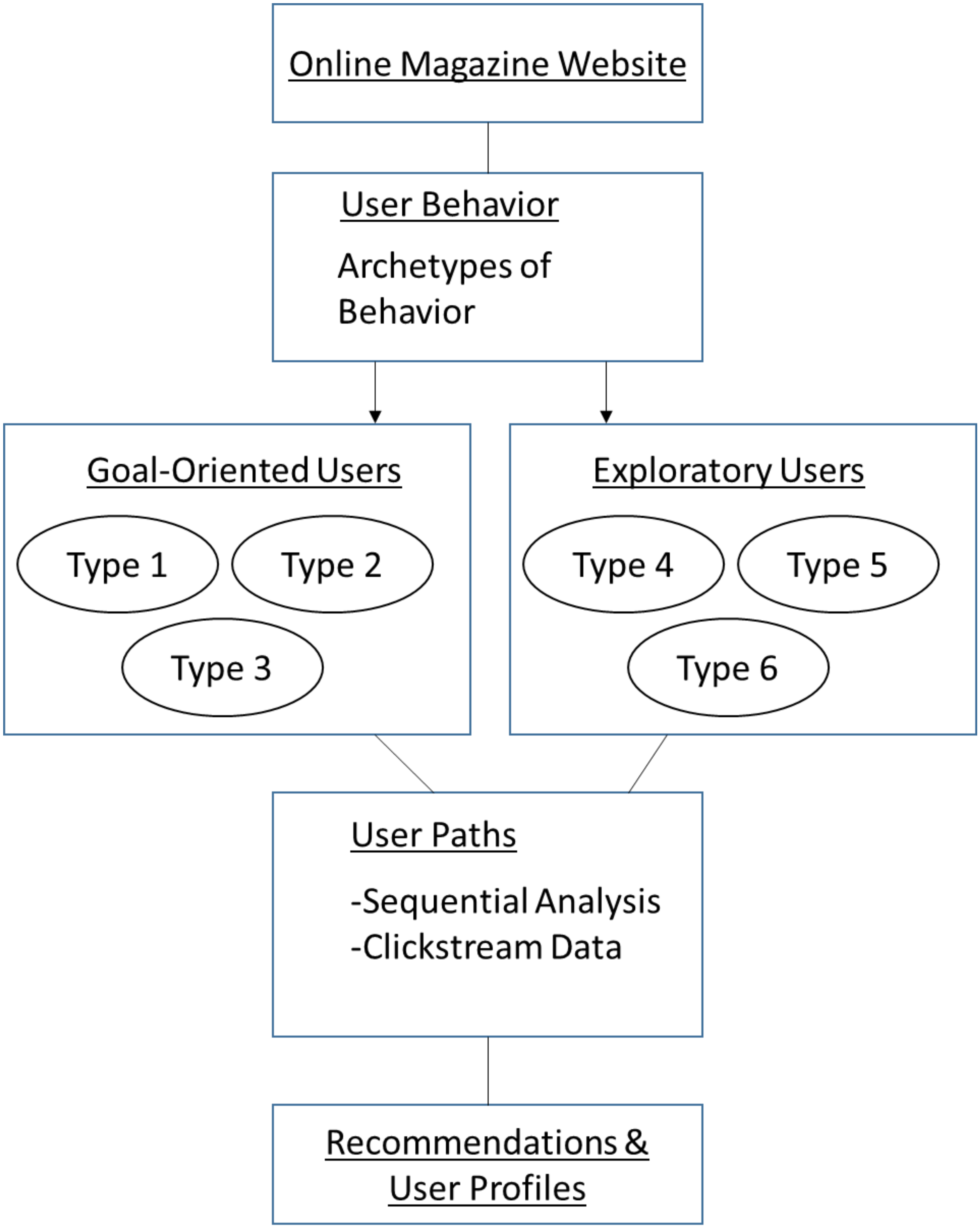

Figure 2: Research Framework.

1.8 Structure of the Thesis

The thesis is structured in the following manner. First the paper will explore the data-driven digital environment, and while considered to be a part of the theoretical portion of the paper, will focus on the current situation, rather than distinct theoretical frameworks. The objective of the chapter will be to familiarize the reader with the background of the digital trends, as well as establish a premise for why the topic is relevant.

The second chapter will introduce the main theoretical framework used in this study, which was selected to be online consumer behavior. The chapter aims to build a lens through which the empirical data can be observed and understood. The theoretical framework leans heavily on traditional consumer behavior in brick-and-mortar shopping environments.

Chapter three shifts the theoretical lens to be applied on the chosen case study, where consumer behavior principles are refined to work in a non-commercial online setting. The content of the chapter involves describing the case and data collection, introducing the methods of analysis used in understanding the data, as well as presenting the results of the case study.

Finally, chapters four and five discuss the relevance of the empirical results when considering the literature used. Additionally, the chapter considers the implications of the study from theoretical and managerial perspectives.

2 ONLINE CONSUMER BEHAVIOR

The key theoretical concepts in this study are based on previously published literature on big data and clickstream analysis. In academic literature, clickstream data analysis tends to focus on the user level, without much emphasis on behavior over time. In general, prior quantitative clickstream studies on the user flows have suffered from a lack of systematic or comprehensive nature, due to a lack of exact conceptual definition of flows. (Moe, 2003; Hoffman et al., 2000) The primary purpose of this chapter is to review current literature on consumer behavior, both online and offline, as well as user behavior in general. Having read this chapter, the reader is equipped with the understanding of how consumer behavior links with online user experience.

Although a unified model for consumer online behavior does not exist, the concept is far from being a novelty in academic research. Research on online consumer behavior has strongly focused on identifying and analyzing how decision-making and consumer behavior manifests in the online environment and which factors influence this behavior while consumers are searching, browsing, finding, selecting, evaluating and comparing information as well as interacting and transacting with websites related to their needs, interests and goals. See Table 4 below for a summary of the literature most influencing the consumer behavior theory of this study.

Table 4: Summary of Primary Literature Sources.

Author	Year	Title	Comments
Hoffman Novak	1996	Marketing in Hypermedia Computer-Mediated Environments	Exploratory vs. Goal-oriented behavior
Novak Hoffman Yung	2000	Measuring the Customer Experience in Online Environments	User experience in the online environment
Moe	2003	Buying, Searching, or Browsing	Browsing Behavior
Brown Pope Voges	2003	Buying or Browsing?	Shopping Orientations; Online Purchase Intention
Montgomery Shiboo Srinivasan Liechty	2004	Modeling Online Browsing and Path Analysis Using Clickstream Data	Path Analysis; Online Browsing; Clickstream Data
Constantinides	2004	Influencing the Online Consumer's Behavior	Consumer Behavior; Web experience
Comegys	2006	Longitudinal Comparison of Finnish and US Online Shopping Behavior Among	Information search; Shopping Behavior
Lee Chen	2010	The Impact of Flow on Online Consumer Behavior	Flow construct; Online Consumer Behavior

Traditional consumer behavior in the retail context is well studied and understood from both psychological and marketing perspectives. It is supported by theoretical literature wherein much of what can be observed in general consumer behavior transfers also to the online context. Koufaris (2002) implies that characteristics are clearly shared, with the distinction of online consumers having additional unique needs and concerns, which reflect their online environment. Hoffman et al. (1996) have proposed that Internet consumers should be considered as an entirely different consumer segment, with its own sub-segments of consumers. This is in stark contrast to the findings of Brown et al. (2003), who challenge this notion by presenting Internet retailing as a similar segment to other non-store retailing, and as such not warranting treatment as a special case.

There is further evidence suggesting the online consumer is generally more convenience-oriented, innovative, and variety seeking (Montgomery et al., 2004). The notion is supported by the thought that the additional dimension of information systems results in online consumer behavior requiring an equal understanding of information systems-related variables, as they are found to be of similar importance to the marketing focus (Lee & Chen, 2010).

2.1 Typologies of Consumer Behavior

Brown et al. (2003) synthesize in their study of consumer behavior existing literature from the field into six core archetypes of consumer. The authors describe the classic economic shopper as having qualities of a “price-bargain-conscious shopper”, “the special shopper”, “the price shopper”, and the “price conscious, value for money consumer”. As a unifying characteristic, these consumers are explained to be concerned primarily with buying at the lowest price or getting the most value for money spent.

The recreational shopper is derived from previously identified “active” and “involved” consumers, who generally enjoy the act of shopping, even if no purchase is made at the end. (Brown et al. 2003)

Consumers who avoid the shopping experience, or try to keep it to a bare minimum, have been referred to as “apathetic” or “inactive” shoppers. The archetype generally incorporates qualities such as convenience-orientedness, and tends to value their time highly. (Brown et al. 2003)

Lastly, the authors identify archetypes of the economic shopper and the personalizing shopper. The economic shopper exhibits characteristics of loyalty to stores or brands, or both. The personalizing shopper on the other hand refers to the consumers who value relations to store personnel above else.

The authors argue that the six archetypes defined above are different enough to avoid redundancy in characteristics, while providing enough diversity to study relevant niches of consumers.

2.2 Driving factors of Behavior

Much of consumer behavior, whether online or offline, can be traced to the fundamentals of customer behavior. Constantinides (2004) describes customer behavior at its most basic form as a process of learning, handling information, and decision-making. The activities within are categorized as follows:

1. Problem Identification
2. Information Search
3. Alternatives Evaluation
4. Purchasing Decision
5. Post-purchase Behavior

The primary distinction within customer behavior is between high and low involvement purchasing. As a rule, involvement is high when products are purchased for the first time, while involvement is low when purchases are done repeatedly. (Constantinides, 2004)

In online web browsing, Catledge (1995) identified early in the history of the Internet three dominant types of behavior on the Web, which still apply today. The first is the so called *search browsing*, which is browsing with a distinct purpose. Characteristic of search browsing is directed and specific search terms, and the goal is clearly in mind. The second behavior type is referred to as *general purpose browsing*, wherein an idea to the purpose of the browsing exists, but the path to the undefined goal is not specified. Typical of general purpose browsing is to consult several different sources, which are believed to hold a high likelihood of containing items of interest. Lastly, there is *serendipitous browsing* behavior, which is described as being truly random browsing behavior. The behavioral types are illustrated in Figure 3 below.

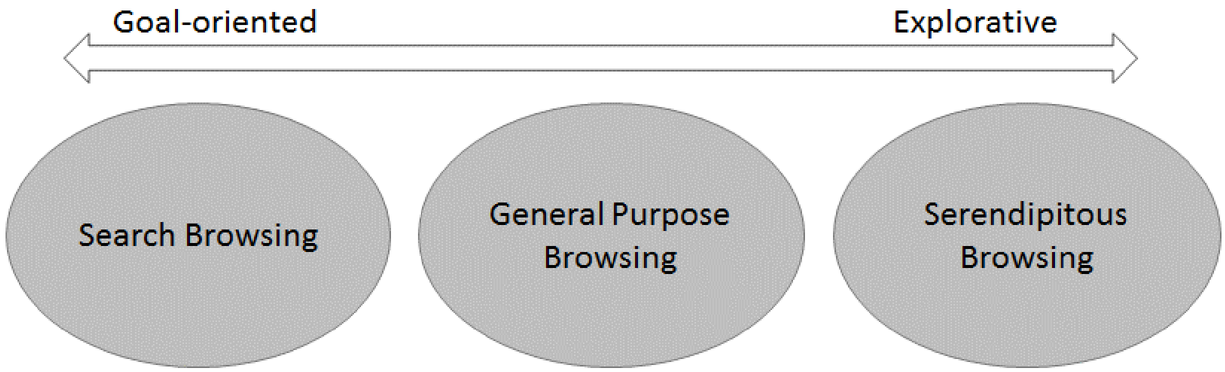

Figure 3: Types of Online Behavior. (Catledge, 1995)

A prominent issue for many companies is a fundamental lack in the basic understanding of how consumers behave. Being a complex and multi-faceted phenomenon, there is no single and all-inclusive model to explain all consumer behavior. Behavior can vary drastically based on the different personalities of customers, which becomes a pronounced problem particularly in international business. Furthermore, product or service characteristics and environmental factors add complexity to the overall consumer behavior. In addition, when considering behavior in the online environment, an entirely new set of rules need to be acknowledged and facilitated to what can be considered as the traditional and accepted patterns of consumer behavior. For instance, one of the classic consumer behavior models used is the Engel-Kollat-Blackwell framework of mapping the decision process stage. The authors originally split the decision process stage into the distinct phases of problem recognition, information search, evaluation of alternatives, purchase decision, and outcome. (Engel, Kollat & Blackwell, 1978) The model has since been adapted for the high-technology context, so as to better satisfy the more complex nature of online behavior, due to target audience being information systems users in addition to being consumers (Koufaris, 2002; Lee & Chen, 2010). Figure 4 below summarizes the key elements of the decision process stage model in the high-technology context.

Figure 4: Key Elements of Decision Process Stage. (Engel, Kollat & Blackwell, 1978)

Problem recognition

The fundamental driver of behavior in the decision process is the recognition of a problem: the desire to resolve the problem initiates the process (also referred to as *Need Recognition* in literature). By definition, this stage is characterized by becoming aware of the difference between one's actual state and the state which is desired. The recognition of a problem can be triggered either through an intrinsic factor (hunger, thirst, exhaustion), or an extrinsic factor (outside stimulus from advertisements, discussion with others). Furthermore, demographic factors such as age, education, marital status, etc. influence the problem recognition stage, as do various psychological factors. (Comegys et al. 2006)

In the field of consumer behavior study, the psychological factor of motivation is the most prevalent. The classic division of motivation splits into physiological needs and psychological needs, with the former encompassing e.g. food and shelter, while the latter is influenced by one's social surroundings. (Kinnear & Bernhardt, 1986) Furthermore, Dubois (2000) identifies differences in behavior based on the consumer's perception. The perception influences, directly or indirectly, how the consumer views themselves and their environment. In the online environment, keeping thresholds for entry and usage as low as possible are of paramount importance.

Information Search

Following the identification of a need or problem, the information search process can begin. According to Comegys et al. (2006), the information search part of the process is characterized by the utilization of multiple channels for gathering of information to resolve the problem.

The level of activity in this stage depends on the level of interest the consumer has. A milder state of arousal is identified as a level of "heightened attention", wherein the consumer is more attentive to relevant inflows of information, for example through advertisements and surrounding conversation. Conversely, in the high involvement *active information search* state, the consumer is proactively searching for more information and actively engaging in conversations of relevance. (Kotler & Keller, 2006)

Kotler & Keller (2006) note that information flows primarily from four distinct sources – personal sources through one's family and friends; commercial sources through advertising and sales people; public sources through mass media and consumer-rating organizations; experimental sources, including examining and using the product. While it is noted that the majority of information is attained through commercial sources, Dubois (2000) argues that the most valuable information is received through personal sources.

Information search further varies depending on the level of expertise of the one searching (Comegys et al., 2006). In the era of online search, level of expertise can be divided into expertise on the product and expertise on search. An expert in a particular topic or product can enjoy benefits in the form of internal search, meaning they are able to effectively utilize previously gathered knowledge, thus needing less new information on a given topic. Similarly, an expert on information search enjoys benefits in the form of more efficient searching habits, needing less time to gather a given amount of high quality and relevant information. (Alba & Hutchinson, 1987)

In addition, Urbany, Dickson and Wilkie (1989) identified an additional influencing element in the information search process in the perceived risk or uncertainty involved in a purchase. Perceived risk was further divided into two main categories: Knowledge uncertainty and choice uncertainty. While knowledge uncertainty involves uncertainty regarding consumer's inherent information regarding the alternatives of what they wish to purchase, choice uncertainty refers to the actual uncertainty regarding which of many alternatives should be selected. What Urban et al. (1989) and other literature agree on is that some level of knowledge on the product or service is required in order for information search to increase significantly, while no prior knowledge or comprehensive knowledge may decrease the amount of information search a consumer is willing to do. The authors suggest that choice uncertainty will generally increase information search behavior, while knowledge uncertainty tends to decrease it.

Finally, in the case of existing prior knowledge, Narayana and Marking (1975) present a conceptualization where all existing knowledge on alternatives exists in an awareness set. The awareness set is further divided into subsets. Brands which the consumer associates with positive experience or knowledge belong to the evoked set, from which the consumer is likely to make the final purchasing selection. Brands or products which the consumer has neither positive or negative feelings of belong to the inert set, from which it is still possible for the consumer to make the purchase selection, though it may require some further information search. Lastly, products which the consumer has negative connotations of belong in the inept set, from which it is very unlikely for the consumer to make a purchase decision from, if no corrective measures appear from the side of the brand.

Evaluation of Alternatives

Once enough information has been gathered on different alternatives to satisfy a given need, it stands to reason that next the consumer must narrow down the pool of possible choices to the most viable ones. For this, Comegys et al. (2006) suggest that the consumer forms a set of minimum acceptable levels for the traits they wish, and exclude the alternatives that don't match these levels. Interestingly, in the online context, price rarely seems to be a primary determinant in the evaluation of alternatives (Bhatnagar & Ghose, 2004). The authors hypothesize that the consumers may expect prices online to be more or less similar to each other, and thus not require much explicit attention.

Comegys et al. (2006) establish that consumers tend to invest their resources (e.g. time) in evaluating alternatives up to the point where acquiring new information causes more trouble than what the new information is worth. Generally, when the consumer feels additional research to no be worth the effort, the search process ends. The value of time is not the only determinant in the evaluation process, however. The authors suggest that equally important is the ease of information processing regarding a brand, where two concepts are presented. The first, where a brand is easily recognizable due to its physical characteristics, is referred to as perceptually fluent, while a brand or product that easily comes to mind when a problem is identified, is referred to as conceptually fluent. The authors note that conceptual fluency is improved if brand related products can be observed in the same context. Lee and Labroo (2006) further define perceptual fluency as a data-driven, bottom-up form of processing that is not particularly sensitive to elaboration. Conceptual fluency on the other hand is elaborated as being a conceptual, top-down form of processing, which benefits greatly from elaboration. In their study, the authors found both forms of processing fluency to enhance product evaluation and brand choice. Lee and Labroo (2004) further note, that conceptual fluency can also be adversely impacted by negative connotations of relevant stimuli. For instance, the authors point out that if a ketchup billboard is reinforced with a preceding billboard advertising french fries, conceptual fluency can be improved due to relevant product categories. However, consumers who feel negatively about the health or taste aspects of French fries may also carry the negative sentiment to the ketchup ad, though otherwise they would not feel badly about the product.

The concept of perceptual and conceptual fluency is easier to keep in control in the traditional offline marketing space, but in online marketing further challenges are introduced, with the online arena being saturated with different advertising stimuli. Conceptual fluency is, however, possible to utilize with improving technologies for ad placement customization.

Purchase Decision

As per the initial framework, following the evaluation of alternatives comes the final purchase decision. However, Comegys et al. (2006) note that there are two additional factors which may influence the choice beyond what the consumer personally feels would be the best alternative (see Figure 5). The first factor is the influence of peers, namely the buyer's family and friends. While initially a certain product may be deemed as the best for the need, the lack of acceptance from people whose opinion the buyer respects can greatly influence the choice. Secondly, there may arise unexpected situational factors, such as the product being sold out, or the price increasing considerably. In the online context, the attitudes of others have less of an impact due to the generally private environment of the point of purchase.

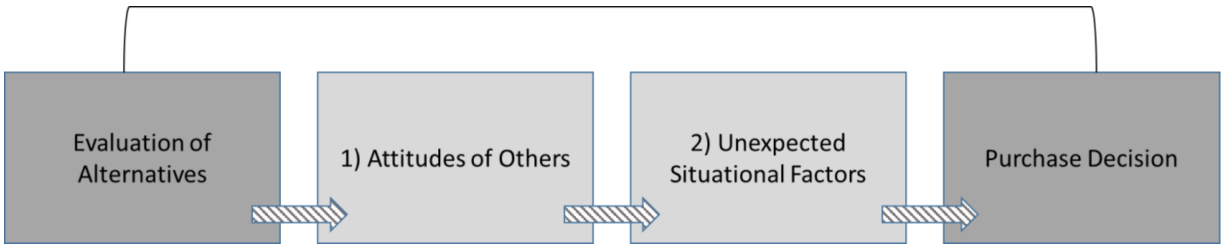

Figure 5: Factors Influencing Purchase Decision

Arriving at the final purchase decision typically requires the consumer to make several sub-decisions prior to purchase. Comegys et al. (2006) summarize these sub-decisions as: price range, point of sale, time of purchase, and method of payment. However, there are times when prevalent consumer behavior theory fails to apply. The authors refer to a phenomenon known as “impulse buying”, where the purchase decision is made purely based on the impulses and emotions of the buyer. Baumeister (2002) suggests that certain irresistible impulses are generated by physiological needs, and while they cannot be ignored indefinitely, do not always lead to purchases. The author expands that consumers have three primary characteristics which drive the individual’s self-control, which in turn helps them as a consumer to avoid unplanned and unbudgeted purchases. Failing in even one, on the other hand, can result in impulse purchasing, if other circumstances are fitting to make the purchase. The first characteristic the author identifies are the standards of the individual, meaning that if the person has clearly set goals and norms and they know what they want, then they are not as prone to making purchases at sudden impulse. Strong standards also make the consumer less vulnerable to pushy sales personnel and advertisements. The second characteristic is referred to as monitoring, where it is suggested that people who track their own relevant behavior closer, are less prone to losing control of themselves. Finally, the consumer’s capacity to change – even when failing to retain self-control through their own standards and monitoring of self, the consumer must still have an inherent willingness to make a change and by going along with the impulse purchase. See Figure 6 below for a summary of the characteristics.

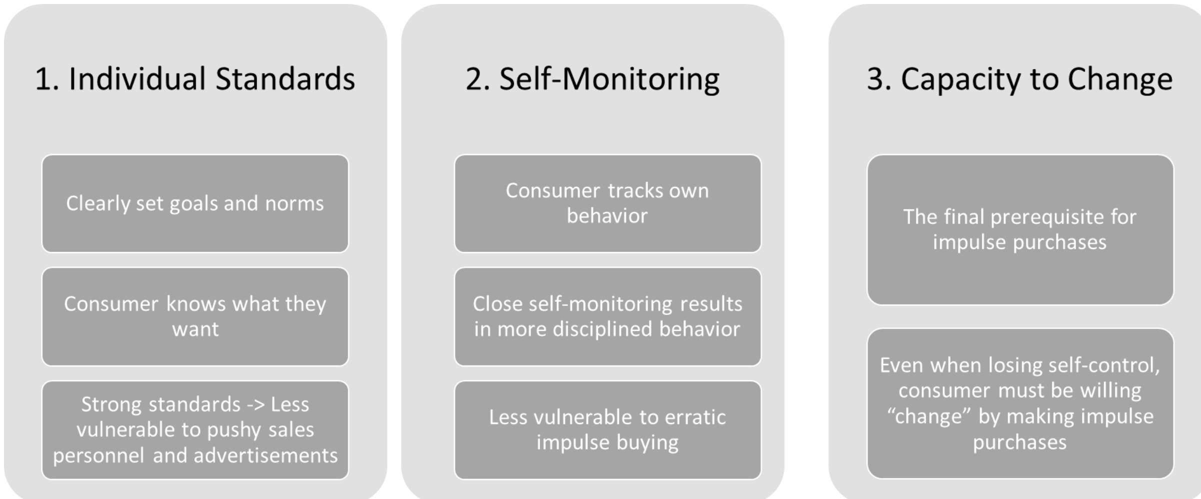

Figure 6: Summary of characteristics of self-control (Beumeister, R. 2002).

Outcome

Also referred to as the post-purchase behavior, the final step of the purchasing process involves everything after the actual purchase is made. Current consumer research divides the post-purchase behavior into *satisfaction* and *action* categories. Post-purchase satisfaction is believed to withhold different thresholds, which influence the customer loyalty following a purchase. The concepts of customer loyalty (action) and customer satisfaction are used frequently in literature, and the two are closely related. IT is important to note the difference, however: while loyal customers tend to be satisfied, satisfaction does not necessarily produce loyalty. (Comegys et al., 2006; Mittal & Kamakura, 2001; Oliver, 1999) Herein some authors disagree, stating that their research indicates that satisfaction does indeed directly produce loyalty (Auh & Johnson, 2005; Ball et al. 2004).

Satisfaction is important in both offline and online environments, as well as in situations where a product is being sold or when a service is being offered. The factor which is considered important above all else in the online environment is convenience. Convenience is typically a strength in online commerce, as information on product variety are more readily available to the consumer. Furthermore, in addition to satisfaction and loyalty, brand preference and repurchase intent play crucial roles. Prevailing literature has hypothesized there to be a multi-tiered level of interaction between the different factors. In summary, it is believed that satisfaction can positively impact customer loyalty, which in turn may positively influence brand preference, which finally impacts the consumer's intent of repurchase. See Figure 7 below for a graphical representation.

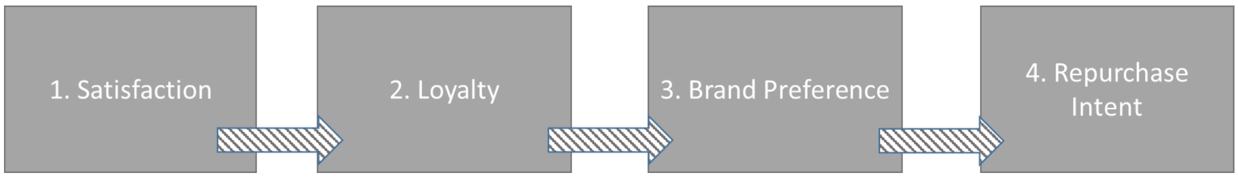

Figure 7: Hypothesized influence of factors in post-purchase outcome. (Hellier et al., 2003)

2.3 Exploratory vs. Goal Oriented Behavior

Academic study of online consumer behavior has identified two primary categorical behaviors – goal oriented and exploratory behavior. Particularly fruitful has been research leading to distinction between the behaviors in the Web context, where the exploratory process is considered to be of greater importance for many individuals than the final result. In user behavior studies, "flow" is often measured to quantify experience. However, due to the broad and general nature of "flow", it has been difficult to incorporate into more precise investigations of flow in comparative investigations between goal oriented and exploratory behavior. (Hoffman & Novak, 1996; Novak et al. 2003)

In a marketing context, the difference between goal oriented and exploratory behavior has been formally studied and documented, as it directly involves in the majority of the purchase and consumption process. Focal to the distinction is the comparison of extrinsic and intrinsic motivation and the level of involvement (situational vs. enduring). (Novak et al. 2003)

Furthermore, when considering the consumer search process, it is evident that behavior is primarily directed or non-directed. The process leading to the choice process on the other hand, can be identified as goal-directed or more oriented towards exploring different choices. (Hoffman & Novak. 1996; Novak et al., 2003) Finally, consumers have attitudes exhibiting both hedonic and utilitarian components (Novak et al., 2003).

Lastly, the core shopping process has been well analyzed in both traditional retail scenarios as well as online store settings, which has led to a general distinction between goal oriented and exploratory behavior being characterized as either “work” or “play” type behavior. (Novak et al., 2003)

The division of goal-oriented and exploratory behavior is supported by the research of Brown et al. (2003), who argues that the general predisposition to shopping manifests in varying patterns of information search, alternative evaluation, and product selection. While focusing on the earlier research in retail shopping, much of the behavior translates to online behavior, as noted by the authors. Brown et al. (2003) synthesize existing topical literature into six primary behavioral orientations for the conventional shopper, based on the most frequently appearing typologies. The typologies are as follows: 1) Economic Shopper – concerned with buying products at the lowest price or getting the best value for the money they spend; 2) Recreational Shopper – enjoys shopping as a pass-time, and does not need to make a purchase to enjoy the experience; 3) Apathetic Shopper – the so-called “inactive shopper” does not enjoy shopping, and does so only out of need; 4) Convenience-oriented Shopper – conceptualized as a time-/space-/effort-oriented construct. Shoppers may be motivated by one or all dimensions; 5) Ethical Shopper – known for store and/or brand loyalty; 6) Personalizing Shopper – value inter-personal relations with store personnel.

It is important to note that the division between goal-oriented and exploratory behaviors may occur due to a fundamental and intrinsic characteristic of the browser. That is to say, for example, does the user enjoy online browsing in general, or is the Internet primarily used for information gathering? Bellenger & Korgaonkar (1980) noted that when considering the orientation of people toward shopping, one should consider what their view on alternative uses and expenditures of time are. If the act of shopping is found enjoyable, customers are likely to spend time shopping even if there is no purchase, and not be frustrated in doing so. Conversely, others might wish to get through the “ordeal” as quickly as possible, so that they may direct their time and effort towards a more enjoyable activity. This is equally true for the online arena, where some are certainly more recreational in browsing habits than others, and it is fundamental to understand how to design websites to attract and satisfy both types. Just as with impulse purchasing accounting for considerable portions of sales in many product categories, it can be argued that impulse browsing may eventually drive traffic to other areas of the website, so long as the core need is catered to. (Bellenger & Korgaonkar, 1980; Brown et al., 2003)

3 USER BEHAVIOR OF ONLINE MAGAZINE READERS

The traditional printed media industry has long been under tremendous pressure to change with the digitalization trend picking up speed. The importance of understanding how readers, both new and old, behave online is crucial for the future success of publishers. Ellonen and Kuivalainen (2008) identify five primary objectives for magazine publishers, based on a 2005 study by FIPP including 71 magazine publishers:

1. to expand the readership beyond the print audience by creating an online audience;
2. to attract new readers of the print magazines;
3. to create revenue streams and profits in the long term;
4. to build a community around the magazine brand; and
5. to have the means of communicating with the target audience on a more frequent basis.

Based on the reported objectives, it becomes clear that online presence is a central issue for many publishers. In order to expand and create readership in online formats, it becomes increasingly important to understand the content preferences of users, and indeed, which types of users frequent the website.

This chapter describes in detail the case used in this thesis to examine online user behavior in the online magazine context.

3.1 Case Description

The case company, Aller Media, has commissioned the thesis to be done on one of their websites, Costume.fi. Costume is a women's fashion magazine concept adapted from it's original in the Netherlands. Now Aller Media is relaunching the website, as well as removing from the market the print version of the magazine, leaving only the web version. Aller media is interested in developing the business of the online magazine implementation, and as such wants to collect and analyze the clickstream data generated on the new version of the website, so as to better understand their online userbase. The old website version had some 90.000 unique weekly visitors, 235.000 visits per week, and 400.000 page downloads per week.

Aller Media Oy is a media brand focusing on magazine periodicals aimed at women, with a total reach of some 300.000-400.000 readers weekly (Aller.fi, 2014). Among the most circulated magazines are Olivia, Elle and Costume. In spring 2014, the combined weekly reach of the three magazines was approximately 210.000 people in the category of ages 12+. Recently the printed version of Costume was discontinued, with all content moved to the online website. In Q4 of 2014, the website attracted 90.000 unique visitors per week, with 235.000 visits and 400.000 page requests, with a median session length was 00:01:21. Majority of the traffic focused on the different blog content. (Aller, 2014)

The motives for Aller Media to understand their online user base better is clear, with the discontinuation of the Costume print magazine circulation. Ideally the Costume editorial team can develop the online content to be comprehensive enough to retain their former print readers, who would like to continue reading the magazine material. Furthermore, a wholly-online presence gives the magazine a unique opportunity to engage with their readership on a daily basis, which complements and enriches the magazine. While at best this allows for the magazine to clearly and continuously communicate the values of the magazine to their readership, it does require constant updating and improvement. (Ellonen & Kuivalainen, 2008)

In the scope of this study, data collection and analytic methodologies are implemented in a small-scale online consumer behavior study. Clickstream data is collected from the website of a prominent Finnish media house. Data is to be collected over a period of six weeks using Open Tracker service. The data is gathered from a brand new website implementation, from day one of launch.

.

3.2 Data Collection

To answer the research questions, this study employs an explorative and data-driven approach in categorizing the consumer behavior pattern. The empirical study herein applies quantitative methods, so as to produce generally applicable categories of website users. Categories of consumer behavior patterns have been identified based on a data-set of clickstream data, gathered from the Web site over a period of six weeks. More specifically, the methods of sequential analysis, cluster analysis, and swim-lane models have been applied so as to extract user behavior patterns from the raw data. The study fits the description of a single-case narrative, which represents the classic case study methodology as defined by Yin (2009), and is used to describe and analyze data. According to Yin, the most important questions that case studies often strive to answer are the “how” and the “why” questions. The use of case study methodology can be justified due to its suitability to analyzing and examining contemporary sets of events, where the researcher has little-to-no control. (Yin, 2009)

The research consists of a six-week long period of clickstream data gathering via the use of OpenTracker software, focused on the new launch of the Costume website. The timespan of the data gathering was between March 16th and April 28th. To ensure a statistically valid sample of users, a random sampling methodology was utilized for selecting the users whose data would be used for the analysis. According to Easton & McColl (1997), random sampling is “a technique where a group of subjects is selected for study from a larger group (a population). Each individual is chosen entirely by chance and each member of the population has a known, but possibly non-equal, chance of being included in the sample.” The sampling method utilized in this thesis is slightly modified from a pure random sampling methodology. For the purpose of this study, the overall population was first limited by randomly selecting a day from within the duration of the study where data has been collected, and only then random sampling methodology was applied to select the individual users. However, the potential population was further limited by including only users with a minimum of five visits during the data collection period. While this may introduce a slight bias to the study for not including less frequently visiting users, it was determined as necessary for the scope of the project. The desire of the commissioning company was to understand behavior of somewhat active users, and as such this qualification was introduced.

The visitors were randomly selected from the dataset of user activity on April 14th. See Table 5 for summary of data collection methodology.

Table 5: ***Data Collection Method***

Data Collection Summary		
Unique visitors click-stream data for the new Website Data collected from Day 1 of Website launch	For the purpose of this study, a period of six weeks (March 16th to April 28th) was selected as the time frame.	Randomly selected from unique visistors on specific day (April 14th) Only users with x>4 visits were considered

The goal of this experiment was to collect data from users who are browsing completely independent and unaware of the experiment. Following the identification of the users to be included in the study, the data was collected based on the selected user IP addresses from the open-tracker service. All IP addresses were coded to user numbers after downloading the data to protect the privacy of selected individuals. The collected raw data was then exported into MS Excel for analysis, which was conducted utilizing the theoretical foundations of sequential analysis and clustering methodology.

As mentioned, the form of data collection applied in this study is site-centric collection of click-stream data utilizing OpenTracker online software. The case company has implemented tags by which data is collected from the website users, and the data will then be downloaded and analyzed by the researcher. The collected data contains raw information on which website elements the visitor clicks, how long they spend on the site, whether they return or not, and where they came to the site from. The raw data, collected over a period of six weeks, has been analyzed using Microsoft Excel. The raw data was collected in the form of access log files, wherein each individual log input illustrates a request from a client machine to the server. (Marascu & Masseglia 2006)

The dataset consists of all page requests from 250 randomly selected unique visitors, from a total population of 1.300 users. The result is a total of approximately 12.000 data points, which was then filtered down further based on the self-imposed rule that a single browsing session would be considered as 20 minutes in length. As such, a session with no activity within the frame of 20 minutes is considered an ended session. As per the rule, the 12.000 data points translate to a total of 8.150 unique sessions, of which a total of 76% were sessions with only a single page request. Finally, for the purpose of this report, the analyzed multi-click sessions amount to roughly 2.000 sessions. See Table 6 below for a summary of data collection results.

Table 6: Summary of Data Collection Results.

Total Population	Sample Size	Data Points	Sessions (>=20 mins.)	Single page request	Multi-click Sessions
1.300	250	12.000	8.150	6.150 (76%)	2.000 (24%)

This data provides a comprehensive view into the behavior of individual users of the website. By seeing a visual representation of the paths taken by users, we are able to ascertain the categories of content which drive traffic on the website, as well as which paths lead to desired outcomes. Through an increased understanding of which types of content lead users to desired content, it becomes increasingly possible for the publisher to include content to the site which lead to actual desired business objectives.

3.3 Methodology and Analysis

Following the collection of the data, the log files were manually filtered to include only the information interesting to the study. To answer the question of which paths customers take, the most important data are the IP Address for separating the different users, the timestamp of the last event for the limiting of the sessions, the duration of the visit, and finally the raw URL of the click actions. The IP Addresses were coded to exclude any possibility of tracing users to real people, while raw URLs were coded categorically based on site content. Out of the total 1.300 users included in the data set, 250 unique visitors were considered in the study, which resulted in a total of 12.000 data points. It was decided that a single browsing session would be considered 20 minutes in length. As such, any session with no activity during 20 minutes would be considered as ended. In total, the data set covers approximately 8.150 unique sessions, of which approximately 76% were sessions with only a single page request. Thus, for the purpose of this study, the analyzed multi-click sessions amount to roughly 2.000 sessions.

Table 7: Coding of Categorical URL Content.

URL Category	Content Included in Category	Code
Navigational Pages	Front page, blogs front page	N
Editorial Content	Fashion/Beauty/Lifestyle Articles	E
Blogs Hosted by the Magazine	All the URLs of the four blogs	B
Alternative Blog	External links to other blogs non-hosted blogs	T
Video Blog	URLs to video blogs	V
Stores	Links to commercial content on the website or on the blogs	S
Ads	Links to advertisements/sponsored competitions	A
Social Media	Links to/from Facebook, Twitter, Instagram, online community	M
Comments	Links to/from comment sections	C

The main analysis of the study involves identifying behavioral patterns by categorizing clicks based on the website content, and then arranging the clicks into sequences. This study has two primary areas of interest from the analytical perspective: the first is the transition between different URL categories, and the second is the time perspective –

whether identified user behavior is recurring or not. Table 7 above presents the URL categories according to their content. User paths in the data can be compactly represented using a selected initial from our chosen categories. In the search for clickstream patterns, the focus was narrowed to focus on transitions (e.g. Bakeman & Quera, 2011; Montgomery et al. 2002; Ellonen et al., 2015). As such, sequences are shortened by removing multiple consecutive categorical clicks of the same type. For example, if a sequence comprises of clicks “N, B, B, B, E, T, T”, the final sequence to be considered in the scope of this study would be “N, B, E, T”. See Figure 8 for categorization scheme of website content.

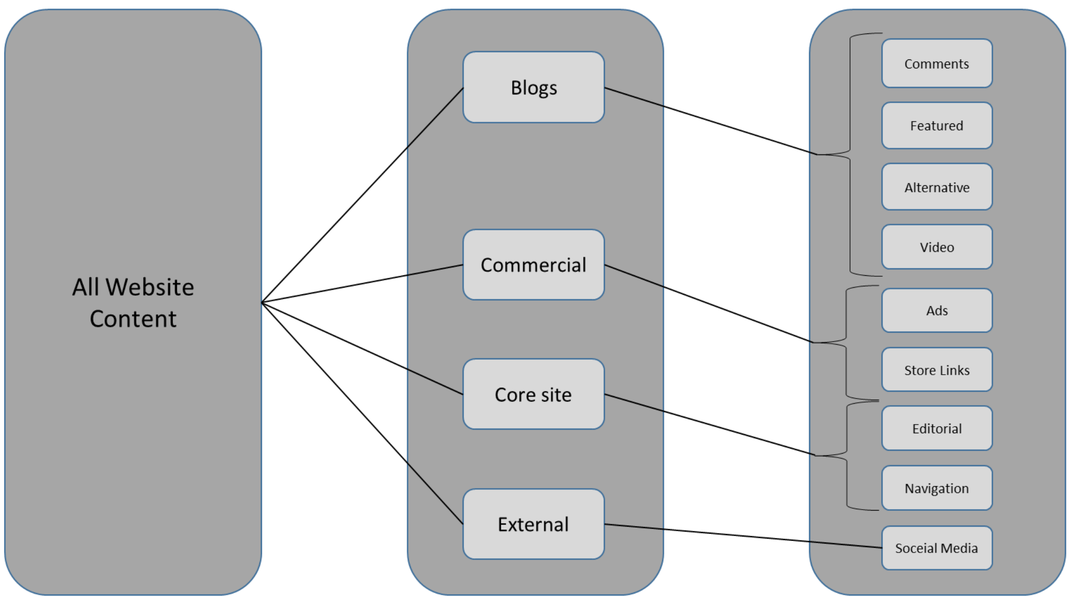

Figure 8: Categorization Scheme for Website Content.

The entire website content was divided into four main categories of content, based on the categorization scheme of Montgomery et al. (2004): 1) Blog(s) content, 2) Commercial content, 3) Core site content, and 4) External content.

The first category includes click-activity in the featured blogs (e.g. Xenia’s Day, Mariannan, Nude, etc.), clicks in the comments of blog posts, clicks to alternative blogs outside the costume.fi site (alternative blogs include blogs kept by the featured bloggers elsewhere, as well as completely unrelated blogs to Costume.fi), as well as the occasional video blog post.

The second category included the ad clicks and store links followed by users. Both are often included in the blog section, which seems to drive the majority of traffic on the website, and thus ad and store link clicks are typically preceded by activity in the overall “Blogs” category.

The third category is the “Core Site” content, including editorial content and navigation. It is important to note that it is the desire of the company to drive activity towards editorial content, and as such already several website sections have been added since the conducting of this particular experiment. In essence, the editorial content is the content which is in control of Aller Media Oy, and as such is the content which can best be influenced to achieving desired business outcomes.

The fourth category was determined to be “External” content, and includes otherwise unsuited content of social media clicks. Please refer to Table 8 below for a breakdown of the shares of total clicks by web content category.

Table 8: Web Content Categories by Size.

Content Category	Share of Total Clicks	Share of Total Clicks, Excluding Navigation	Included Click Categories
Blog Content	20,7 %	80,0 %	B, T, V, C
Commercial Content	2,6 %	10,1 %	S, A
Core Site Content	75,8 %	6,1 %	(N), E
External Content	1,0 %	3,7 %	M

Finally, to complement the individual visitor data that was used as a method to develop paths, the dataset was also used to determine more general observations, such as which pages people typically enter and exit the website, where they were referred from, and which platforms are commonly used to access the website. (See Chapter 3.4.1)

3.4 Results

3.4.1 User Background

For the purpose of this study, it is presumed that the users of the website correspond roughly with the readership profile of the print magazine. Due to the problematic nature of identifying and collecting data on individual users accurately, some overall demographic data from the readership of the print magazine will be used to build a general user profile. The majority of the readers fall within the age category of 15-29 years old, accounting for a combined 61% of total readership. Further, the readers are heavily concentrated in major cities, as over 60% of readers reside in cities with a population of more than 70.000 inhabitants.

An analysis of the users by platform used to access the website showed a fairly even distribution between mobile and desktop operating systems. A generic assumption for a rough difference between mobile and desktop browsing is that on a PC, browsing exploratory browsing can be more easily carried out, while due to delimitations of mobile browsing, it tends to be more goal-oriented. Figure 9 below illustrates the different platform shares of the sample, with roughly 40% being desktop based users, and 60% mobile users. Windows based platforms are dominant on the desktop category, while iOS is the preferred platform for mobile.

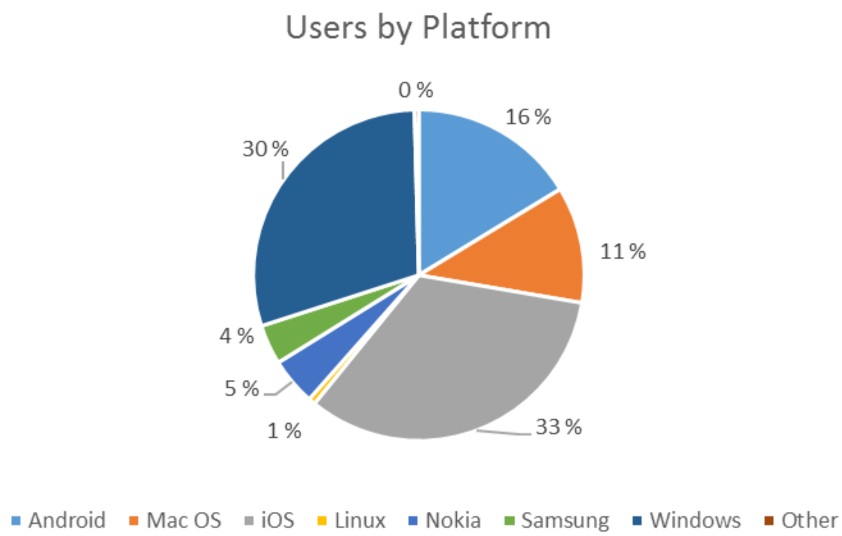

Figure 9: Users by platform.

While the usefulness of knowing the platform can be speculative in understanding behavior, the information can prove useful in website optimization and development. However, the information may also prove insightful when considering certain types of user behavior. For instance, behavioral patterns where banner ads are actively clicked may be explained by the use of mobile platforms, where it is easier to accidentally click on ads, particularly if they load slower than the rest of the site content.

3.4.2 Categories of Site Content

This study utilizes established theoretical frameworks from traditional consumer behavior study to build a premise for assumed user categories. Based on identified behavior of consumers in a traditional setting, this paper will aim to link similar behavior from an online non-commercial setting to the established categories. While the environment is certainly different in many ways from a traditional consumer setting, it can be argued that in terms of user behavior, similarities are sufficient to draw some broad conclusions. From existing literature, five distinct categories of consumers have been identified based on their behavior. As such, five categories of browsers were identified based on their generally similar behavior. Browsing behavior was observed through sequences of categorical clicks. This chapter will discuss the general characteristics of each category, as well as how common they were in the sample population, and whether the behavior was observed to repeat over time.

In terms of content consumed on the website, it was apparent that the preferred content was the blog content of the featured blogs. When observing the overall clicks that were made during the time period of the data collection, approximately 80% took place in blog related content (when omitting basic navigation pages). This situation makes it challenging to identify directly comparable user groups, as the frequency of observed behavior is clearly more dominant for certain user groups than others. (See Figure 10)

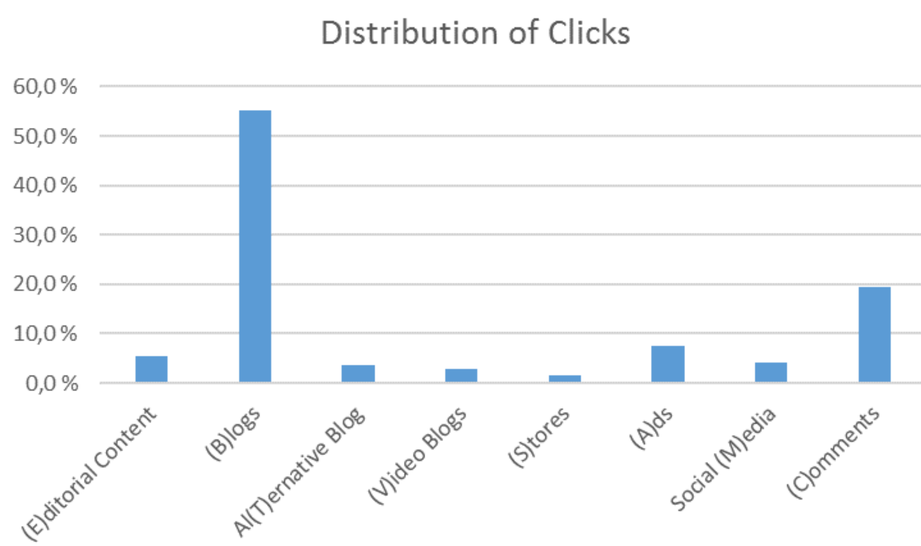

Figure 10: Distribution of Click. (Navigation clicks omitted)

The following section will explore the different identified user types, with descriptions of each group, as well as their size relative to total sessions in the sample.

3.4.3 Editorial Content

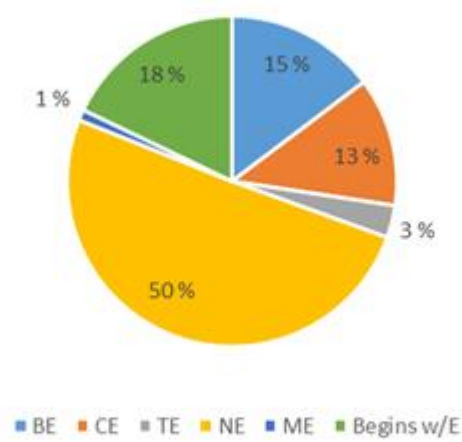

Figure 11: Clicks Preceding Editorial Content

From the perspective of the company, it is important to understand which type of behavior leads to the editorial content created by the publisher, and thus is in the power of the company to influence directly. As can be expected, half of the paths leading to editorial content are preceded by activity in the “Navigation” pages. This should be considered the organic path to editorial content, where the user enters the site through the landing page, and proceeds to navigate to content interesting to the user. The second largest categorical group is one where the user path begins with editorial content. This particular path suggests a higher degree of goal-oriented user behavior, where the reader is finding their way to the content through specific search terms. The remaining significant precedents of editorial content are blog content and commercial content. (see Figure 11)

In addition to understanding which behavior is important to arrive at editorial content, it is equally important to understand where the paths continue immediately after editorial content. Up to 53% of paths end after editorial content, reinforcing the assumption of editorial content being integral part of goal-oriented browsing behavior. A majority of remaining paths continue to either navigational (32%) or featured blog content (8%). (See Figure 12)

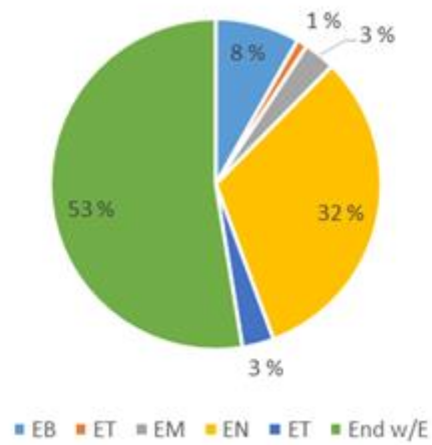

Figure 12: Clicks Following Editorial Content.

3.4.4 Referrals, Entry, and Exit links

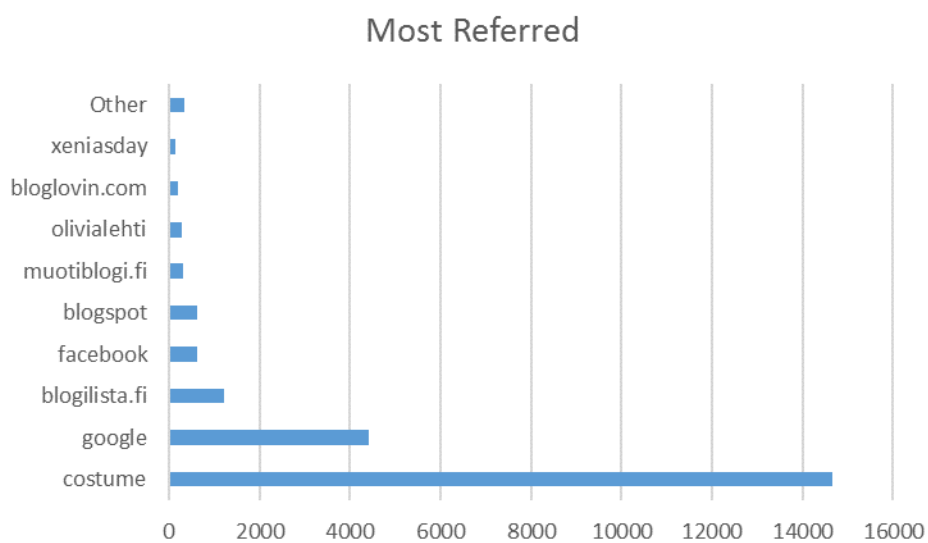

Figure 13: Site Referrals.

Modern web browsers all record certain information regarding browsing behavior. An example of such information are the referrer pages (i.e. what is the last page a user viewed before entering the website) and the entry and exit pages (which provide information on which part of the website the user landed first in the session, and what is the last page viewed before leaving, respectively). (Opentracker.net, 2015; searchengineland.com, 2015) As can be seen from Figure 13 above, the majority of clicks are preceded by content on the actual website, which can be interpreted as a positive sign of traffic, as this means that the overall amount of page views is greater than the overall number of visits, suggesting that browsers stay to read more than one page. Beyond knowing that users stay on the website, referrer data allows a webmaster to form a picture of what are the main sources of traffic to the site. Understandably google is high on the list as the dominant search engine, and the software is able to produce useful information on which search terms bring users to preferred website content. However, such an analysis is beyond the scope of this study. It is also potentially useful to identify which referrers lead to other than search engines lead to preferred content, particularly if there is a targeted campaign or partnership ongoing.

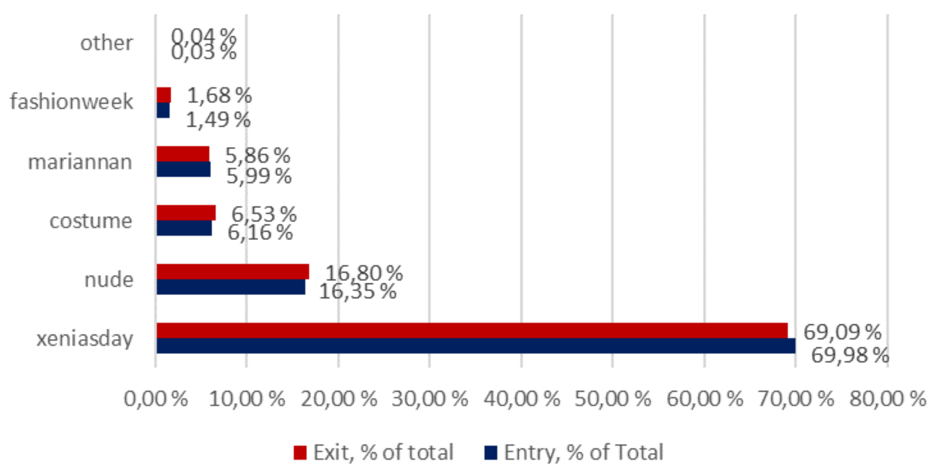

Figure 14: Entry and Exit Pages.

The entry and exit pages are another source of potentially useful information when mapping user behavior on the website. In the case of costume.fi, it is clear that most of the traffic is ongoing in the blogs section, which results in heavily skewed results. It may be useful to realize that only 6% of incoming and outgoing traffic occurs in non-blog related content pages, and to monitor this value as the website is developed to include more featured content. (see Figure 14)

3.4.5 Archetypes of Online User Behavior

Consumer behavior study relies heavily on building generally applicable archetypes of consumers based on their behavioral patterns. The same methodology is applied in this study, with the intent of utilizing established archetypes from consumer behavior study, and applying it to non-commercial online context. Similar key characteristics can be identified from both, which helps to make relevant observations. Table 9 below summarizes the identified key archetypes from consumer behavior study, as well as the proposed counterparts for the purpose of this study.

Table 9: Identified Archetypes

Consumer Behavior Archetypes	**Archetypes Identified in Online Magazine Website**
Economic Shopper	Goal-Oriented Browser
Convenience-Oriented Shopper	Editorial-Content Reader
Recreational Shopper	Recreational Browser
Ethical Shopper	Commercially-Oriented Browser
Personalizing Shopper	Active Contributor
Apathetic Shopper	Non-returning User

Goal-Oriented Browser

Based on identified archetypes by Brown et al. (2003), the goal-oriented browser bears similarity to the *Economic Shopper*. The economic shopper from consumer behavior study is characterized by a keen price-bargain-consciousness, as well as tendency to prefer low-price alternatives. The underlying tendency of the economic shopper revolves around getting the best value for money and time spent, and is highly goal oriented by nature. The parallels drawn to the goal-oriented browser proposed in this study are the purposefulness of actions taken. While the setting of this study does not revolve around purchase decisions, the behavior is similar in that the browser tends to have a clear idea of what they want from the website, exhibited by comparatively clear and repeated behavior, manifesting as direct click-stream paths with little side-tracking.

Table 10: Goal-oriented Browser Descriptives

	N	% of N	MIN	MAX	MEAN
Total volume	1076	100 %	1	69	10
Single-click users	46	4 %			
Multiple-click users	1030	96 %			

It is immediately observable from the dataset that the biggest subset of users belongs to the goal-oriented browser category. Typically, a user of this category finds their way directly to the blogs with little need or desire for navigating the rest of the site. From Table 10 above, we can see that the vast majority of users in the sample had a tendency to read multiple blog posts within the timeframe, with only 4% belonging to the single-click category. On average, users clicked on ten blog links over the period of 6 weeks, while the most active users clicked on up to 69 links. Figure 15 below visualizes a typical path for the goal-oriented browser. Furthermore, the swim-lane diagram illustrates the highly goal-oriented behavior of the archetype.

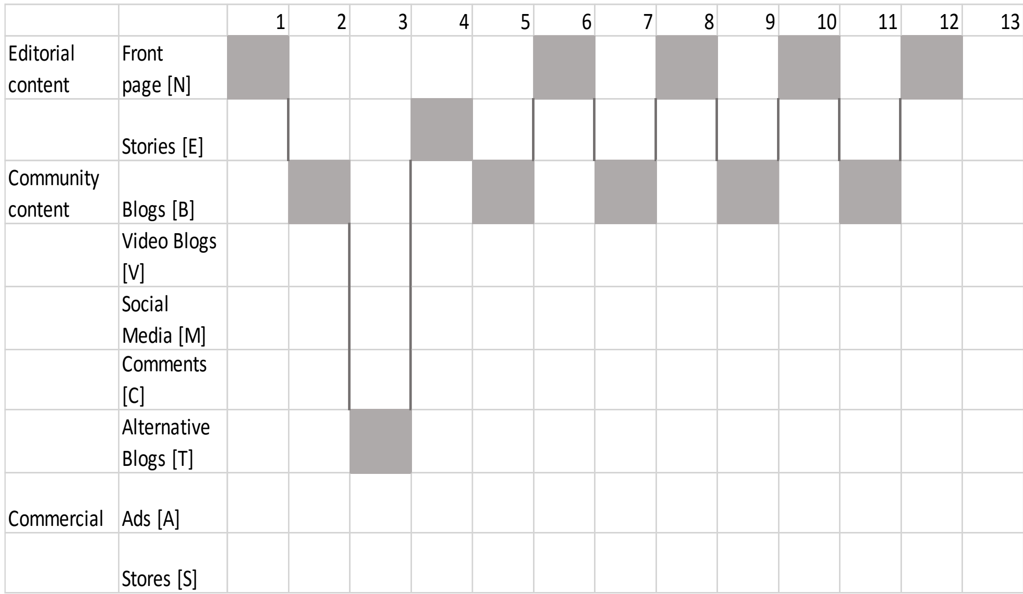

Figure 15: Example Path for Goal-oriented Browser.

Active Contributor

The second user archetype is based on Brown et al. (2003) consumer type *personalizing shopper.* Characteristic to this sub-type is a high emphasis on interpersonal relations with store personnel. While obviously the online environment lacks face-to-face interaction, a similar behavioral pattern can be observed in the Costume.fi virtual community, where there are regular readers for certain bloggers. For the purpose of this study, they will be referred to as *active contributors*, who in addition to being active readers, exhibit behavior of participation in the websites community. Currently this primarily shows in the blogs sections, where this user group actively comments and discusses topics with the bloggers. This user group is a potentially valuable one to reach, as they may be active also in spreading word-of-mouth regarding the website, thus ideally bringing in even more traffic.

Table 11: Active Contributor Descriptives.

	N	% of N	MIN	MAX	MEAN
Total volume	535	100 %	1	43	5
Single-click users	42	8 %			
Multiple-click users	493	92 %			

As the second most common user archetype, the active contributor frequently has multi-click sessions, and is predominantly goal oriented in behavior. On average, active contributor types followed five relevant links, with 92% enjoying multiple clicks per session. The most active users of this archetype clicked 43 links over the test period. (see Table 11)

Figure 16 below illustrates the largely exploratory behavior, where the typical user of this category is shown to browse an assortment of interesting content, while also ending in the comments section frequently.

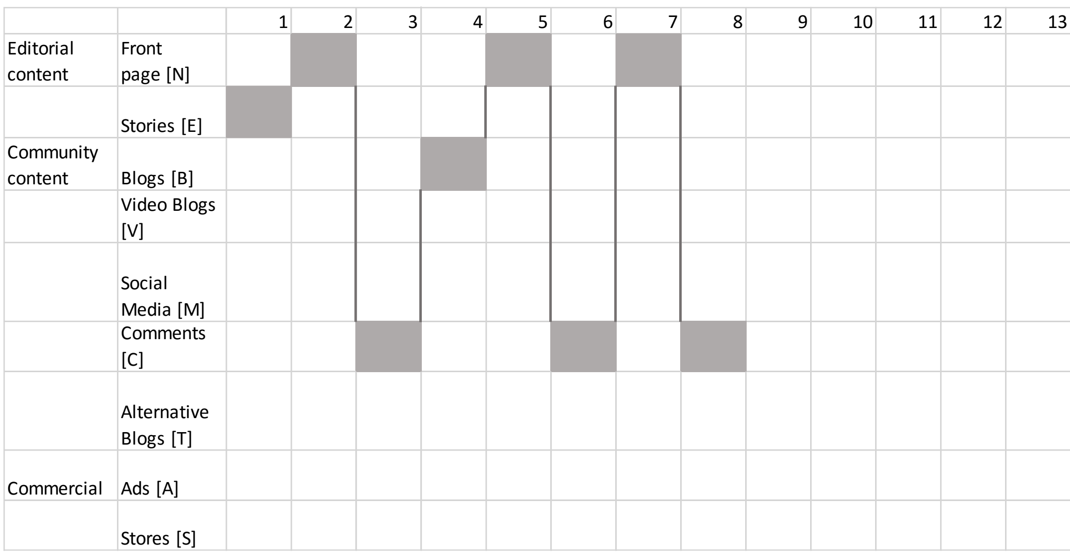

Figure 16: Example Path of Active Contributor.

Commercially-oriented Browser

Characteristic of the *Commercially-Oriented Browser* category is the tendency of clicking store links and advertisements. This archetype is based on the *Ethical Shopper,* which Brown et al. (2003) describe as being brand or store loyal consumers. While on the website of costume.fi it is impossible to determine whether the clicks result in purchases (or indeed, if the clicks were intentional or accidental), it is reasonable to assume that users are at the very least open to finding out more about products that interest them, which they can find access to via the website. In the sample used for this thesis, the user type consisted of only 43 clicks, with an even distribution of single store visits to multiple store visits over the time frame. On average, users of this type clicked commercial links once, while the most active clicked a maximum of five times. (see Table 12)

Table 12: Commercially-oriented Browser Descriptives.

	N	% of N	Min	Max	Mean
Total visits	43	100 %	1	5	1
Single store visits	21	49 %			
Multiple store visits	22	51 %			

In the sample of this study, the shopper category of user did not typically repeat behavior over time, making it a difficult user group to follow actively. However, if this group is included in future research, it is possible that with a larger sample size and longer study period the behavior may be found to repeat. The behavior in general is considered exploratory, as the behavior does not exhibit clear goals in content to be reached. See Figure 17 for an example path of this user type.

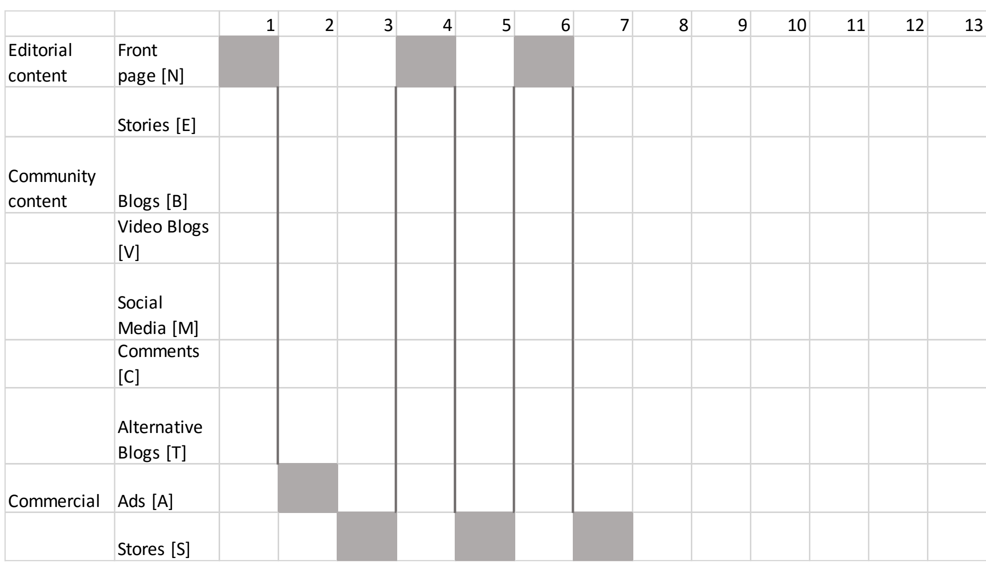

Figure 17: Example Path of Commercially-oriented Browser.

In regards to this category, it can be useful for Aller Media to also know which store links are followed. Table 13 below summarize the clicks to each of the different ad- and store links. It is evident from the data which ad-links attract the most clicks, as well as which store links have driven the most traffic.

Table 13: Summary of Ad- and Store Link Clicks.

Ad-link	N	% of Total
ad.zanox.com	92	34 %
clkuk.tradedoubler.com	177	66 %
track.webgains.com	1	0 %

Store-link	N	% of Total
www.designer-vintage.com	7	16 %
www.verkkokauppa.com	4	9 %
www.vestiairecollective.com	10	23 %
www.iwhiteinstant.com	2	5 %
www.nyxcosmetics.se	7	16 %
www.kosmetik4less.de	1	2 %
www.ebay.com	1	2 %
www.boozt.com	8	19 %
www.vivaladiva.se	1	2 %
fi.oriflame.com	1	2 %
www.netanttila.com	1	2 %

Recreational Browser

The recreational browser is observed to be a media consuming category, utilizing most or all of the available media content on the website. This includes following links to video material, reading blogs (featured and alternative), and overall is eager to consume media content. The browsing behavior as presented by the path analysis is highly erratic and impulsive, characteristic to a high degree of an exploratory user behavior. In the retail context, recreational shoppers are observed to be oriented towards information seeking, and the same is true in the browsing scenario. Brown et al. (2003) further describe the *recreational shopper* archetype, which the recreational browser is based on, to enjoy the shopping process as an experience. Even when no purchase is made, the consumer feels energized from simply being able to go shopping. It is also remarked that this category is widely believed to represent a sizeable portion of consumers. It is also among the largest user groups in the context of costume.fi. Out of the total 263 users exhibiting sufficient behavior to be considered belonging to this user group, 76% visited more than one media-type content page. The behavior is not very typical, as the average number of media content clicks is only two, while the more active media content users reached a high of 17 clicks over the test period. (see Table 14)

Table 14: Recreational Browser Descriptives.

	N	% of N	Min	Max	Avg
Total visits	263	100 %	1	17	2
Single media content visits	64	24 %			
Multiple media content visits	199	76 %			

As mentioned previously, the recreational browser is exploratory in nature, as can be seen in the swim-lane diagram of an example path in Figure 18 below.

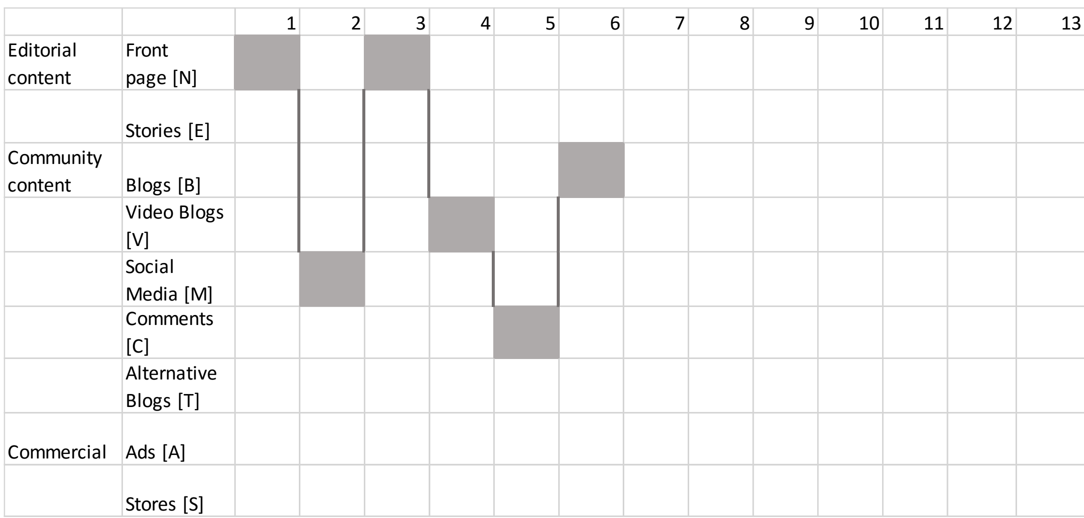

Figure 18: Example Path of Recreational Browser.

Editorial content reader

As per the interest of the company, the editorial content reader is the most valuable user group, and one the company wishes to understand the most. As more content is included in the editorial section of the site, it is suggested to keep an eye on the development of this user group. Currently, the category consists of just under 5% of total sessions. Within the user type, some 12% followed links to editorial content only once, while 88% did so multiple times over the test period. On average, users of this type followed links to editorial content three times, with the most active users doing so 22 times. (see Table 15)

Table 15: Editorial Content Reader descriptives.

	N	% of N	Min	Max	Avg
Total visits	190	100 %	1	22	3
Single editorial content clicks	22	12 %			
Multiple editorial content clicks	168	88 %			

Non-Returning User

As with most behavioral scenarios, there are those whom are interested, and those who are not. While in consumer behavior Brown et al. (2003) identified a class of shopper they call the *apathetic shopper*, who in general is disinterested in the shopping experience, only doing it out of necessity. Similarly, in the costume.fi context there are users who make their way to the website, but for whatever reason do not continue to do so. While it can be argued that a part of these may be due to visiting the website from a public computer, such as from a library, thus explaining why only one visit was made from the IP address. Otherwise, it may be that the users simply found their way to the website, but failed to find content relevant to them, and thus never returned.

3.5 Discussion About the Case Results

Based on Brown et al. (2003) classification of archetypes in consumer behavior, this study considered five active user archetypes, drawing from similarities in behavior to the user types identified by Brown. When observing the results, it became clear that the goal-oriented blog reader category was the dominant archetype, with over 51% of users belonging to this group. Users percentages were calculated based on the relative share of sessions involving any of the identified user categories. The active contributor was identified as the second largest user group with 25% of sessions belonging to this archetype, followed by recreational browser with 12%. Of the three most common archetypes, goal-oriented blog reader and active contributor were observed to be recurring behaviors, while the recreational browser was not.

Table 16: Summary of User Archetypes.

User category	Recurring behavior (Y/N)	Percentage of total sessions	Example click paths
Goal-oriented blog reader	Y	51%	B; BNB; NBNB
Active contributor	Y	25%	NBCB; BC; NCB;
Prospective shopper	N	2%	NASNSNS; NAS; NA; NS
Recreational Browser	N	12%	NMVCB; NBVM; NMNVCB;
Editorial content reader	Y	9%	ENE; BNENE; E;

4 DISCUSSIONS

The aim of the current study is to understand and extend the understanding of user behavior online by examining the paths taken by website users through the theoretical “lens” of traditional consumer behavior study. This chapter represents a summary of the theoretical and empirical sections, and reflects how the findings of the study fulfill the original research questions, what propositions the study will leave the reader with, as well as the research gaps of the study. This section also discusses the theoretical implications of the study.

4.1 Summary of Key Findings

Following the commission for the work by Aller Media Oy, this study investigates actual click-stream paths of website users, with the purpose of understanding how users behave, which behavioral archetypes are observable, and what sort of behavior leads to desired outcomes. As an outcome of the study, six traditional behavioral archetypes (Brown et al. 2003) were identified and linked to related categorical behavior found from the data set. Using the consumer behavioral archetypes as a basis for the study, prevalent online user behavior patterns were identified from within the data, largely using the findings of Ellonen et al. (2015) as a foundation for identifying key patterns. They key presumption prior to the study was that similar behavioral patterns can be observed in the online (non-commercial) environment, as could be identified in a traditional retail consumer setting.

This chapter examines the findings of the thesis from the perspective of the original research questions by discussing the results from the chosen academic literature and the empirical study through answering the original research questions.

How do Web users behave on an online magazine Website?

In traditional retail environments, the study of consumer behavior has had a tremendous impact on how business is done. Based on the study of behavior, some examples of putting a greater understanding of the consumer to use are (supermarket) store design and layout planning, introduction of emotional ideas into marketing, and exhibiting an air of exclusivity in marketing to target the consumers desire to boost self-esteem, based on Maslow's hierarchy of needs. (Ohta & Higuchi, 2013; Kotler & Keller, 2006) Targeting of marketing efforts to the most beneficial consumers has improved rapidly as a result of consumer study, resulting in development of consumer behavioral typologies. Generic behavioral typologies have, in turn, helped develop methods to reach niche customer groups in more effective ways. (Brown et al. 2003)

As a concrete manifestation of online user experience study, path analysis has emerged as a popular tool for measuring and presenting how users interact with websites. Lee & Chen (2010) state that path analysis is among the most important statistical tools used in testing structural models. While path analysis can be done from a statistical analysis basis as suggested by Lee & Chen (2010), this study utilizes a more practical and explorative methodology, incorporating elements of path analysis with sequential analysis, and as a result is difficult to compare results with prior works in academia.

Ellonen et al. (2015) conducted a similar study, which was used as a basis for this thesis, and reached similar conclusions in many issues. In the context of this study, the business model implications are not considered, though the findings mirroring those of Ellonen et al. (2015) certainly reinforce the prevalent thinking that a web-version of a magazine may not necessarily contribute to the business model in as effective a way as print. Despite being done for an online version of a different magazine, the user profiles remained much the same, both in terms of behavior, and in relative size. See Table 17 below for a comparative summary of behavioral pattern subset sizes.

Table 17: Comparison of Behavioral Pattern Subsets by Size.

Behavioral Pattern	Ellonen et al. (2015)	current study
Pure Blog	86 %	80 %
Core Site Content	min. 0,91%*	6,10 %
Commercial	0,36 %	10 %
External (e.g. Social Media)	1,40 %	3,70 %

*different manner of reporting core site content clicks

What behavioral archetypes can be observed in the user base of the Website??

The six behavioral archetypes presented by Brown et al. (2003) were used to identify behavioral patterns of behavior in the online user of a non-commercial website. Through modifying the typological user types to fit the different scenario of the study, many proved still valid. Based on the results of this thesis, however, not all were proven to exhibit recurring behavioral patterns, and as such should not be considered as de facto user groups, when considering precise targeting of users. However, finding similar user behaviors nonetheless suggests that Brown et al. (2003) proposal of finding different user subsets within the existing practice of study can be justified.

Seeing as actual consumption patterns cannot be utilized in order to find similarities, this study focused instead on the experience shoppers/browsers go through in the respective scenarios (see Table 18). As a result, this study proposes new behavioral typologies for the non-commercial, online environment, as discussed in chapter 4.2.

Table 18: **Characteristics of Behavioral Archetypes.**

Behavioral Orientation	Characteristics	Online Equivalent Orientation for Non-Commercial Setting	Characteristics
Economic Shopper	"price-bargain-conscious"; "low-price shopper"; concerned with best value for money; goal-oriented	Goal-Oriented Browser	values easy access to desired content; doesn't care for navigating the part of site which doesn't contain relevant information; goal-oriented behavior
Convenience-oriented Shopper	time-oriented; value effortlessness	Editorial-content Reader	requires some content within website; has a clear purpose to achieve; goal- oriented behavior
Apathetic Shopper	"inactive shopper"; presented as largest segment of consumers;	Non-returning User	finds way to site, but goes inactive after; disinterested in site content; goal-oriented behavior
Recreational Shopper	"active shopper"; enjoys the shopping experience; no purchase required for feeling of fulfilment; believed to represent a sizeable proportion of consumers	Recreational Browser	enjoys most or all available site content; explores in search of interesting content; prone to impulse browsing; exploratory behavior
Ethical Shopper	loyal – store or brand, or both;	Commercially Oriented Browser	prone to following store-links and advertisements; exploratory behavior
Personalizing Shopper	values relationships with store personnel;	Active Contributor	values interaction with content creators; typically active commentators; may become an inactive user, if favored content is discontinued; exploratory behavior

What "paths" within the Website can be identified from the user data?

Based on the data collected from the users over the time period of the study, sequential analysis methods were applied to determine the paths users take on the Website. The visual representations of example cases within each of the identified archetypes helped to determine which typologies are exploratory in nature, and which are goal-oriented.

4.2 Theoretical Implications

From a theoretical perspective, the goal of this study was to participate and contribute to the academic discussion around online consumer behavior and data analytics. As a theoretical foundation of the study, archetypes of consumer behavior as summarized by Brown et al. (2003) were used. Based on observations in traditional brick-and-mortar shopping scenarios, certain key characteristics have been observed time and again in consumers. Based on the underlying characteristics of each behavioral type, this study builds a framework of similar proposed behavioral archetypes for the non-commercial online setting used in the case. The goal is to determine whether the behaviors known to exist in traditional shopping environments can be confirmed to exist in a similar enough form online. In addition to the basic definitions of consumer behavioral categories proposed by Brown et al. (2003), the taxonomy introduced by Montgomery et al. (2004) was utilized to structure clickstream data is utilized based on categories of pages on the Web site.

5 CONCLUSIONS

The case study provided an interesting hands-on opportunity to implement existing theory into a real data-set in order to come up with something new. A challenging aspect in making enduring recommendations for action is that the Web site is very much still a work-in-progress, meaning that much what is suggested here based on the state of the site at the beginning of the thesis work may already be obsolete. Furthermore, with added features existing already, the data collected would have likely provided vastly different insights if taken now. Having said that, it is likely that the identified archetypes of behavior on the website can be utilized in future research, so long as they are adapted to new content.

5.1 Managerial Implications

When considering the direct suggestions for the management of Aller Media, the primary advice to be offered based on the research results is to build the understanding of who are the frequent visitors, and how to best cater to these loyal customers. Furthermore, it is important to understand what kind of behavior leads to desired outcomes on the web site, and then to fine tune the site to be pleasing to the target groups. As a result of this study, we have identified that the majority of return visitors on the site are blog readers and recreational browsers, while the most interesting groups to grow from the company's perspective are the editorial content readers, as well as the visitors ending their sessions with following targeted ad links.

For the purpose of catering to the identified groups of site visitors, Table 19 below provides concrete suggestions for better serving the target groups. Suggestions are partly based on the research of Brown et al. (2003) in approaching web users in a commercial setting, fine-tuned for the online magazine setting in the present study based on my personal view.

Table 19: Summary of Recommended Practical Strategies

Browser Type	Practical Strategies
Goal-oriented Browser	Provide easy access to various content Subscription options for different blogs
Editorial-content Reader	Regular e-mail updates for new editorial content Develop social media content
Recreational Browser	Design and develop visually appealing pages Include entertainment aspects, such as competitions, raffles, chat rooms, comment sections, notice boards
Commercially Oriented Browser	Offer easy access to partner-sites offering store-services Develop an aggregator for recommended store-/product links Targeted ads
Active Contributor	Experiment with ways to connect content creators and browsers (Social Media, Chat capabilities, etc.)

As has been established, the **goal-oriented browser** forms the majority of visitors in the defined categories. The user type is further identified as valuing easy access to desired content, not appreciating a need to navigate excessively through uninteresting content. As such, the suggested approach to satisfy this browser type is to improve easy-access to blog content, for instance through easy subscription functionality.

The next user type is the important **editorial-content reader**, identified as highly goal-oriented in terms of behavior, often only coming to visit the site to read editorial content. As such, it is crucial to keep the user type visiting as regularly as possible. As per the previous user type, the editorial-content reader also benefits from receiving up-to-date information on content being added to the site. Currently there is no clear way receive news on added content, which could be done for instance through e-mail newsletters. Additionally, developing the social media integration can make reaching regular users easier.

The third subset of browser is the **recreational browser**, who is characterized best by a high degree of explorative behavior in browsing. As such, the users of this type are susceptible to all content on the website, as there is a degree of randomness in this behavior. Due to the impulsive browsing, it is important to ensure that the path can be guided to the content which is important for Aller to have reached. As such, if the goal is to have users read editorial content, it should be ensured that the editorial content is visible from anywhere on the site, without becoming intrusive and unappealing to the browser. Furthermore, to make the recreational browser feel stimulated and appeased, it should be considered to offer them forms of engaging entertainment, such as competitions and raffles, chat rooms, and notice boards for interesting events related to the topics of interest. Lastly, seeing as it cannot be predicted where the recreational browser will go on the website, it should be made sure that the website feels equally “complete” on all its sub-pages.

Next to be discussed is the **commercially-oriented browser**. Despite being identified in the analysis, this subset was not particularly significant in terms of size, nor did the behavior prove to be recurring over the time period of the study. Since the data was collected, a new section called “Shop” has been added to the website, so likely this group will grow in size significantly. To encourage commercially-oriented users to reading the site, Aller can provide easy access to partner websites, where purchases can be made. Additionally, it is suggested to provide an aggregator like service, which the recently implemented “Shop” site seems to aim at doing.

Finally, the **active contributor** is a user subset which enjoys engaging in commentary on the site. As such, it can be believed that the users of this type are a high potential group for spreading word-of-mouth awareness regarding the site also in their own personal networks. Motivated by engagement, to satisfy the needs of the active contributor, Aller should ensure that communication between this user subset and the website content creators is possible.

In conclusion, as the case company appears to have a fairly good idea of what they want to achieve with the website, the overall recommendation would be to focus on two or three of the most important user groups. Nurturing their behavior to following desired paths within the site should prove to be an effective way to ensure the website is developed in an optimal way.

5.2 Limitations and Suggestions for Future Research

The primary limitation to this study was the fact that the website for the case study has been undergoing rapid development since the website was launched. With a rapid development cycle adding new features regularly, it is difficult to accurately assess the needs and requirements of the site, as the collected data is already to be considered "old" at the time of finishing the study. Much of what is deduced from the data is thus no longer entirely relevant, as issues may have been addressed already. As such, many of the findings in this study are kept to a general level, to not be as impacted by changes in the website.

From a practical point of view, it is recommended in future studies to utilize more robust tools for analysis of large data sets, such as R or Python. The programs were evaluated in the scope of this research, and while packages exist that can be employed in sequential clickstream analysis, the usage of said programs require a more comprehensive understanding of coding.

6 REFERENCES

Alba, J. & Hutchinson, J. 1987. Dimensions of Consumer Expertise. Journal of Consumer Research, Vol. 14, Iss. 1, pp. 83-95.

Auh, S & Johnson, M. 2005. Compatibility Effects in Evaluations of Satisfaction and Loyalty. Journal of Economic Psychology, Vol. 26, Iss. 1, pp. 35-37.

Ball, D., Coelho, P. & Machas, A. 2004. The Role of Communication and Trust in Explaining Customer Loyalty. European Journal of Marketing, Vol. 38, Iss. 9/10, pp. 1272-1293.

Baumeister, R. 2002. Yielding to Temptation: Self-Control Failure, Impulsive Purchasing and Consumer Behavior. Journal of Consumer Research, Vol. 28, Iss. 4, pp. 670-676.

Bertolucci, J. 2013. Big Data Analytics: Descriptive Vs. Predictive Vs. Prescriptive. [online document]. [Accessed 23 November 2014]. Available: http://www.informationweek.com/big-data/big-data-analytics/big-data-analytics-descriptive-vs-predictive-vs-prescriptive/d/d-id/1113279

Bhatnagar, A. & Ghose, S. 2004. A Latent Class Segmentation Analysis of e-shoppers. Journal of Business Research, Vol. 57, Iss. 7, pp. 758-767.

Bloomberg Businessweek. 2011. The Current State of Business Analytics: Where Do We Go from Here? [online document]. [Accessed 16 November 2014]. Available at http://www.sas.com/resources/asset/busanalyticsstudy_wp_08232011.pdf.

Bogue, R. 2014. Towards the Trillion Sensors Market, Sensor Review, Vol. 34 Iss 2 pp. 137-142.

Borgia, E. 2014. The Internet of Things Vision: Key Features, Applications and Open Issues. Computer Communications, vol. 54, pp. 1-31.

Burby, J., Brown, A., & WAA Standards Committee. 2007. Web Analytics Definitions. Web Analytics Association. [online document]. [Accessed 4 February 2014]. Available at http://www.digitalanalyticsassociation.org/Files/PDF_standards/WebAnalyticsDefinitionsVol1.pdf

Brown, M., Pope, N. & Voges, K. 2001. Buying or Browsing?: An Exploration of Shopping Orientations and Online Purchase Intention. European Journal of Marketing. Vol. 37, Iss. 11/12, pp. 1666-1684.

Catledge, L. & Pitkow, J. 1995. Characterizing Browsing Strategies in the World-Wide Web. Computer Networks and ISDN Systems. Vol. 27, pp. 1065-1073.

Cerrato, D., Crosato, L. & Depperu D. 2016. Archetypes of SME Internationalization: *A Configurational Approach.* International Business Review, Vol. 25, Iss. 1, pp. 286-295. Available from: <http://dx.doi.org/10.1016/j.ibusrev.2015.05.010>

Chen, H., Chiang, R. & Storey, V. C. 2012. Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly, (36: 4) pp.1165-1188.

Clark, L., Ting, I.-H., Kimble, C., Wright, P. C. & Kudenko, D. 2006. Combining ethnographic and clickstream data to identify user Web browsing strategies. Information Research, 11(2) paper 249 [Available at http://InformationR.net/ir/11-2/paper249.html]

Comegys, C., Hannula, M. & Väisänen, J. 2006. Longitudinal Comparison of Finnish and US Online Shopping Behavior Among University Students: The Five-Stage Buying Decision Process. Journal of Targeting, Measurement and Analysis for Marketing. Vol. 14, Iss. 4, pp. 336-356.

Constantinides, E. 2004. Influencing the Online Consumer's Behavior: the Web Experience. Internet Research, Vol. 14, Iss. 2, pp. 111-126.

Dubois, B. 2000. Understanding the Consumer. Prentice Hall, Upper Saddle River, NJ.

Dyer, W.G. and Wilkins, A.L. 1991. Better stories, not better constructs, to generate better theory: a rejoinder to Eisenhardt. Academy of Management Review, Vol. 16, Iss. 3, pp. 613-19.

Ellonen, H-K & Kuivalainen, O. 2008. Exploring a successful magazine web site. Management Research News, Vol. 31 Iss 5 pp. 386 – 398.

Ellonen, H-K. Tarkiainen, A. & Kuivalainen, O. 2010. The Effect of Magazine Web Site Usage on Print Magazine Loyalty, International Journal on Media Management, 12:1, 21-37.

Ellonen, H-K., Wikstrom, P. & Johansson, A. 2015. The Role of the Website in a Magazine Business: *Revisiting Old Truths*. Journal of Media Business Studies. Vol 12, Iss. 4, pp. 238-249.

Engel, J., Kollat, D. & Blackwell, R. 1978. Consumer Behavior. 3rd Edition. Dryden Press.

Eriksson, P. & Koistinen, K. 2005. Monenlainen tapaustutkimus (Diverse Case Study). Kuluttajatutkimuskeskuksen julkaisuja, Helsinki, April.

Hellier, P., Greuse, G., Carr, R. & Rickard, J. 2003. Customer Repurchase Intention: *a General Structural Equation Model.* European Journal of Marketing, Vol. 37, Iss. 11/12, pp. 1762-1800.

Hoffman, D. & Novak, T. 1996. Marketing in hypermedia computer-mediated environments: conceptual foundations. Journal of Marketing, 60, 50-68.

Hoffman, D. & Novak, T. 2009. Flow Online: Lessons Learned and Future Prospects. Journal of Interactive Marketing. Vol. 23, Iss. 1, pp. 23-34.

Ketchen, D. & Shook, C. 1996. The Application of Cluster Analysis in Strategic Management Research: *An Analysis and Critique*. Strategic Management Journal, Vol. 17, Iss. 7, pp. 441-458.

Kinnear, T. & Bernhardt, K. 1986. Principles of Marketing, 2nd Edition. Scott, Foresman and Co., Glenview, IL.

Kolesar, M. & Gabraith, R. 2000. A Services-Marketing Perspective on E-Retailing. Internet Research: Electronic Networking Applications and Policy, Vol. 10, Iss. 5, pp. 424-438.

Kotler, P. & Keller, K. 2006. Marketing Management. 12th edition. Prentice Hall, Upper Saddle River, NJ.

Koufaris, M. 2002. Applying the Technology Acceptance Model and Flow Theory to Online Consumer Behavior. Information Systems Research. Vol. 13, Iss. 2, pp. 205-223.

Lee, A. & Labroo, A. 2004. The Effect of Conceptual and Perceptual Fluency on Brand Evaluation. Journal of Marketing Research, Vol. 41, Iss. 2, pp. 151-165.

Lee, H., Ahn, H., Kim, H. & Lee. J. 2014. Comparative Analysis of Trust in Online Communities. Procedia Computer Science, Vol. 31, pp. 1140-9.

Lee, S. & Chen, L. 2010. The Impact of Flow on Online Consumer Behavior. The Journal of Computer Information Systems. Vol. 50, Iss. 4, pp. 1-10.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H. 2011. Big Data: The Next Frontier for Innovation, Competition, and Productivity McKinsey Global Institute. [online document]. [Accessed 18 November 2014]. Available at http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

Marascu, A. & Masseglia, F. 2005. Mining Sequential Patterns from Temporal Streaming Data. Conference paper

Mittal, V & Kamakura, W. 2001. Satisfaction, Repurchase Intent and Repurchase Behavior: *Investigating the Moderating Effect of Customer Characteristics.* Journal of Marketing Research, Vol. 38, Iss. 1, pp. 131-142.

Moe, WW. 2003. Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream. *Journal Of Consumer Psychology (Lawrence Erlbaum Associates)*, Vol. 13, Iss. 1/2, p. 29, Business Source Complete, EBSCO*host*, viewed 23 October 2015.

Montgomery, A, Shibo, L, Srinivasan, K, & Liechty, J. 2004. Modeling Online Browsing and Path Analysis Using Clickstream Data, *Marketing Science*, Vol. 23, Iss 4, pp. 579-595, Business Source Complete, EBSCO*host*, viewed 23 October 2015.

Narayana, C. & Marking, R. 1975. Consumer Behavior and Product Performance: *an alternative conceptualization*. Journal of Marketing, Vol. 39, Iss. 1, pp. 1-6.

Natschläger, C. & Geist, V. 2013. A Layered Approach for Actor Modelling in Business Processes. Business Process Management Journal, Vol. 19, Iss. 6, pp. 917-932.

Novak, T., Hoffman, D., & Yung, Y. 2000. Measuring the Customer Experience in Online Environments: A Structural Modeling Approach. Marketing Science, Vol. 19, Iss. 1, pp. 22-45, Business Source Complete, EBSCOhost, viewed 27 October 2015.

Ohta, M. & Higuchi, Y. 2013. Study on the Design of Supermarket Store Layouts: *The Principle of "Sales Magnet".* International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering. Vol. 7, Iss. 1, pp. 209-212.

Oliver, R. 1999. Whence Consumer Loyalty. Journal of Marketing, Vol. 63, Iss. 4, pp. 33-44.

Porter, M. 2001. Strategy and the Internet. Harvard Business Review. March 2001.

Press, P. 2013. A Very Short History of Big Data. [online document]. [Accessed: 22 November 2014]. Available at http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/.

Ruh, B. 2012. The Industrial Internet: Even Bigger Than Big Data. [Accessed: 15 November 2014]. Available at http://www.forbes.com/sites/ciocentral/2012/10/04/the-industrial-internet-even-bigger-than-big-data/.

Rummler, G. & Brache, A. 1995. Improving Performance: How to Manage the White Space in the Organization Chart. 2nd edition. Jossey-Bass.

Ståhlstedt, S. 2014. The Role of Non-Coercive Influence Tactics in Online Marketing. Master's Thesis. Lappeenranta University of Technology, School of Business.

Swim Lane Diagrams: Mapping and Improving the Processes in Your Organization, 2015. Available from: <https://www.mindtools.com/pages/article/newTMC_89.htm>

TechAmerica Foundation. 2012. Demystifying Big Data: a Practical Guide to Transforming the Business of Government. [online document]. [Accessed 23 November 2014]. Available at http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf.

United States Patent Application Publication 2004 US 2004/0133671 A1, Jul. 8, 2004.

Urbany, J., Dickson, P. and Wilkie, W. 1989. Buyer Uncertainty and Information Search. Journal of Consumer Research, Vol. 16, Iss. 2, pp. 208-215.

Van der Zee, E. & Scholten, H. 2014. Spatial Dimensions of Big Data: Application of Geographical Concepts and Spatial Technology to the Internet of Things. In “Big Data and Internet of Things: A Roadmap for Smart Environments”, Pp. 137-168. Springer International Publishing.

Yin, R.K. 2003. Case Study Research. Design and Methods, 3rd ed., Sage Publications, Thousand Oaks, CA.