Information Search and Personalization in Electronic Commerce

Shreya Basu

Doctoral thesis, to be presented for public examination with the permission of the Faculty of Social Sciences of the University of Helsinki, in Porthania, PIII, on the 10th of March, 2022 at 12 o'clock.

Supervisor(s): Professor Klaus Kultti

Examiner(s): Mika Kortelainen

University of Helsinki

Faculty of Social Sciences

Helsinki Graduate School of Economics

Supervisor: Professor Klaus Kultti
Pre-examiners: Professor Janne Tukiainen and Otto Kässi

The Faculty of Social Sciences uses the Urkund system (plagiarism recognition) to examine all doctoral dissertations.

ISBN 978-951-51-7912-8 (softcover)
ISBN 978-951-51-7913-5 (PDF)

Unigrafia
Helsinki, March, 2022

Abstract

This thesis studies consumer search behaviour online and its implications on firm performance. The first chapter introduces the overarching topic, providing an overview of the research methodology and key findings. The second chapter examines behavioural implications of consumer types on information search and choice of smartphones online, using demographic, behavioral, browsing history, and detailed product data under laboratory settings. The key finding suggests that opposing personal traits such as conformism and self direction are both associated with extensive search, where the former is steered by bandwagon effects and the latter, by snob effects (demand for a good by individuals of a higher income level is inversely related to its demand by those of a lower income level). Additionally for conformists, price of the purchased good is not reflective of the searched levels, which may be driven by their propensity to choose the most popular alternative rather than the cheapest. This is indicative of conspicuous motives, especially relevant for luxury goods.

The third chapter investigates optimal search paths of online shoppers for *experience* versus search goods, as they engage in continuous sequential search for product information. An optimal stopping rule is designed, based on reservation utilities where the instantaneous utility at each search is modelled as a continuous stochastic process. Furthermore, an empirical model validates the theoretical finding using browsing and purchase data from a Finnish multi-product retailer. The main finding is that, experience goods are associated with three times lower search intensities as compared to search goods. A proxy for the agents' prior information is calculated based on historic search data via novel methodology from the field of information retrieval, such as *Text frequency-Inverse document frequency*, which exhibits an estimated twelve percent increase in search for search goods, while having no effect on experience goods. Finally, the role of personalised recommendations is studied in the context of online search and choice, which has completely opposing effects on the two product types.

The fourth chapter investigates the incentives of e-commerce platforms to show personalized recommendations and its effects on performance. A theoretical framework is developed that characterizes the optimal decision policy of a firm, given current state of shoppers. The key finding is that the firm must always show recommendations to shoppers in the high state above a certain price or value threshold. In the low state, recommending is optimal if the "salience effect" is above a threshold that maximizes discounted future stream of profits. An empirical model provides support to the theoretical findings, highlighting the reputation effects of platform recommendations, using browsing and purchase data from a Finnish multi-product platform. While recommendations are associated with a 29% increase in firm revenue, relevance of such recommendations potentially boost revenue by a significant 30%. Furthermore, strong evidence is presented that consumer state is endogenous in firm revenue regressions. A three-step IV process extracts the direct effect of consumer state on revenue which shows positive association between reputation effects and firm performance.

Acknowledgement

In my PhD journey of many peaks and troughs I have received unflinching support from several institutions, colleagues, friends and family. I wish to express my sincere gratitude to each one of them for the inspiration and encouragement that helped me carry on.

First, I would thank my supervisor, Klaus Kultti for his utmost patience, astute guidance and particular humour all through the project. I continue to carry your teachings and perspectives not only in the field Economics, but life in general.

I feel grateful to have worked closely with Topi Miettinen, your insightful comments and feedback have helped me tremendously. I would like to thank my pre-examiners Janne Tukiainen and Otto Kässi for their invaluable feedback and especially to Mika Kortelainen for agreeing to be my oppnonent.

I feel lucky to have met many wonderful collegaues, who acted both as sparring partners and encouraging friends during my stay in Finland, Kristine, Annika, Michaela, Marlene, Olena, Min, Anustup, Saara, Tuomas, to name a few, as well as participants of HECER seminar series and Hanken lunch seminars who provided insightful feedback. Thank you for creating such a safe and stimulating work environment that certainly contributed to my progress. I'm thankful to the Yrjo Jahnssen Foundation, OP Pohjola and HECER for providing me financial support on this journey.

Writing a PhD next to a full time job would have been impossible without the support of my friends, Marieke and Yue. Thank you both for having my back and keeping my stress levels at bay. Finally, I want to thank Shubh as you are the reason I embarked upon this journey that taught me resilience, more than anything else. I'm truly lucky to have my best friend, my biggest critic and my fiercest cheerleader, all wrapped in a life partner. Even though the PhD comes to a conclusion, I hope we continue to have our discussions on Ito processes and recommender systems.

I cannot end this note, without thanking my parents for their unconditional support, monumental faith and teaching me to have my feet firmly on the ground, always.

Contents

Chapter 1

Introduction

1.1 Motivation

Electronic commerce has created an interactive, networked economy where communications and market processes are synergistic and immediate. This thesis is a contribution to the vast body of literature studying demand and supply side behaviour in the internet markets. In 2019, retail electronic commerce sales worldwide amounted to 3.53 trillion US dollars and electronic retail revenues are projected to grow to 6.54 trillion US dollars by 2022[1]. Commercial interactions online have drastically increased in the past two decades thus motivating multi-disciplinary theoretical as well as empirical academic research in fields spanning across economics, computer science and marketing, to name a few. Online search and transaction data is as rich as it is structurally complex, and can be harnessed to understand and predict consumer behaviour, as agents aim to maximize payoffs in the long run.

Economic literature on consumer search in the internet markets spans across several areas, such as, price dispersion (Sorensen, 2000; Baye et al., 2006; Chandra and Tappata, 2011; Richards et al., 2016), search cost estimation (Hong and Shum, 2006; Moraga-González and Wildenbeest, 2008; Moraga-González et al., 2013), attribute search and learning (Moorthy et al., 1997; Huang et al., 2009; Koulayev, 2009; Branco et al., 2012; De los Santos et al., 2013) or information obfuscation (Muir et al., 2013; Ellison and Ellison, 2009) in a diverse

[1]Source: https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/

spectrum of consumer goods or services. The primary focus of this thesis is to study the intensity of online search for various class of consumer goods, when varying behavioural motivations or retailer instruments are at play. In that, I build on the existing research in three distinct ways, as described below.

Firstly, behavioural characteristics of individuals, obtained in a laboratory setting, are mapped onto their unique search paths in order to investigate shopper motivations or values driving online search and choice. While personal traits have been amply linked to consumption patterns in the past (Corneo and Jeanne, 1997; Hirschman and Holbrook, 1982; Doran, 2009; Kastanakis and Balabanis, 2014), there has been relatively less focus on its association with search behaviour. A part of this thesis shows the significance of personal traits and behavioural characteristics as individuals engage in everyday search activities, which in turn, provided basis for firms to differentiate selling strategies across consumer types.

Secondly, *personalized recommendations* on retailer websites and their impact on consumer decision making and firm performance are studied at length. Personalized recommendations are generated by algorithms that analyze shoppers' browsing history, transaction data as well as latest digital media trends. Such tools are increasingly used by online platforms to not only inform potential buyers of what is available but also influence their choice of purchase. There is a multitude of studies capturing how firms can attract consumer attention online by means of advertising (Grossman and Shapiro, 1984; Economides and Salop, 1992; Chiou and Tucker, 2010; Haan and Moraga-González, 2011; Lewis and Nguyen, 2015), online reviews (Mudambi and Schuff, 2010; Cui et al., 2012) and search rankings (Athey and Ellison, 2011; Blake et al., 2015; Ursu, 2018), to name a few. However, there is a gap in literature with regards to personalized recommendations, which is a crucial aspect of pricing or promotion strategy in the online retail space today[2].

And thirdly, novel methodology from the field of information retrieval and text data mining is used to disentangle product attribute information obtained by consumers and quantify their match quality of sampling each alternative, in the context of information search. In the past, information search has been limited to a discrete set of product attributes,

[2]"More than 35 percent of what consumers purchase on Amazon and 75 percent of what they watch on Netflix come from personalized product recommendations" - Mckinsey Report *How retailers can keep up with consumers*

which does not take all relevant information into account in the context of consumer learning. Matching entire product descriptions of every searched good and the final purchase not only allows us to gauge shoppers' product knowledge prior to purchase, but also, determine the relevance of recommended products to their preferences.

The first two chapters of this thesis study consumer behaviour online: while the first chapter focuses on the behavioural motivations of search and choice for smartphones, the second chapter explores information search considerations for a broad range of *experience* versus *search* goods (Nelson, 1970). The final chapter focuses on the firm's optimal decision policy in designing personalized recommendations which influence consumer search patterns online and ultimately choice of goods purchased.

1.2 Research Methodology

This dissertation is a collection of three self-contained articles that are based on two unique datasets: 1) experimental and survey data on shopper motivations along with their exhaustive search paths on the internet 2) observational clickstream data[3] constrained to a large multi-product retail platform based in Finland[4].

The first article examines behavioural motivations in potential buyers that drive certain paths of search and purchase decisions. A novel experiment was designed and executed in multiple steps to not only capture exhaustive search paths of the subjects prior to choosing their preferred smartphones, but also their values, behavioural motivations and demographic characteristics. The experiment had three distinct phases: 1) a pre-questionnaire based on the Portrait Value Questionnaire (PVQ) (Schwartz, 2003), Frederick's (2005) Cognitive Reflection Test (CRT) (Frederick, 2005) and Arnett's (1994) Sensation Seeking scale (AISS) (Arnett, 1994) which outlined motivational goals, value systems and risk attitudes of the subject pool; 2) a search assignment, conducted in a computer laboratory weeks after the pre-questionnaire was completed, involving three sequential tasks with several treatment variations and the end goal of choosing the preferred smartphone; 3) a post-questionnaire to measure subjects' overarching affinity towards shopping online. The primary objective

[3]Detailed log of how individuals navigate through the web site during a task. The log typically includes the pages visited, time spent on each page, how they arrived on the page, and where they went next

[4]Due to non-disclosure agreements with the data provider, any identifiable information, such as name of the platform is not mentioned in the thesis

of this experiment was to study the impact of personality traits on the length and depth of search and choice online. This article builds on the work of Bronnenberg et. al on consumer search for cameras online (Bronnenberg et al., 2016), by introducing detailed behavioural attributes of representative shoppers. While the length of search was measured by number of domains searched, depth of search was measured by search queries on any search engine, such as Google. In order to estimate the effect of behavioural attributes and gender, an OLS specification is used for each task in the the search assignment as well as the combined data set. Findings in task A are leading as later tasks were designed with the treatments as focal features. As the nature of the data collected exhibits panel structure, standard errors are clustered at the task level. Building on and adding to the extensive past research on search for the 'best price', my co-author and I study shoppers' behavioural implications on search for smartphones in Finland, based on similar modelling framework. This article, furthermore, maps search paths to choice of good to be purchased to investigate the empirical relationship between search and choice given shopper characteristics and values. To that end, we test if search is largely predictive of choice, and if specific behavioural traits could be associated with any deviation from the standard prediction. We start by estimating the mean effect of each of the three continuous product attributes, namely, price, memory and shipping cost on their chosen levels. Next, we investigate basis for convergence of search to choice by replacing mean searched attribute levels with recency-weighted mean searched attribute levels, and finally include interaction terms with personality values and searched attribute levels to the baseline model that allows a better understanding of varied search paths that leads to the ultimate choice of good purchased, depending on consumer types.

The empirical analyses in the second and final article are based on unique browsing and transaction data of online shoppers, from a Finnish multi-product retail platform. Several studies in the past two decades have elicited the potential that exists in studying an individual's behavior as they navigate from one webpage to another, with the intent to find the best alternative (Hoffman and Novak, 1996; Moe and Fader, 2004; Johnson et al., 2004; Kim et al., 2010; De los Santos, 2018). Clickstream data, as has been used for the purpose of this thesis, is structurally complex and poses several operational challenges around pre-processing, however, provides fascinating insights into a potential buyer's journey on internet platforms, from search to choice. Additionally, the data set observes when browsers click on firm-generated product recommendations.

The second article studies the varying intensities of consumer search associated with search and experience goods on an online retail platform, as they receive product recommendations based on their search history. The key objective is to investigate differences in search patterns for experience versus search goods as shoppers engage in information search to find the best alternative. A theoretical model is developed that pins down optimal stopping rules of shoppers for the two classes of differentiated goods, while the instantaneous utility they derive at each search event is modelled as a generalized Brownian motion or an Ito process (Dixit and Pindyck, 1994; Ito, 1944, 1957). The empirical analysis is based on the OLS specification that estimates the effects of search variance, prior product knowledge and quality of personalized recommendations along with a set of control variables, on the extent of search online. The key variables of interest are calculated as discussed next. The extent of search is largely dependent on its informativeness as defined by the inverse of the shoppers' search variance, which is higher for experience goods as compared to search goods. Discrete choice models can be applied to evaluate choice probabilities depending on the specifications of density of the unobserved factors. I calculate search variance by taking the average of the difference between the choice probabilities across two sequential search events, where choice probabilities are derived from a standard logit specification. Additionally, consumer search and learning is captured via novel methodology from the field of information retrieval and text analysis. To derive a measure for prior product knowledge, I start by using *text frequency-inverse-document frequency* (tf-idf) (Sparck Jones, 1988) to translate product descriptions into attribute vectors. This method essentially allows the most relevant features to be retrieved from complete product descriptions. Next, to determine how similar searched products are to each other, the *cosine similarity* between two vectors are calculated. Cosine similarity is a comparison metric between two product descriptions on a normalised space, which not only takes the magnitude of each word count (tf-idf) of each vector into consideration, but also the angle between each pair of vectors representing product attributes. This is a simple way of handling text data which can be used to measure how similar products are to each other, based on their descriptions. This, in turn, informs of the shoppers' attribute knowledge of the bought good, in the sessions prior to purchase. A similar approach is followed in order to measure the match quality or relevance of product recommendations, in the event that a shopper clicks on such a recommendation.

The final article focuses on firm's incentives that motivate the use personalized recommendations, conditional on the state of the consumer. Online platforms use personalized recommendations to direct potential buyers to their desired products, however in designing these recommendations, firms face a trade-off between earnings and relevance to the buyer. This article deals particularly with such a trade-off which is shown not only to impact short-run firm performance, but have significant reputation effects impacting firm earnings in the long-run. A theoretical framework is presented that outlines optimal policies of the firm based on current profits and transition probabilities of shoppers switching between states. Given the two consumer states (High, Low), the decision variable of the firm is based on four exhaustive policies. Pair-wise comparisons of these policy combinations (Bellman equations) lead to examining each case in detail and ultimately pinning down price or value thresholds above which it is optimal for the firm to recommend. The empirical analysis is designed such that the reputation effects of firm recommendations over time are taken into account, as consumer states are largely dependent on the firm's reputation. As a first step in understanding the relationship between firm revenue and consumer state, I examine the OLS estimators of firm revenue with consumer state and relevance of recommendations as primary regressors, along with a set of control variables. A measure for relevance of product recommendation is calculated using similar methodology as described in the first article, namely tf-idf and cosine similarity. The OLS estimates establish the degree of association between consumer state and firm performance, but do not elucidate causation. As the goal is to estimate revenue impact to exogenous changes to consumer state, which is likely to be dependent on reputation of firm generated recommendations, I use instrumental variables to address possible endogeneity and isolate the effects of consumer state on firm revenue from any other sources of variation. The endogenous variable, consumer state, is binary in nature hence the three-step IV procedure (Renee Adams and Ferreira, 2009; Wooldridge, 2001) is used. This eliminates the possibility of a *forbidden regression* (Angrist and Pischke, 2009) which occurs when the standard 2SLS method is applied to a non-linear model.

1.3 Summary of findings

This section summarizes the key findings in all three articles in this thesis, which aims to provide a deeper understanding of consumer search and choice in the internet markets. This thesis attempts to approach the subject holistically both in terms of demand side considerations and supply side instruments, that are pivotal in determining commercial interactions online today.

The first article sheds light on the differences in search patterns for smartphones, based on inherent motivational goals and value systems of shoppers. While conformism and self-direction represent contrasting personality traits, individuals with these traits exhibit similar search patterns in our data. Though surprising at the outset, both of these traits are positively associated with the extent of search, stemming from bandwagon effects for the former and conspicuous motives for the latter. This has interesting implications for retailers as it warrants the use of appropriate instruments to influence purchase behaviour of the different consumer types, despite their search behaviour indicating that they might belong to the same cohort of individuals. Generally, search is shown to be predictive of choice, with shoppers exhibiting hedonistic tendencies being the only exceptions. Furthermore, we show increased search leads to a higher likelihood of choice and improved firm performance. This finding particularly emphasizes the importance of observing and analyzing the path of convergence of search to choice for retailers, such that they are able to show product recommendations at specific points of the search path where probability of purchase is the highest.

The second article illuminates the structural differences in search patterns across a diverse set of experience and search goods. One of the findings is that personalized recommendations boost search intensity for search goods by approximately 17%, but reduces it by 9% for experience goods. This further provides basis to some of the managerial implications in the first article. Interestingly, similar search behaviour is observed for a typical search good in subjects participating in a controlled laboratory experiment where search prevails across several domains, in the first article, and observational click-stream data from a single multi-product platform in the second article. As search goods enable shoppers to determine true match quality prior to purchase, both the theoretical model's optimal stopping rule and empirical evidence on the extent of information search show that for a typical search good, shoppers search at least three times more than a typical experience good.

The third article studies firm incentives to show personalized recommendations to potential buyers and how reputation effects play a significant role in the design of long-run recommendation policies of online platforms. First and foremost, it is shown that recommendations are associated with improved firm earnings. This, in combination with findings from the second article, indicates a positive association between consumer search intensity and firm performance, especially for search goods. This motivates online retailers and platforms to judge their performance based on shopper engagement. Additionally, I also find that firm performance improves as the quality or relevance of recommendations improve. However, it is optimal for the firm to recommend only above a certain price threshold. Given the two consumer states, I start by laying out policy options for the firm, and then examine several cases in detail which pins down optimal price or value thresholds. The empirical model investigates the unobserved reputation effects of firm-generated recommendations on their earnings. The three-step IV treatment allows isolating this effect showing a significant positive relationship between firm reputation and earnings. Furthermore, I show that the long-run performance of the firm is positively associated with consumers remaining in the high state in every period, that is, purchasing via recommendation and not search. This highlights key determinants in developing long-term profit-maximizing recommendation strategy for online platforms. This study implicitly points to the value that can be created by online platforms for potential buyers, which in turn influences the probability of purchase. For example, we find that depth of search or the time spent per product pages are both positively associated with firm earnings, while the total number of pages viewed do not have any meaningful impact. Therefore, the quality of product information and relevance of recommendations are key in converting browsers to paying customers.

References

Joshua Angrist and JÃ¶rn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. 01 2009. ISBN 9780691120348 (hardcover : alk. paper).

Jeffrey Arnett. Sensation seeking: A new conceptualization and a new scale. *Personality and Individual Differences*, 16(2):289 – 296, 1994.

Susan Athey and Glenn Ellison. Position auctions with consumer search. *The Quarterly Journal of Economics*, 126(3):1213–1270, 2011.

Michael R Baye, John Morgan, Patrick Scholten, et al. Information, search, and price dispersion. *Handbook on economics and information systems*, 1:323–375, 2006.

Thomas Blake, Chris Nosko, and Steven Tadelis. Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica*, 83(1):155–174, 2015.

Fernando Branco, Monic Sun, and J Miguel Villas-Boas. Optimal search for product information. *Management Science*, 58(11):2037–2056, 2012.

Bart J. Bronnenberg, Jun B. Kim, and Carl F. Mela. Zooming in on choice: How do consumers search for cameras online? *Marketing Science*, 35(5):693–712, 2016.

Ambarish Chandra and Mariano Tappata. Consumer search and dynamic price dispersion: an application to gasoline markets. *The RAND Journal of Economics*, 42(4):681–704, 2011.

Lesley Chiou and Catherine Tucker. How does pharmaceutical advertising affect consumer search? *Retrieved from June*, 10:2018, 2010.

Giacomo Corneo and Olivier Jeanne. Conspicuous consumption, snobbism and conformism. *Journal of Public Economics*, 66(1):55 – 71, 1997.

Geng Cui, Hon-Kwong Lui, and Xiaoning Guo. The effect of online consumer reviews on new product sales. *International Journal of Electronic Commerce*, 17(1):39–58, 2012.

Babur De los Santos. Consumer search on the internet. *International Journal of Industrial Organization*, 58(C):66–105, 2018.

Babur De los Santos, Ali Hortaçsu, and Matthijs R Wildenbeest. Search with learning. *Available at SSRN 2163369*, 2013.

Avinash K. Dixit and Robert S. Pindyck. *Investment under Uncertainty*. Number 5474 in Economics Books. Princeton University Press, 1994.

Caroline Josephine Doran. The role of personal values in fair trade consumption. *Journal of Business Ethics*, 84(4):549–563, 2009.

Nicholas Economides and Steven C Salop. Competition and integration among complements, and network market structure. *The Journal of Industrial Economics*, pages 105–123, 1992.

Glenn Ellison and Sara Fisher Ellison. Search, obfuscation, and price elasticities on the internet. *Econometrica*, 77(2):427–452, 2009.

Shane Frederick. Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4):25–42, December 2005.

Gene M Grossman and Carl Shapiro. Informative advertising with differentiated products. *The Review of Economic Studies*, 51(1):63–81, 1984.

Marco A Haan and José L Moraga-González. Advertising for attention in a consumer search model. *The Economic Journal*, 121(552):552–579, 2011.

Elizabeth C. Hirschman and Morris B. Holbrook. Hedonic consumption: Emerging concepts, methods and propositions. *Journal of Marketing*, 46(3):92–101, 1982.

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

Han Hong and Matthew Shum. Using price distributions to estimate search costs. *The RAND Journal of Economics*, 37(2):257–275, 2006.

Peng Huang, Nicholas H Lurie, and Sabyasachi Mitra. Searching for experience on the web: an empirical examination of consumer behavior for search and experience goods. *Journal of marketing*, 73(2):55–69, 2009.

Kiyosi Ito. Stochastic integral. *Proceedings of the Imperial Academy*, 20(8):519–524, 1944.

Kiyosi Ito. On stochastic differential equations. *Matematika*, 1:78–116, 1957.

Eric J. Johnson, Wendy W. Moe, Peter S. Fader, Steven Bellman, and Gerald L. Lohse. On the depth and dynamics of online search behavior. *Management Science*, 50(3):299–308, 2004.

Minas N. Kastanakis and George Balabanis. Explaining variation in conspicuous luxury consumption: An individual differences' perspective. *Journal of Business Research*, 67 (10):2147 – 2154, 2014.

Jun B. Kim, Paulo Albuquerque, and Bart J. Bronnenberg. Online demand under limited consumer search. *Marketing Science*, 29(6):1001–1023, 2010.

Sergei Koulayev. Estimating search with learning. *Available at SSRN 1285619*, 2009.

Randall Lewis and Dan Nguyen. Display advertisingâs competitive spillovers to consumer search. *Quantitative Marketing and Economics*, 13(2):93–115, 2015.

Wendy W. Moe and Peter S. Fader. Capturing evolving visit behavior in clickstream data. *Journal of Interactive Marketing*, 18(1):5 – 19, 2004.

Sridhar Moorthy, Brian T Ratchford, and Debabrata Talukdar. Consumer information search revisited: Theory and empirical analysis. *Journal of consumer research*, 23(4):263–277, 1997.

José Luis Moraga-González and Matthijs R Wildenbeest. Maximum likelihood estimation of search costs. *European Economic Review*, 52(5):820–848, 2008.

José Luis Moraga-González, Zsolt Sándor, and Matthijs R Wildenbeest. Semi-nonparametric estimation of consumer search costs. *Journal of Applied Econometrics*, 28(7):1205–1223, 2013.

Susan M Mudambi and David Schuff. Research note: What makes a helpful online review? a study of customer reviews on amazon. com. *MIS quarterly*, pages 185–200, 2010.

David M. Muir, K. Seim, and Maria Ana Vitorino. Price obfuscation and consumer search : An empirical analysis. 2013.

Phillp Nelson. Information and consumer behaviour. *Journal of Political Economy*, 78(2): 311–329, 1970.

Heitor Almeida Renee Adams and Daniel Ferreira. Understanding the relationship between founder ceos and firm performance. *Journal of Empirical Finance*, 16(1):136 – 150, 2009.

Timothy J Richards, Stephen F Hamilton, and William Allender. Search and price dispersion in online grocery markets. *International Journal of Industrial Organization*, 47:255–281, 2016.

Shalom Schwartz. A proposal for measuring value orientations across nations. *European Social Survey*, pages 259–290, 01 2003.

Alan T Sorensen. Equilibrium price dispersion in retail markets for prescription drugs. *Journal of Political Economy*, 108(4):833–850, 2000.

Karen Sparck Jones. *A Statistical Interpretation of Term Specificity and Its Application in Retrieval*, page 132â142. Taylor Graham Publishing, 1988.

Raluca M Ursu. The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, 37(4):530–552, 2018.

Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*, volume 1 of *MIT Press Books*. The MIT Press, February 2001.

Chapter 2

personal traits and online footprint in smartphone search[1]

2.1 Introduction

Information search online is potentially a very complex process when there are multiple product attributes that the consumer must evaluate. This complexity challenges researchers attempting to analyze such processes. Several pioneering methodological approaches have been proposed to study consumer search over the internet, in particular (Stigler, 1961; McCall, 1970; Burdett and Judd, 1983; Janssen and Moraga-GonzÃ¡lez, 2004; De Los Santos et al., 2012; De los Santos, 2018). The complexity of the process has necessitated simplifications for instance, by looking at search on a particular retailer's website (Kim et al., 2010), or merely listing particular retailer domains that the consumer has visited (Johnson et al., 2004; Park and Fader, 2004; Huang et al., 2009). Rarely is there data available of the entire search process, including how consumers use comparison sites and search engines during the process and what a typical search path look like. There are a few exceptions, as the challenges facing the researchers in this field are described by Bronnenberg et al.(Bronnenberg et al., 2016) as follows: "A comprehensive collection of online searches and purchases requires casting a net over a very large number of consumers over an extended period of time. The resulting browsing data captured in this way is potentially enormous, impeding its procurement and

[1]This chapter is based on an article jointly written with Topi Miettinen and Jaakko Aspara

processing. Second, URLs browsed, containing the characteristics of the products searched are typically dynamic or perishable. This requires extracting the information displayed on the pages requested by consumers concurrently with their search and choice activity".

In this paper, we propose a novel and complementary laboratory method that allows us to tackle several of those challenges. We invited participants to a computer laboratory and engage them in an ecologically valid internet search task with real incentives, yet of limited duration. The laboratory approach allowed us to collect data of the entire search paths across websites and keeping track of everything that appears on the screen during the process. Furthermore, we collected data on detailed consumer characteristics, such as values or cognitive reasoning styles. This enables us to profoundly understand how the characteristics correlate with search patterns and choice of smartphones in Finland. Our evidence complements the field evidence studying consumer search behavior in Bronnenberg et al. (Bronnenberg et al., 2016).

The experiment had three distinct phases: phase one involved participants filling a pre-questionnaire eliciting personal characteristics, risk attitudes and cognitive reasoning styles. Phase two was conducted in a computer laboratory about a week later, which involved three sequential search tasks, where each participant was to engage in online search with the aim of choosing their preferred smartphone. Finally, in phase three participants were required to fill a post-questionnaire eliciting individual preferences with regards to online shopping. In each search task, each participant was given a 1000 euro budget to spend on a mobile phone which they needed to search over the internet and place in a virtual shopping cart at any website selling mobile phones. One of the participants and one of her/his three tasks was randomly drawn to receive the phone placed in the cart at the price listed. The participant would also receive the residual budget net of a 10% commission[2].

We find that the search patterns of the students in the laboratory are closely reminiscent of the patterns observed in larger scale in the field studies (Johnson et al., 2004; De Los Santos et al., 2012; De los Santos, 2018). The students visit several domains during the search process using search engines, occasionally using comparison sites, but after all, they visit a small number of retailer sites. While Bronnenberg et al. found that on average 3 brands

[2]This basic principle applied in all three tasks. There was some additional variation in the incentives across the three task. Details in Section 2.2

and 6 models were compared during the two-week process addressed in their study, our participants compare on average 2.4 brands and 3.4 models during the 8-minute search process. In Bronnenberg et al's study, more than 70% only visited one retailer site, while in our study 60%-70% of the participants do so. Nevertheless, we observe rich and variant search sequences in our data across search engines, comparison sites, retailer sites, etc.

The key advantage of this study is that we have access to detailed personality traits and demographic data across all participants. These were elicited, through an internet questionnaire a week prior to completion of the incentivized search tasks in a computer lab. When associating search behaviour with personal traits, we show empirically that extensive search is associated with the values of conformity, self-direction and hedonism. Prior research has identified these as key factors positively associated with bandwagon effects or snob effects leading to conspicuous consumption associated with high-end goods (Leibenstein, 1950; Corneo and Jeanne, 1997; Vigneron and Johnson, 1999, 2004; Wiedmann et al., 2009; Kastanakis and Balabanis, 2014). This paper shows how value motivations are reflected in consumer search and purchase behaviour. Arguably, both snobbism (driven by self-direction) and bandwagon effects (driven by conformism) necessitate more careful consideration of the product that matches the needs generated by the inter-dependent values. Human values are conceived as static constructs that involve any criteria or standards of preference (William A Darity, 2008)and therefore can be used to forecast behaviour and choice patterns of consumers (Kamakura and Novak, 1992; Doran, 2009). We find that although conformism and self direction represent opposing value motivations, namely conservatism and self-enhancement, respectively (Schwartz, 1992, 2003), they have a directionally similar association with search for smartphones. Additionally, hedonistic tendencies too lead to extensive search, although it is of recreational nature rather than being task-oriented.

We find in our laboratory settings, that the smartphones shoppers in Finland exhibit search patterns comparable to findings in prior search literature (Johnson et al., 2004; De Los Santos et al., 2012; Bronnenberg et al., 2016; De los Santos, 2018). One such finding is that, search is predictive of choice and generally speaking, leads to choice, over a shopper's online journey. Interestingly, this study sheds light on some exceptions to the above: 1) Hedonistic values typically induce search behaviour that is recreational rather than being goal-oriented (Hirschman and Holbrook, 1982; Griffin et al., 2000; Chaudhuri et al., 2010) and therefore, may not be predictive of choice; 2) For conformists, extensive attribute search does not necessarily lead to the purchased price. The key contribution of this paper lies in the pinning down the behavioural implications of consumer traits on search and choice.

The rest of the paper is organized as follows. An elaborate description of the experiment design and methodology is provided in Section 2. Section 3 presents the descriptive results and Section 4 tests empirically several hypotheses relating to consumer types, search and choice based on experimental data.

2.2 Data collection design and methodology

There has been significant research in the recent past about online search behaviour, primarily focusing on the size of search costs or the search strategy adopted. The objective of this paper is to study the link between consumers' commonplace traits and personality attributes, on the one hand, and their search and purchase behaviors, on the other hand. The richness of the compiled dataset lies in that it covers the end-to-end process, starting from search to purchase. For instance, 26% of the total search domain visits across all tasks can be attributed to Google, followed by 21% and 15% to two of the biggest retailers for electronics in Finland namely, Verkkokauppa and Gigantti, respectively. The design of the experiment not only allows us to observe activity within a store, but also across stores, brands and products, including cases where subjects search on one store or domain and eventually purchase from another. For instance, 28% of the population went directly to a retailer's domain to make a purchase, whereas 12% searched on Google first before they landed on their purchase domain.

The data was collected by monitoring search behaviour of 69 students[3] in a laboratory setting, where the task was to search and choose a smartphone. In addition to search behaviors, also personality traits were collected. The experiment was conducted in three different stages: a pre-questionnaire to be filled over the internet before coming to the laboratory, a search assignment experiment in a decision making laboratory and a post-questionnaire over the internet immediately after the experimental task in the laboratory. A pre-survey and post-survey were conducted with the aim of gathering information about consumer characteristics and their preferences towards shopping online. We will now explain in detail the procedures in each of the three parts of this study.

[3]Participants were students at the Hanken School of Economics

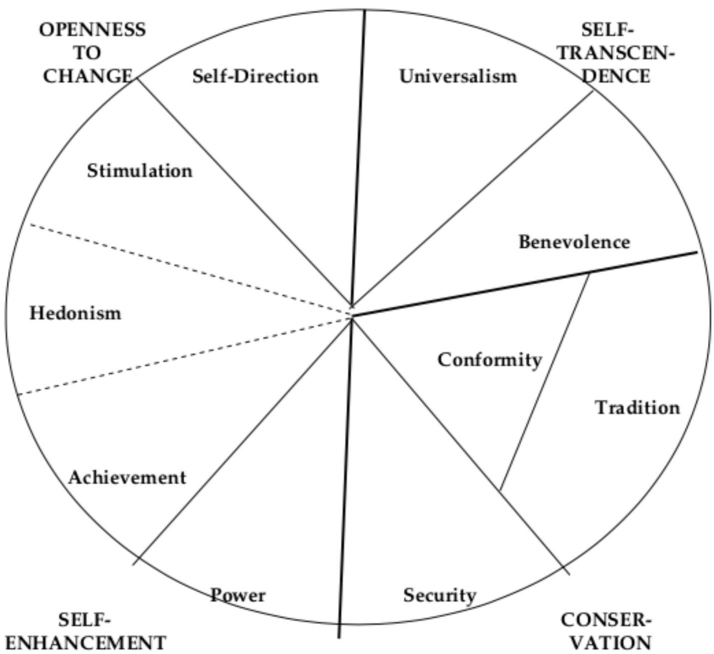
2.2.1 Pre-questionnaire

The online pre-questionnaire was designed to collect data on subjects' personality traits and demographics. Personality traits were determined via a combination of Schwartz's Portrait Value Questionnaire (PVQ) (Schwartz, 2003), Frederick's Cognitive Reflection Test (CRT) (Frederick, 2005) and Arnett's Sensation Seeking scale (AISS) (Arnett, 1994). While PVQ uses a set of questions to outline personal traits and thereby motivational goals of individuals, CRT quantifies the reasoning style (deliberative or intuitive). Additionally, a measure from Arnett's sensation seeking study was used seeking both novel and intensive stimulation through two measures, namely, "Novelty" and "Intensity". The key explanatory variables in the analysis are derived from the PVQ, while the other surveys provide basis for the control variables.

Portrait Value Questionnaire

Each portrait in the Portrait Values questionnaire (PVQ) by Schwartz (2003) describes a person's goals, aspirations, or wishes that point implicitly to the importance of a single value type. By describing each person in terms of what is important to her and the goals and wishes she pursues, the portraits capture the person's values without explicitly identifying values as the topic of investigation. The PVQ was used in this study to identify value goals of subjects which likely influences their search and purchase behaviour online. According to Schwartz (2003) there are ten universal values that guide the principles of how people live their lives . The 21-item questionnaire (Figure 2.10 in the Appendix) asks subjects to evaluate the resemblance of a given statement to their own values. The strength of each value is determined by two or more questions that subjects rate on a 5-point Likert scale. Respondents differ systematically in their tendencies to report that certain values are more important to them than others. While some subjects report that most values are highly important, others use the middle of the scale and others tend to rate only a few values highly. Such differences in use of the response scale also appear in ratings of other persons as more or less similar to self in the PVQ. To retain accuracy of the value measurement when comparing individuals or groups, it is critical to correct for individual biases in use of the response scale. It essentially displays tradeoffs between relevant values that influence behavior and attitudes, so it is the relative importance of the ten values to an individual, that should be measured. In order to retain accuracy in the empirical analyses, we compare the relative importance of the ten values (indicated in Figure 2.1) to every individual's values, by mean centering. Mean centering is performed by subtracting the mean of an individual's response to all 21-items from each item. Thereafter the mean is computed for each value from the items that index it (Schwartz, 2003). As Figure 2.1 exhibits, human values are associated with inherent motivations such as self-enhancement, openness to change, self-transcendence and conservatism, which may lead to contrasting personalities. For example, *Self-direction* and *Conformity* represent opposing values as the former is motivated by openness to change, while the latter by conservatism.

Figure 2.1: Schwartz (2003) Theoretical model of relations among ten motivational types of values



Cognitive Reflection Test

Cognitive tasks are performed with two different forms of processing (Frederick, 2005). These are defined by Kahneman (Kahneman, 2003) as *system 1* and *system 2* (Stanovich and West, 2000). The cognitive-processes that occur by default are intuitive and spontaneous (system 1), while those tasks that require a more rule-based, analytic and deliberate process are defined as system 2. The system 1 and system 2 processes are defined in this paper according to the grading criteria included in Table 2.7 in the Appendix. Frederick (2005) evaluates the cognitive-processes to generate two cognitive reasoning styles: high (scoring 3 out of 3) and low (scoring 0 out of 3), with distinct differences in risk preference between these groups.

Arnett's Sensation Seeking Scale

Arnett Inventory of Sensation seeking (1994), can be used to evaluate how likely subjects are to seek new experiences and take risks to achieve it. It is a measure based on a questionnaire that includes 20 items focusing on intensity and novelty as components for sensation seeking (Questionnaire in Tables 2.8 and 2.9 included in the Appendix). Novelty refers to openness to experiences and intensity as to how intensively senses get simulated. The items consist of multiple choice questions in which subjects were required to rate on a 5-point likert scale how well a statement relates to them. Six additional questions were worded negatively and scaled reversely in order to alleviate any affirmation biases. The scaling was conducted with a total score and two subscales that measure novelty and intensity; higher the score, more the subject's personality coincides with the measured trait (Arnett, 1994).

2.2.2 Search experiment

Several days after successful completion of the pre-questionnaire, the invited students participated in the search experiment on site in a decision making laboratory, which included three sequential tasks (detailed instructions to participants are included in the Appendix). In task A they were asked to add their chosen phone to basket without consulting each other. The participants were subjected to five treatment variations through tasks B and C, for example, participants had the option of buying information on the most popular phones in Finland. For each task, there was an upper limit of 8 minutes to complete the task. The incentive scheme was designed in a way that each participant had a chance to win a reward of 1000 euros minus the listed price of their chosen phone. The participant would also receive the residual budget net of a 10% commission. After collection of the data, one participant and one of his/her three search tasks was randomly drawn. The person was rewarded with her/his chosen mobile phone + the residual of 1000 euros once the price of the phone at the website where the phone had been found was subtracted. The incentive scheme was designed uniformly across all treatments.

In tasks B and C, there were some exogenous/experimental variation in the incentive schemes. Details regarding the experimental variation in the incentive scheme can be found in the Appendix (table 2.11). Additionally in task C, all subjects were shown a randomly drawn smartphone among the 15 most popular smartphones sold in Finland. This randomly drawn smartphone belonged to the brand Huawei. Furthermore, subjects were informed that the Huawei smartphone would be shown to all other participants. The subjects were incentivized to purchase the Huawei phone with cash deductions applicable across several treatments in task C. The incentive was the following: the less the subject's purchase differed from the other participants, the higher the possible cash reward. As the sample size is relatively small, the treatment differences were not statistically different, so we do not focus on the analysis of treatment differences in this study.

2.2.3 Post-questionnaire

Following the search task, subjects were instructed to fill out an online questionnaire as the final step of the experiment. The post-questionnaire (included in Table 2.10 in the Appendix) collects information on the subjects' knowledge of online stores and choice of the purchased phone, so as to gauge their overall market awareness and product knowledge.

2.3 Descriptive results

2.3.1 Demographic characteristics

The dataset includes a rich set of user demographics and personal traits in addition to browsing and transaction information, that are used to estimate several search metrics. Demographic data was collected via the pre-questionnaire, some of which is presented as follows. We observe that birth years vary from 1965 to 1997 with an average age of 25 years. Students come from business disciplines majority from marketing (17), economics (14), and finance (13) and 60% of the participants are male.

2.3.2 Search and choice

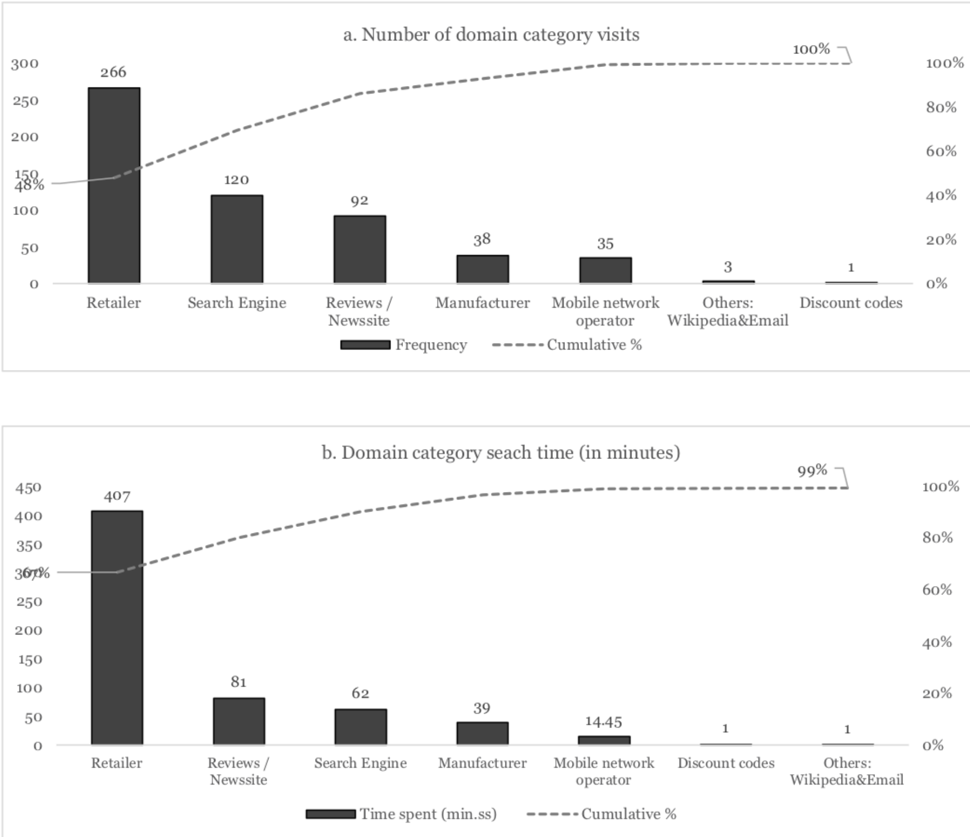
Search patterns provide insight into the nature of consumer awareness, brand recognition, and preference for some retailers over others. Figures 2.2 and 2.3 summarize search and purchase decisions across domains. Figure 2.2a,b exhibits the number of unique domain visits and search time for all tasks. Evidently, more than 50% of the visits were on retailer sites, followed by search engines with 21% of the total visits. The two biggest electronic retailers in Finland, Verkkokauppa and Gigantti were the most popular retail domains, while Google emerged as the most popular search engine used across all tasks. Apple was the only manufacturer associated with direct search on its own website; the other popular brands were searched strictly via retailers. Only 6% of the domain visits can be attributed to price comparison sites (such as Hintaseuranta and CNET). In line with literature, significant percentage of the participants visited only one domain prior to purchase, therefore search for mobile phones, much like cameras (Bronnenberg et al., 2016) or books (De Los Santos et al., 2012; De los Santos, 2018) is fairly limited. We further observe that the intensity of search diminishes over time. In general, subjects engage in more extensive search in task A compared to the later tasks as they learn about their match quality over time. Furthermore, incentives in task A were a close to perfect match to field incentives, more so than in some of the treatment conditions in B or C (see appendix for details on treatment conditions). For these two reasons we focus our analysis on tasks A and use task B and C for robustness checks.

Time spent collectively at Verkkokauppa and Gigantti account for 49% of the total search time. Search time spent on a domain is measured by taking the time difference between two subsequent URLs and amassing these differences to the respective domains.[4] Although subjects had 8 minutes to complete each task, approximately half of the population searched for less than two and a half minutes, while a third engaged in less than one and a half minutes of active search only. This indicates that opportunity cost of time is quite high for at least half the population; only 3% of the population searched for the whole 8 minutes, which implies that the artificial restriction is not binding and thus is likely not to influence our results.

Figure 2.3 details the number of purchases across domains. Unsurprisingly, the highest number of purchases were recorded in domains that were most searched. However, the price distribution across domains, as reported in Table 2.1, shows that the average spend for the most visited retailer platforms is relatively low. This may be due to the fact that retailers tend to offer a fairly large assortment with better deals and cheaper alternatives than individual manufacturers. Comparing all smartphone models that were added to cart, Apple as a brand

[4]Saunalahti is excluded due to the domain's technical properties that would provide an inaccurate measurement of the time spent in the domain.

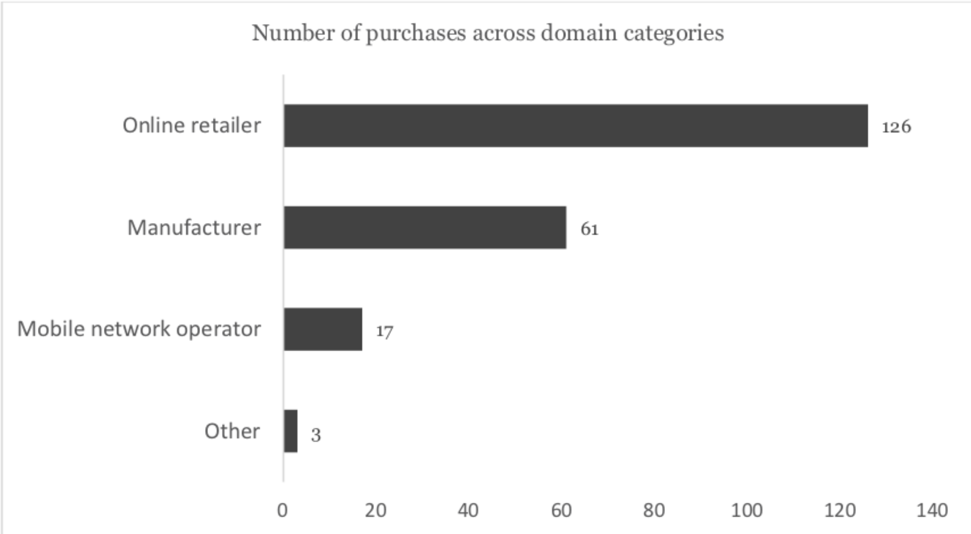
Figure 2.2: Domain search



has the highest mean price and the third highest number of purchases. This is contrary to a large body search literature where shoppers essentially sample a fixed number of stores and choose to buy the lowest priced alternative (Stigler, 1961; Burdett and Judd, 1983; Janssen and Moraga-GonzÃ¡lez, 2004). However, this evidence points to the existing Veblen effects that are significant in luxury products such as smartphones, as identified by prior research (Leibenstein, 1950; Bagwell and Bernheim, 1996; Vigneron and Johnson, 2004; Kastanakis and Balabanis, 2014).

Figure 2.4 exhibits brand performance across all tasks. Apple is evidently the most popular brand as the highest number of page visits, purchases and average spend have been

Figure 2.3: Purchases across domain categories



recorded with this brand. Apple is followed by Samsung and Huawei, although Huawei is a comparatively lower priced alternative. This represents a clear dichotomy in what the subjects value and how it relates to distinct traits in their personality. Apple was the most popular brand with 47% of the total purchases, followed by Samsung accounting for 25% and Huawei for 10%. It is to be noted here that the majority of subjects chose the two of most high priced brands as observed from the brand price distributions reported in Table 2.2. This presents a similar picture as observed across domain search, contrary to the classical search model prediction. This indicates conspicuous motives fueled either by conformism or hedonism, that we study in detail in the following section. Table 2.3 shows no significant difference in price search between task A and the following task B.

Figure 2.5a-2.5b exhibits distribution of search queries and the most popular query names used by subjects. Search queries are captured in the data via URLs and video recordings. The key take-away remains consistent, in that, subjects do not necessarily search across majority of the alternatives available. Data shows relatively limited active search as the most popular query names happen to be most commonly chosen phones across all three tasks.

Table 2.1: Price set across chosen domains

Domain	Mean Price	Std. Deviation Price	Number of buys
Verkkokauppa	609,54	246,33	80
Gigantti	729,20	190,16	49
Apple	956,83	113,27	17
Elisa	792,64	169,78	11
Amazon.de	524,87	242,52	9
Amazon.com	573,64	220,27	8
Power	302,85	336,05	8
Sonera	850,67	66,46	6
Oneplus	345,00	-	4
Amazon.co.uk	356,64	246,55	3
Ebay	882,71	2,53	3
CDON.com	592,95	349,31	2
DNA	594,00	360,62	2
Expert	749,00	-	2
Knaitek	662,50	173,24	2
MyTrendyPhone	825,00	-	1

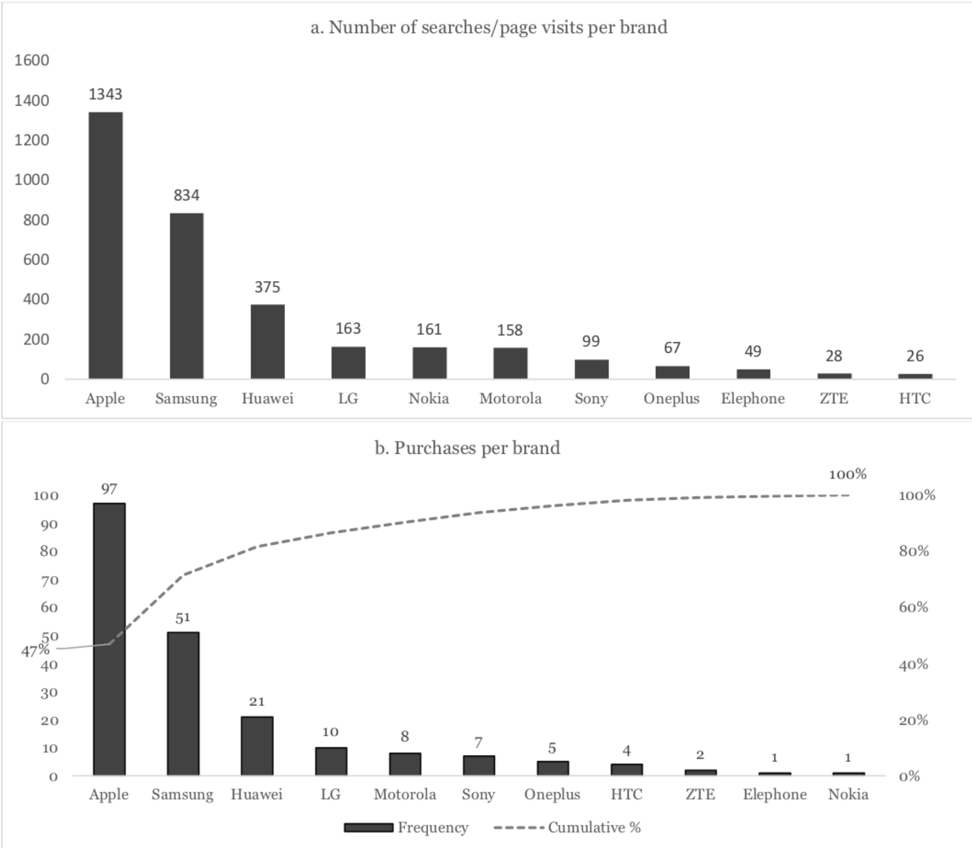
Figure 2.6 shows the search sequences based on domain changes for all tasks. It details the search paths of subjects across domain categories prior to adding their chosen phone to basket, P. While 43% percent of the population went directly to the retailer's domain, 19% were directed to a retailer's domain via Google, where they eventually made the purchase. This is consistent with the key finding of search being fairly limited from a large perecentage of the sample.

Figure 2.7a1,a2 shows the evolution of brand views in time deciles during search, while figure 2.7b1,b2 shows the evolution of model views. In order to normalize the length of search activity across subjects, we divide each search session into ten equal parts following Bronnenberg et al. (2016), where individual search deciles, d are defined as,

$$d(t, N_j) = ceil\left(\frac{10 * (t - r(0,1))}{N_j - 1}\right) \tag{2.1}$$

where $t = 1, ..., N_j - 1$ is the number of searches made by subject j, with choice $t = N_j$, r(0,1) is a random uniform number on [0,1] and the *ceil* operator rounds up to the next whole number. The x-axis are time deciles, divided into ten parts of the total time used to complete

Figure 2.4: Search and purchases across brands



the task.

The graphs in figure 2.7 are plotted based on the characters appearing in the URLs and site title information which is collected manually by checking the URLs and video recordings of the entire search sessions up to the point of purchase for each subject for each of the tasks. They display the link between search and choice. Figure 2.7a1 represents the number of unique brands viewed by subjects for each task, where more than half of the subject pool visited only one brand prior to purchase. Figure 2.7b1 shows the number of unique models that were sampled by subjects across all tasks, where 81% of the total user search activity accounts to sampling between one and three models. Figures 2.7a2 and 2.7b2 display the

Figure 2.5: Search queries

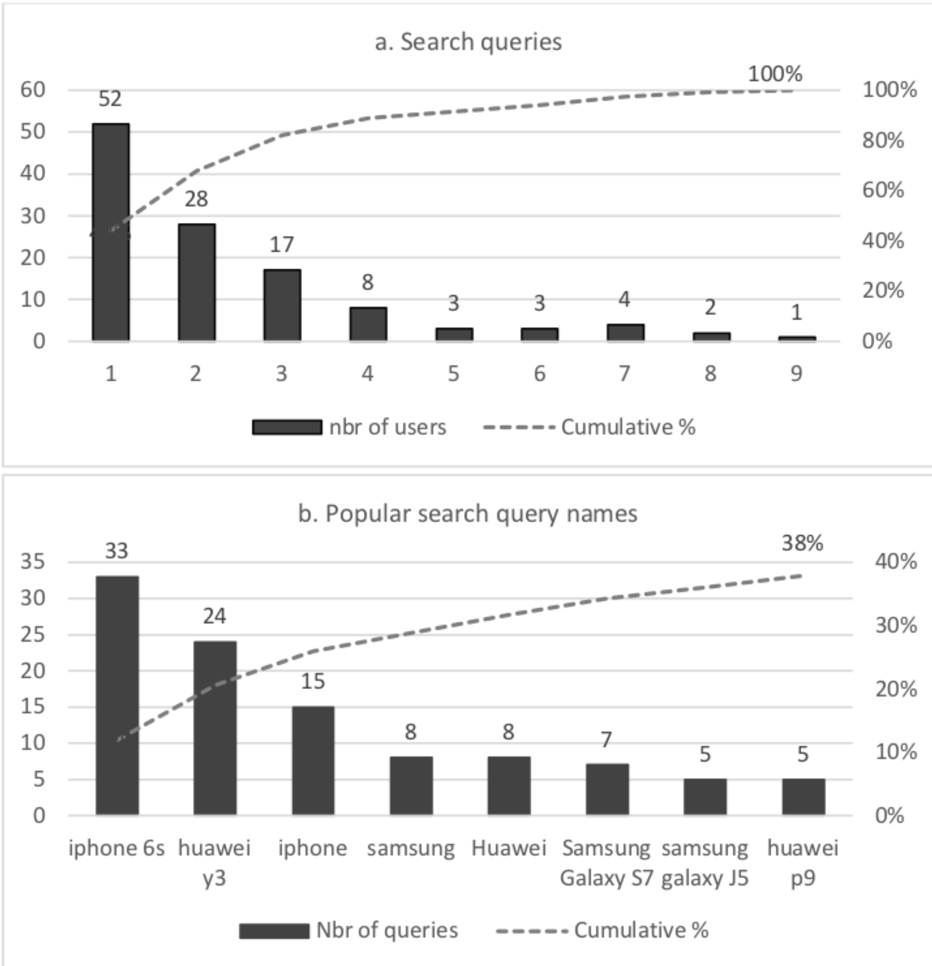


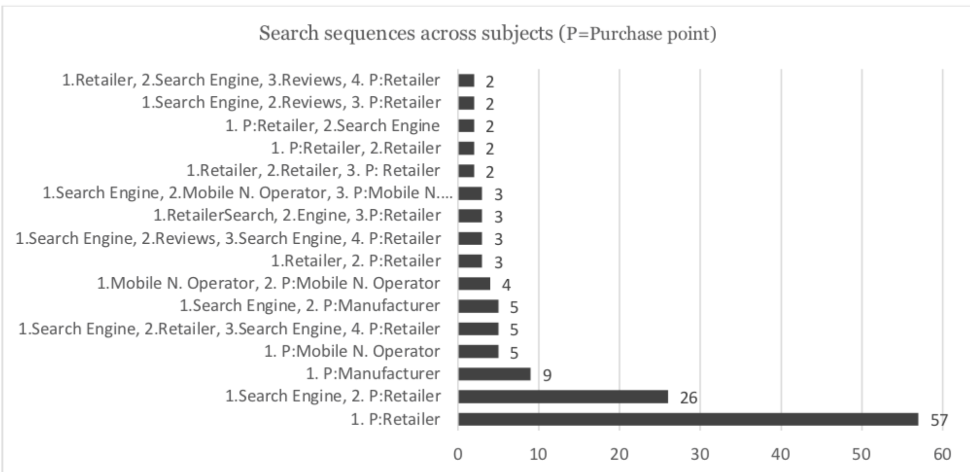
Table 2.2: Price set across chosen brands

Brand	Mean Price	Std. Deviation Price	Number of buys
Apple	792,02	144,45	97
Samsung	704,86	201,83	51
Huawei	495,19	218,31	21
LG	617,18	167,95	10
Motorola	529,65	204,14	8
Nokia	122,00	41,04	7
Oneplus	339,80	11,63	5
Sony	513,40	104,21	4
ZTE	129,90	-	2
Elephone	244,99	-	1
HTC	731,50	-	1

Table 2.3: Price set across tasks

Task	Mean Price	Std. Deviation Price	Number of buys
TASK A	667,87	238,32	69
TASK B	663,74	259,71	69

Figure 2.6: Search patterns



evolution of brand and model searches respectively. The black line counts the total number of views of the purchased brand (figure 2.7a2) and model (figure 2.7b2) in a particular time decile. The grey line counts the total number of views of other brands (figure 2.7a2) and models (figure 2.7b2) in a particular time decile. Brand information is separated into two different categories which are *Purchased Brand* and *Other Brand*. The former consists of brand information only of the product added to basket and the latter represents all other brands viewed. It can be observed from figure 2.7a2 that search activity after the third decile is limited to the brand from which subjects have made their purchase, that is, the subjects were able to identify their most preferred brand quite early on during the length of their search. On the other hand, figure 2.7b2 shows that, subjects increasingly discover the most preferred model later in their search journey. This may imply that the range of preferred product attributes narrows as search proceeds for the majority of subjects. After the seventh search decile, a larger proportion of the population were observed to view the models that they eventually purchased. Furthermore, 71% of the population chose to visit only one domain with the same brand-model combination where they finally made their purchase. These views combined provide motivation to test if late search is a better predictor of the chosen alternative. We empirically examine this hypothesis further by overlaying personal traits in Section 4.

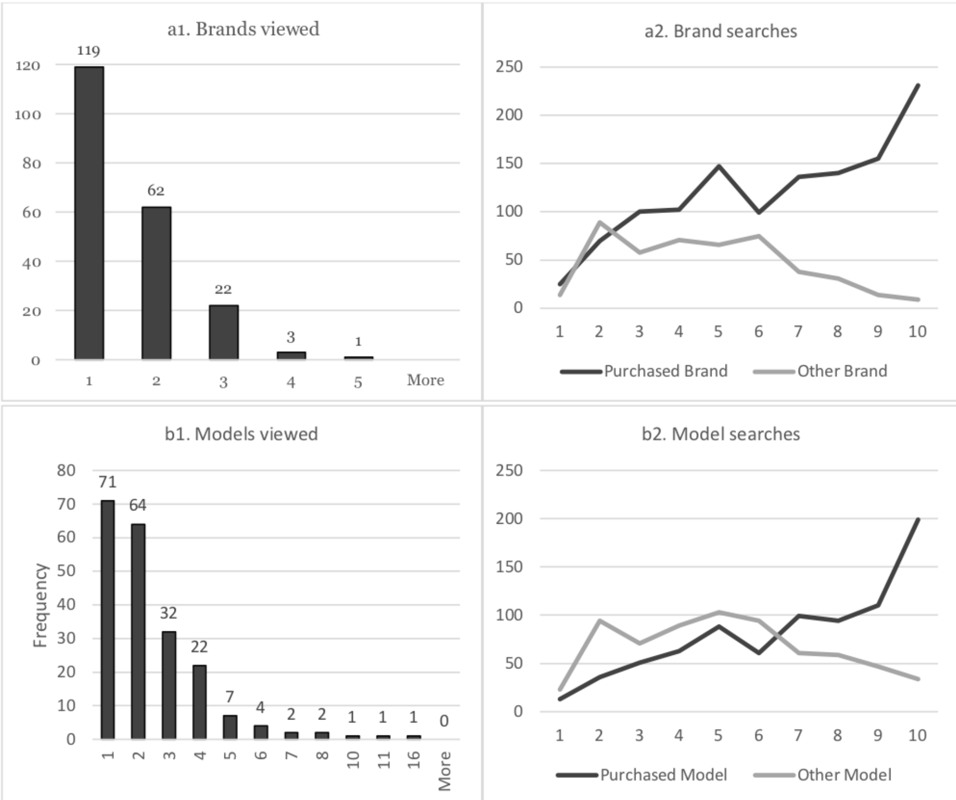
2.4 Empirical Analysis: Consumer types, search and choice

In this section we show several sources of empirical evidence that concludes search is informative of choice and consumer value orientations have significant correlation with online search behaviour. Linking personality traits to demographics and ultimately browsing data, a rich sample was constructed to study the effects of consumer value systems on search and choice.

2.4.1 Do personal traits impact search?

In this study, search is quantified by its length as well as depth. Length is measured by the time devoted to search and depth, or the rigour and specificity of search, is measured by the number of queries. The number of unique domains searched provides an alternative measure of how long subjects engaged in search. This latter analysis is provided in the Appendix as an additional robustness check. As mentioned in Section 2, personal traits are derived via mean centering subjects' responses of the PVQ (Schwartz 2002). Tables 2.4 and 2.12 exhibit the OLS estimates of the number of unique search queries and number of unique domains

Figure 2.7: Convergence of search



searched, respectively, on personality traits. We present OLS estimates for each of the three tasks in columns (1)-(3), and all tasks combined in column (4). However, we focus on the results from task A which reflects participants natural search behaviour closely. We chose three relevant personal traits illustrated in Schwartz (2003) namely, conformism, hedonism and self direction, based on Pearson product-moment correlations. Surprisingly, the values that represent opposing motivational goals (Figure 2.1) tend to show similar associations with search.

One of the key findings is that, conformity is significantly positively associated with both search variables. This is intuitive, as shoppers that follow trends and adhere to social expectations are likely to search more extensively. Conformity relates to conservatism as per

Schwartz (2003) which tends to exhibit a higher degree of herding behaviour. This typically motivates potential buyers to search longer and more rigorously, such that their chosen option does not deviate from what is popular. Additional evidence is presented to this claim via one of the treatments, where, subjects were shown a popular phone in Finland at the time in Task C, namely Huawei. A binary variable that takes value 1 if the chosen phone in Task C was Huawei, 0 otherwise is included that interacts with the set of PVQ values (Hedonism, Conformity and Self Direction). We find subjects choosing Huawei that value conformity highly are negatively associated with both search variables[5]. This points to the incentives of information search: once a conformist becomes aware of the most popular product with certainty, she will choose this option without further search for possible alternatives, as the marginal value of an additional search at this point is fairly low. Therefore, conformists primarily engage in extensive search with the aim of finding most popular alternatives within their social cohort.

Contrary to conformism, self direction relates to independent action, snobbism and openness to change (Schwartz, 2003). And yet, on both search variables, the OLS estimates of self direction are highly positive, which is directionally similar to the effect of conformism. An interpretation of this result may be, that self-direction leads to more self-initiated search behaviour, driven by inherent curiosity and willingness to learn. Finally, hedonism, that primarily relates to self-interest and independent action too has a highly significant positive association with the extent of search. Hedonistic shoppers may enjoy search largely due to non-functional motives, self-gratification or simply to learn about newest trends and novelty features (Childers et al., 2001)[6]. Moreover, the aim of this type of search is not necessarily to obtain the lowest possible price, although there is a consensus in economic literature that increased search leads to the cheapest alternative (Stigler, 1961; Rothschild, 1974; Baye and Morgan, 2001; Hong and Shum, 2006; Ellison and Ellison, 2009). Our findings show that search time has no impact on price at any meaningful level of statistical significance. This raises relevant questions on the motives of search, for example, subjects with a higher degree of sensation seeking behaviour chose smartphones, that are, on average, 44% higher priced (supporting results in Table 2.14 included in the Appendix). This may be driven by the

[5]Complete results controlling for Huawei are included in the Tables 2.13 in the Appendix

[6]OLS estimates in Task C (Conformity, Hedonosim in Table 2.12 and CRT in Tables 2.12 and 2.4) are not significant at a meaningful level, which may be due to the chosen treatment as it would expectedly alter subjects' natural search behavior.

novelty aspect of sensation seeking behaviour which values niche features or specifications and are positively correlated with hedonistic motivations. While price can be a legitimate motivation to search, other product attributes or behavioural constructs and preferences make search equally as necessary for online shoppers today.

Table 2.4: OLS: Unique search queries on personal characteristics

	Dependent variable: Number of search queries			
	TaskA	TaskB	TaskC	All tasks
	(1)	(2)	(3)	(4)
Conformity	0.776**	0.807**	0.740**	0.774***
	(0.295)	(0.364)	(0.279)	(0.248)
Hedonism	0.725**	1.374***	0.873***	0.991**
	(0.345)	(0.427)	(0.328)	(0.402)
CRTIntuitive	−1.050***	−0.911**	−0.481	−0.814***
	(0.362)	(0.448)	(0.344)	(0.302)
SelfDirection	0.870*	1.429**	1.309***	1.203**
	(0.449)	(0.555)	(0.426)	(0.546)
Patience	0.038	−0.383	−0.507	−0.284
	(0.415)	(0.514)	(0.394)	(0.363)
Gender	0.194	−0.407	−0.138	−0.117
	(0.553)	(0.684)	(0.524)	(0.537)
Constant	2.788***	2.678***	2.274***	2.580***
	(0.542)	(0.671)	(0.514)	(0.645)
Observations	66	66	66	198
R^2	0.274	0.275	0.267	0.251
Adjusted R^2	0.200	0.201	0.193	0.227
Residual Std. Error (df = 59)	2.133	2.639	2.023	2.232
F Statistic (df = 6; 59)	3.714***	3.721***	3.586***	10.665*** (df = 6; 191)
Note:				*p<0.1; **p<0.05; ***p<0.01
Discussion based on estimates in Task A only				

2.4.2 Is search predictive of choice across consumer types?

We further investigate the empirical relationship between the searched and chosen smartphones, given rich data on product attributes searched and detailed behavioural values. Intuitively, search behaviour of potential buyers should be informative of their final choice.

We test if certain human values exhibit notably different relationship between search and choice. During the experiment, subjects search through several sources of information about smartphone features online, such as, price, memory, operating system, shipping costs etc. Let x_{ijk} be the level of the k_{th} attribute of smartphone model i, searched by shopper, j and y_{jk} be the attribute level of the mobile phone added to basket. Then we estimate the following:

$$y_{jk} = \alpha + \beta \bar{x}_{jk} + \epsilon_{jk} \tag{2.2}$$

where, $\bar{x}_{jk}$ = mean searched attribute level.
For our analysis, we consider three continuous product attributes searched and chosen during Task A, namely, price, memory and shipping cost. The results are reported in Table 2.5 columns 1-3. We find that across all attributes the mean searched attribute level has a strong association with the chosen attribute level, homogeneously across all three attributes at 1% level of significance. Searched price shows a very strong association with the chosen attribute levels, followed by phone memory. Typically, potential buyers acquire information on product attributes via search to determine their own match quality and they stop search either when they find the best match in their consideration set (Stigler, 1961) or when the marginal benefit of an additional search exceeds the marginal cost (McCall, 1970; Burdett and Judd, 1983). In either case, search is shown to be informative of the final choice, which is in line with literature (Bronnenberg et al., 2016).

Furthermore, we hypothesize that as search evolves, attribute levels become increasingly more predictive of the chosen levels. Figure 2.7 in Section 3 motivates this hypothesis, which displays convergence of search to choice for smartphone models, however not brands. In order to test whether late search is predictive of choice, we use recency weighted mean (RWM) searched attribute levels as regressors in Equation (2), where $\bar{x}_{jk}$ is replaced by $\bar{x'}_{jk}$

$$\bar{x'}_{jk} = \sum_{t} (1 - e^{\lambda d}) * x_{jkt}$$

where, d is calculated per Equation (2.1), t is time and λ =0.5[7].

Table 2.5 columns 4-6 report the OLS estimates of RWM searched attribute levels

[7]We used the exponential distribution to model recency weights across search deciles with varying values of λ which do not alter the OLS estimates dramatically, therefore 0.5 was chosen arbitrarily

on chosen attribute levels in Task A. Interestingly, for most of the attributes this does not significantly improve performance of the model as observed in the respective R^2. The RWM coefficients remain directionally comparable, but magnitudinally smaller than the mean search attribute levels. We observe the largest association (47%) between RWM searched price and the final price chosen, which was followed by 16% for shipping costs and 9% for memory[8]. Overall, this specification implies that searched product attribute levels are a strong predictor of chosen attribute levels, which in turn, concurs with observations in the descriptive results of the chosen smartphones as well.

Generally speaking, the empirical relation between search and choice is strong and in favour of search being predictive of choice. In order to study the effects of individual personal traits on search and choice, we include interaction effects in Equation (2.2). Table 2.6 columns 1-3 report the results of the following model:

$$y_{jk} = \alpha + \beta \bar{x}_{jk} + \gamma_i V_q \bar{x}_{jk} + \epsilon_{jk} \tag{2.3}$$

where, $q = [Conformity,\ \ Hedonism,\ \ Self\ direction,\ \ CRT]$

Replacing the $\bar{x}_{jk}$ or mean searched attribute level by $\bar{x'}_{jk}$ or recency-weighted mean searched attribute level, we arrive at the estimates reported in Table 2.6 columns 4-6. Including interaction terms with personal traits and searched attribute levels in the baseline model leads to two interesting findings. Firstly, mean search and RWM search price for conformists are statistically insignificant. One interpretation consistent with this finding may be that conformists engage in a fixed-sample search strategy with recall. At the outset, they have a high degree of awareness of the popular brands and models, which is ultimately what they choose, to be on par with social trends. However, as we have seen earlier in Tables 2.4 and 2.12, they do have a high propensity to search, so as to enrich their information sets regarding the most relevant alternatives and corresponding specifications at any given point in time. It may be that these shoppers do not have a high marginal utility from an incremental price search, as price is not the only determinant of choice, in this case. RWM search across other attributes, such as shipping cost and memory have a significant positive effect on choice, as this is a price characteristic non-observable to peers when the phone is being used.

[8]Phone memory is homogeneously measured in gigabytes in our dataset

Secondly, subjects with hedonistic tendencies exhibit a negative relationship between searched and chosen attribute levels, that is, as recently searched attribute levels go down, chosen attribute level of the purchased good goes up. Hedonism can be associated with festive, ludic or even epicurean shopping; it is related to fun rather than task completion, and thus reflects the experiential side of shopping, comprising of pleasure, curiosity, fantasy and fun, rather than task-oriented and rational (Hirschman and Holbrook, 1982; Griffin et al., 2000; Chaudhuri et al., 2010). As shoppers with hedonistic tendencies search online with the goal of recreation rather than necessity, it is likely that their search paths are not entirely representative of choice.

2.4.3 Probability of choice as a function of search

In Section 3, figure 2.7a1 and 2.7b1 exhibit the density of brands and models viewed, respectively with 34% viewing only one model prior to purchase. However, 58% of the subject pool searched between two to four models, which leads to the following hypothesis: repeat search or "repeated fixations to the same item" (Reutskaja et al., 2011) is associated with higher likelihood of a successful purchase. Furthermore, we test if there exists a positive association between the extent of fixations or time spent on contemplating one model and the probability of purchase of that model. In order to investigate these two hypotheses, we chose some of the most popular models added to basket across all tasks.

Let the probability of adding model i to basket at n_{th} visit be P_{ni}. As the unobserved factors may be correlated over choices, a fixed effects binary choice specification is used to estimate the probability of purchase as a function of 1) number of views; 2) number of repeat views. The weights of the control variables obtained from a standard logistic regression are used to calculate P_{ni} for each model, as displayed in Figures 2.8 and 2.9, where 2.8 exhibits the probability of choice as a function of overall search length and 2.9 exhibits the probability of choice as a function of repeat views of the particular model chosen. The regression accounts for consumer fixed effects to account for differences in tasks as well as the search set size. Majority of the chosen smartphones were from Apple and Samsung, where we observe a monotonically increasing probability of being added to basket as overall search increases. Furthermore, an increase in repeat views of the chosen model is consistent with an increase in the probability of purchase as well. For the sake of robustness, this analysis is performed with other randomly chosen models from the above mentioned brands, as well as other brands

Table 2.5: 1-3: Effects of mean searched attribute levels on choice 4-6: Effects of recency-weighted mean (RWM) searched attribute levels on choice

	Dependent variable:					
	Price	Shipping cost	Memory	Price	Shipping cost	Memory
	(1)	(2)	(3)	(4)	(5)	(6)
Mean search price	0.969***					
	(0.066)					
Mean search shipping		0.193***				
		(0.045)				
Mean search memory			0.424***			
			(0.102)			
RWM search price				0.471***		
				(0.046)		
RWM search shipping					0.160***	
					(0.048)	
RWM search memory						0.090*
						(0.050)
Constant	0.034	4.567***	0.998***	1.558***	5.228***	1.523***
	(0.185)	(0.472)	(0.154)	(0.116)	(0.406)	(0.064)
Observations	216	230	177	216	230	177
R^2	0.502	0.074	0.090	0.331	0.047	0.018
Adjusted R^2	0.500	0.070	0.085	0.328	0.043	0.012
Residual Std. Error	0.262 (df = 214)	4.517 (df = 228)	0.286 (df = 175)	0.304 (df = 214)	4.584 (df = 228)	0.297 (df = 175)
F Statistic	216.000*** (df = 1; 214)	18.287*** (df = 1; 228)	17.257*** (df = 1; 175)	106.002*** (df = 1; 214)	11.170*** (df = 1; 228)	3.191* (df = 1; 175)

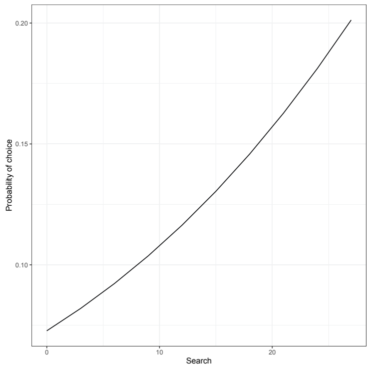
Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 2.6: 1-3: Mean searched attribute levels on choice w/ personality traits 4-6: RWM searched attribute levels on choice w/ personality traits

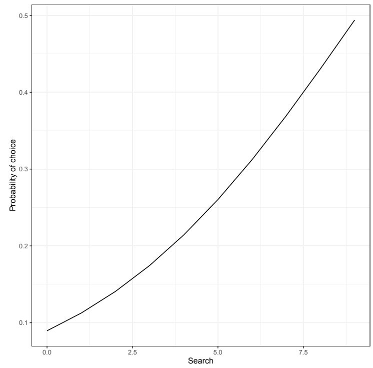
	Dependent variable:					
	Price	Shipping cost	Memory	Price	Shipping cost	Memory
	(1)	(2)	(3)	(4)	(5)	(6)
Mean search price	1.643^{***}					
	(0.174)					
Mean search shipping		0.291^{***}				
		(0.053)				
Mean search memory			0.606^{***}			
			(0.126)			
RWM search price				0.147^{*}		
				(0.086)		
RWM search shipping					0.161^{***}	
					(0.056)	
RWM search memory						0.090
						(0.068)
x_j : Conformity	0.026	0.451^{***}	0.510^{***}	−0.118	0.342^{***}	0.160^{**}
	(0.151)	(0.058)	(0.117)	(0.071)	(0.063)	(0.065)
x_j : Hedonism	0.301	-0.310^{***}	0.826^{***}	-0.375^{***}	-0.372^{***}	0.122
	(0.232)	(0.057)	(0.140)	(0.089)	(0.065)	(0.076)
x_j : Self Direction	0.615^{***}	0.507^{***}	0.588^{***}	0.564^{***}	0.232^{*}	0.174
	(0.262)	(0.126)	(0.193)	(0.141)	(0.127)	(0.119)
Constant	1.912^{***}	4.103^{***}	0.671^{***}	2.343^{***}	5.242^{***}	1.489^{***}
	(0.490)	(0.474)	(0.190)	(0.222)	(0.413)	(0.087)
Observations	210	224	171	210	224	171
R^2	0.611	0.349	0.369	0.500	0.248	0.183
Adjusted R^2	0.597	0.328	0.342	0.483	0.224	0.148
Residual Std. Error	0.239 (df = 202)	3.873 (df = 216)	0.245 (df = 163)	0.270 (df = 202)	4.163 (df = 216)	0.279 (df = 163)
F Statistic	45.236^{***} (df = 7; 202)	16.576^{***} (df = 7; 216)	13.627^{***} (df = 7; 163)	28.891^{***} (df = 7; 202)	10.200^{***} (df = 7; 216)	5.220^{***} (df = 7; 163)
Note:						*p<0.1; **p<0.05; ***p<0.01

in the dataset. Not only is the empirical relation between choosing a phone that has been viewed multiple times homogeneously valid in the data, but we also find positive relation between the extent of search, in general, and the probability of choice.

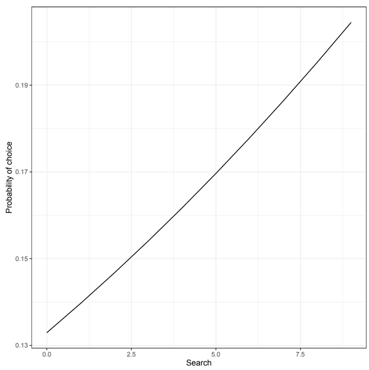
Figure 2.8: Probability of choice as a function of search



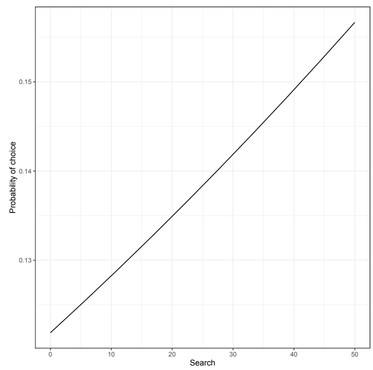
(a) Apple iPhone 6s 64GB



(b) Samsung Galaxy S7 Edge 32GB



(c) Apple iPhone 6s 16GB



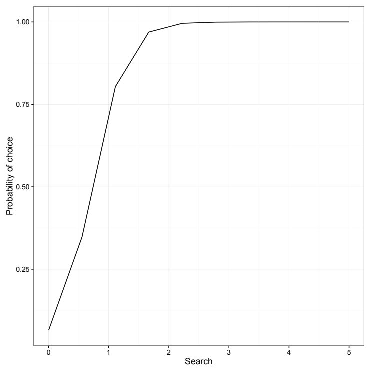
(d) Samsung Galaxy S7 32GB

2.5 Discussion

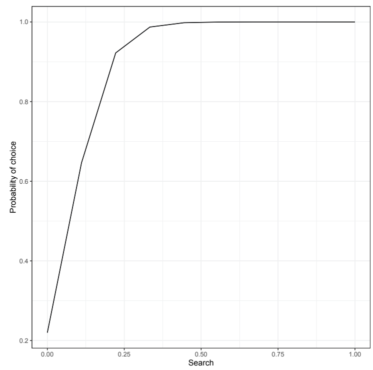
2.5.1 Search

This paper sheds light on the behavioural implications of consumer types on search and choice. In general, we find search to be fairly limited across domains, where most visits were restricted to retailer websites and search engines. Interestingly, only a small fraction of the

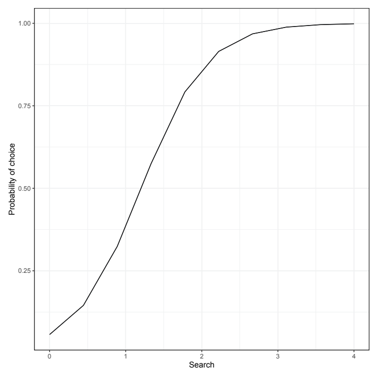
Figure 2.9: Probability of choice as a function of repeat search



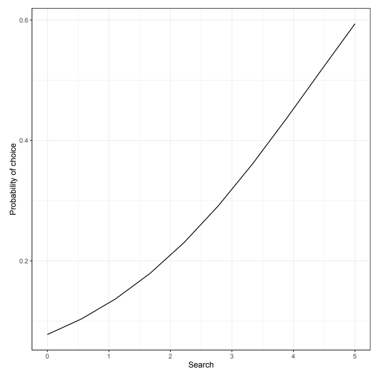
(a) Apple iPhone 6s 64GB



(b) Samsung Galaxy S7 Edge 32GB



(c) Apple iPhone 6s 16GB



(d) Samsung Galaxy S7 32GB

population visited price comparison sites. Detailed browsing data, shopper demographics and personality traits together provide rich information on consumer heterogeneity. We show that conformism, hedonism and self direction, which clearly represent contrasting consumer motivations and personalities, have a directionally similar relationship with information search for smartphones. Our findings point to the conspicuous motives of shoppers that stem from either bandwagon effects or snobbism, but ultimately lead to comparable outcomes in terms of search for smartphones. It can be observed that self direction, as a trait, is strongly associated with extensive search, both in terms of length and depth, which stems from self-driven behaviour. As shown by Schwartz (2003), self direction characterizes curiosity and openness to change, hence justifying the high intensity of search as observed in this study.

Hedonistic and novelty seeking behaviour too are associated with extensive search, where the objective is to primarily seek out "new and discrepant" information (Hirschman, 1980), that may be useful in the long run. Although, comparable in terms of association with search patterns, conformists are motivated by radically different goals compared to self-direction and hedonism. Increased search leads to potentially discovering the most popular alternative within a social cohort an individual belongs to, as there is are clear bandawagon effects in play.

2.5.2 Choice

Complete search paths of the subjects are mapped into choice, overlaying personality traits to study convergence of search. Overall, we find time spent on contemplating an option is predictive of choosing that option, with hedonistic shoppers being the only exceptions. This may stem from an inherent hedonistic disposition which may lead to the pursuance of sensory gratification rather than efficiency (Hirschman and Holbrook, 1982; Wolfinbarger and Gilly, 2001; Wan, 2011; Scarpi, 2012). Additionally, mean searched price levels are shown to be highly informative of the chosen price levels. While late search also exhibits a high degree of association with choice, we find no strong evidence that late searches are necessarily a better predictor of choice. The strong association between search and choice supports a number of recent works that aims to establish a link between the two, in demand side models (Bronnenberg et al., 2016; Honka and Chintagunta, 2017; Kim et al., 2017; Chen and Yao, 2017). This primarily suggests that consumers' search and choice decisions are based on the same utility function, contrary to consideration set literature that assumes separate functions for consideration and choice (Bronnenberg and Vanhonacker, 1996; Moe, 2006). We also find evidence of search with recall (Koulayev, 2014; Santos et al., 2017) wherein a higher purchase likelihood is associated with increasing product views.

2.5.3 Managerial implications

This study takes a deeper look at the heterogeneity of online search for a highly differentiated product across varied consumer types, which allows retailers and marketeers to design tools that can ultimately influence purchase decisions. As depicted in Figures 2.8 and 2.9, increased search increased search efforts on a specific alternative is associated with higher likelihood of purchasing that alternative. Therefore, it is ideal to keep shoppers engaged on retailer or manufacturer websites for a longer period of time. This may be done via improving search engine algorithms, employing chat-bots or showing personalized recommendation that help present relevant information to potential buyers and in turn, enlarge their consideration sets. It is important to note however, that search quality matters. Information search enables individuals to map their preferences to product attributes, therefore it is intuitive that search converges to choice. It is critical for marketeers to study this path of convergence in order to show product recommendations at specific points of the search path where the probability of purchase is the highest.

Tables 2.5, 2.6 and Figure 2.7 summarize search being predictive of choice, generally speaking. This poses an interesting challenge for retailers showing personalized recommendations to potential buyers: the trade-off between relevance and price of the recommended good. Based on specific consumer types combined with search histories, retailers must recommend products that on one hand are sufficiently relevant for shoppers to click on and continue their search, on the other hand maximizes firm earnings.

Furthermore, some consumer types search online largely for recreational motives, therefore, their expectations and preferences for interactive shopping may differ from those held in the physical retail shopping environment for identical products. Improving the aesthetics of online platforms to improve user experience is necessary to target this specific consumer type. Similarly, focusing on novelty features of a product over and above standard set of attributes increases search, which in turn may improve the likelihood of purchase.

References

Aesthetics and the online shopping environment: Understanding consumer responses. *Journal of Retailing*, 87(1):46 – 58, 2011.

Jeffrey Arnett. Sensation seeking: A new conceptualization and a new scale. *Personality and Individual Differences*, 16(2):289 – 296, 1994.

Laurie Simon Bagwell and B. Douglas Bernheim. Veblen effects in a theory of conspicuous consumption. *American Economic Review*, 86(3):349–73, 1996.

Michael R. Baye and John Morgan. Information gatekeepers on the internet and the competitiveness of homogeneous product markets. *American Economic Review*, 91(3):454–474, June 2001.

Bart J. Bronnenberg and Wilfried R. Vanhonacker. Limited choice sets, local price response, and implied measures of price competition. *Journal of Marketing Research*, 33(2):163–173, 1996.

Bart J. Bronnenberg, Jun B. Kim, and Carl F. Mela. Zooming in on choice: How do consumers search for cameras online? *Marketing Science*, 35(5):693–712, 2016.

Kenneth Burdett and Kenneth L. Judd. Equilibrium price dispersion. *Econometrica*, 51(4): 955–969, 1983.

Arjun Chaudhuri, Khaled Aboulnasr, and Mark Ligas. Emotional responses on initial exposure to a hedonic or utilitarian description of a radical innovation. *Journal of Marketing Theory and Practice*, 18(4):339–359, 2010.

Yuxin Chen and Song Yao. Sequential search with refinement: Model and application with click-stream data. *Management Science*, 63(12):4345–4365, 2017.

Terry L. Childers, Christopher L. Carr, Joann Peck, and Stephen Carson. Hedonic and utilitarian motivations for online retail shopping behavior. *Journal of Retailing*, 77(4):511 – 535, 2001.

Giacomo Corneo and Olivier Jeanne. Conspicuous consumption, snobbism and conformism. *Journal of Public Economics*, 66(1):55 – 71, 1997.

Babur De los Santos. Consumer search on the internet. *International Journal of Industrial Organization*, 58(C):66–105, 2018.

Babur De Los Santos, Ali HortaÃ§su, and Matthijs R. Wildenbeest. Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, 102(6):2955–80, May 2012.

Caroline Josephine Doran. The role of personal values in fair trade consumption. *Journal of Business Ethics*, 84(4):549–563, 2009.

Glenn Ellison and Sara Fisher Ellison. Search, obfuscation, and price elasticities on the internet. *Econometrica*, 77(2):427–452, 2009.

Shane Frederick. Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4):25–42, December 2005.

Mitch Griffin, Barry Babin, and Doan Modianos. Shopping values of russian consumers: The impact of habituation in a developing economy. *Journal of Retailing - J RETAIL*, 76: 33–52, 03 2000.

Elizabeth C. Hirschman. Innovativeness, Novelty Seeking, and Consumer Creativity. *Journal of Consumer Research*, 7(3):283–295, 12 1980.

Elizabeth C. Hirschman and Morris B. Holbrook. Hedonic consumption: Emerging concepts, methods and propositions. *Journal of Marketing*, 46(3):92–101, 1982.

Han Hong and Matthew Shum. Using price distributions to estimate search costs. *The RAND Journal of Economics*, 37(2):257–275, 2006.

Elisabeth Honka and Pradeep Chintagunta. Simultaneous or sequential? search strategies in the u.s. auto insurance industry. *Marketing Science*, 36(1):21–42, 2017.

Peng Huang, Nicholas H. Lurie, and Sabyasachi Mitra. Searching for experience on the web: An empirical examination of consumer behavior for search and experience goods. *Journal of Marketing*, 73(2):55–69, 2009.

Maarten C. W. Janssen and JosÃ© Luis Moraga-GonzÃ¡lez. Strategic pricing, consumer search and the number of firms. *The Review of Economic Studies*, 71(4):1089–1118, 2004.

Eric J. Johnson, Wendy W. Moe, Peter S. Fader, Steven Bellman, and Gerald L. Lohse. On the depth and dynamics of online search behavior. *Management Science*, 50(3):299–308, 2004.

Daniel Kahneman. Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5):1449–1475, December 2003.

Wagner Kamakura and Thomas Novak. Value-system segmentation: Exploring the meaning of LOV. *Journal of Consumer Research*, 19:119–32, 02 1992.

Minas N. Kastanakis and George Balabanis. Explaining variation in conspicuous luxury consumption: An individual differences' perspective. *Journal of Business Research*, 67 (10):2147 – 2154, 2014.

Jun B. Kim, Paulo Albuquerque, and Bart J. Bronnenberg. Online demand under limited consumer search. *Marketing Science*, 29(6):1001–1023, 2010.

Jun B. Kim, Paulo Albuquerque, and Bart J. Bronnenberg. The probit choice model under sequential search with an application to online retailing. *Management Science*, 63(11): 3911–3929, 2017.

Sergei Koulayev. Search for differentiated products: identification and estimation. *The RAND Journal of Economics*, 45(3):553–575, 2014.

H. Leibenstein. Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand. *The Quarterly Journal of Economics*, 64(2):183–207, 05 1950.

J. J. McCall. Economics of Information and Job Search*. *The Quarterly Journal of Economics*, 84(1):113–126, 02 1970.

Wendy W. Moe. An empirical two-stage choice model with varying decision rules applied to internet clickstream data. *Journal of Marketing Research*, 43(4):680–692, 2006.

Young-Hoon Park and Peter S. Fader. Modeling browsing behavior at multiple websites. *Marketing Science*, 23(3):280â303, August 2004.

Elena Reutskaja, Rosemarie Nagel, Colin F. Camerer, and Antonio Rangel. Search dynamics in consumer choice under time pressure: An eye-tracking study. *American Economic Review*, 101(2):900–926, April 2011.

Michael Rothschild. Searching for the lowest price when the distribution of prices is unknown. *Journal of Political Economy*, 82(4):689–711, 1974.

Babur De Los Santos, Ali HortaÃÂ§su, and Matthijs R. Wildenbeest. Search with learning for differentiated products: Evidence from e-commerce. *Journal of Business & Economic Statistics*, 35(4):626–641, 2017.

Daniele Scarpi. Work and Fun on the Internet: The Effects of Utilitarianism and Hedonism Online. *Journal of Interactive Marketing*, 26(1):53–67, 2012.

Shalom Schwartz. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. volume 25 of *Advances in Experimental Social Psychology*, pages 1 – 65. Academic Press, 1992.

Shalom Schwartz. A proposal for measuring value orientations across nations. *European Social Survey*, pages 259–290, 01 2003.

Keith E. Stanovich and Richard F. West. Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, 23(5):645â665, 2000.

George J. Stigler. The economics of information. *Journal of Political Economy*, 69:213–213, 1961.

Franck Vigneron and Lester Johnson. A review and a conceptual framework of prestige-seeking consumer behavior. *Academy of Marketing Science Review*, 2:1–15, 01 1999.

Franck Vigneron and Lester Johnson. Measuring perceptions of brand luxury. *Journal of Brand Management*, 11, 07 2004.

Klaus-Peter Wiedmann, Nadine Hennigs, and Astrid Siebels. Value-based segmentation of luxury consumption behavior. *Psychology & Marketing*, 26(7):625–651, 2009.

Jr. William A Darity. *International encyclopedia of the social sciences*. Detroit, Mich. : Macmillan Reference USA, 2008.

Mary Wolfinbarger and Mary C. Gilly. Shopping online for freedom, control, and fun. *California Management Review*, 43(2):34–55, 2001.

Appendix

Laboratory experiment details

Below is the complete set of instructions that were provided to participants in the second phase of the experiment:

"Welcome to the laboratory part of the Great smartphone quest!
In this second part of the experiment, you work individually on your computer. You are not allowed to talk to others. If you have any questions, please raise your hand and the supervisor will come and answer your question. Before you start, please read these instructions carefully. The experiment consists of three tasks, A, B and C, plus a short questionnaire. You have 8 minutes time to complete each task and 10 minutes for the questionnaire. You complete these tasks using Mozilla Firefox browser. You can use the browser entirely as you like. Note, however, that all computer activity will be recorded and used for research purposes. The tasks are described below.

Rewards

You will receive your 20 euros gift voucher when you leave the laboratory today. In addition, on Monday May 23rd, we will randomly draw one participant (i.e. one 6-digit identifier). The randomly chosen participant will receive the reward for one of tasks A, B, or C (each task equally likely) as explained in the task instructions later.

Task A

You have a budget of 1000 euros. Using your web browser, find your preferred smartphone in any online store and add the phone in the store's shopping basket. However, do not complete the purchase. Instead, take a screen print of the shopping basket (using the

PrtSc key) and place the file to the folder with your identifier.
You have 8 minutes to
1) put your preferred phone in the basket
2) take a print-screen shot of the basket with your preferred phone in it
3) and placing the print screen file into the folder renamed after your identifier
Use only Firefox, do not close the timer tab. If you do not comply with these instructions you run the risk of losing the reward. Do not close Firefox before the supervisor advises to do so after the task.
As a reward for this task, we will give you
a) the phone you place in the basket.
b) a gift voucher to an online telecom store with a value of 90% of the 1000 euros. However, the price of the phone and possible expenses (such as shipping cost) will be deducted from the 1000 euros. We apply the price and the expenses of the online store where you made the 'purchase'.
In conclusion: if you are the winning participant drawn in the lottery on May 23rd and Task A is drawn as the relevant task, you will receive
a) the phone you chose
b) a gift voucher of value 90% x (1000 euros â price â expenses)
In this case your identifier will be announced as the winning code and you will be asked to collect the phone and the voucher.
Example:You choose Phone Z in Task A. The web store you decided to 'buy' the phone from sells it for 700 euros and there are no shipping costs. On May 23rd your personal identifier wins the lottery and Task A is randomly chosen as the one to be rewarded. You get the Phone Z and a 0.9 x (1000 â 700) = 270 euros gift voucher.

Task B

You have a budget of 1000 euros. Using your web browser, find your preferred smartphone in any online store and add the phone in the store's shopping basket. However, do not complete the purchase. Instead, take a screen print of the shopping basket (using the PrtSc key) and place the file to the folder with your identifier. You have 8 minutes to
1) put your preferred phone in the basket
2) take a print-screen shot of the basket with your preferred phone in it
3) and placing the print screen file into the folder renamed after your identifier

Use only Firefox, do not close the timer tab. If you do not comply with these instructions you run the risk of losing the reward. Do not close Firefox before the supervisor advises to do so after the task.

You can get information about popular smartphones in Finland. You do this by placing the red paper outside the sidewall of your cabin, after which the supervisor comes and hands you a chart showing the 15 most popular smartphones sold in Finland in the last week of April (in the order of popularity). Buying this information costs you 3 euros. This cost will be deducted from your reward for this task. Please, place the green paper outside the wall if you do not want the information.

As a reward for this task, we will give you

a) the phone you place in the basket.

b) a gift voucher to an online telecom store with a value of 90% of the 1000 euros. However, the price of the phone and possible expenses (such as shipping cost) will be deducted from the 1000 euros. We apply the price and the expenses of the online store where you made the 'purchase'.

Furthermore, for each participant (there are 81 participants including you) whose choice in Task A differs from your choice in Task B an additional 2 euros will be deducted from the 1000 euros gift voucher, and if you bought the piece of information an additional 3 euros will be deducted.

In conclusion: if you are the winning participant drawn in the lottery on May 23rd and Task B is drawn as the relevant task, you will receive

a) the phone you chose

b) a gift voucher of value 90 % x (1000 euros â price â expenses â 2 euros x choices in A that differ from your choice in B â 3 euros if you bought the info)

In this case your identifier will be announced as the winning code and you will be asked to collect the phone and the voucher as explained above.

Example: You choose Phone Z in Task B. The web store you decided to 'buy' the phone from sells it for 700 euros and there are no shipping costs. You decide not to buy the information. On May 23rd your personal identifier wins the lottery and Task B is randomly chosen as the one to be rewarded. In Task A 60 participants did not choose Phone Z. You get the Phone Z and a 0.9 x (1000 â 700 â 2x60) = 162 euros gift voucher.

Task C

You have a budget of 1000 euros. Using your web browser, find your preferred smartphone in any online store and add the phone in the store's shopping basket. However, do not complete the purchase. Instead, take a screen print of the shopping basket (using the PrtSc key) and place the file to the folder with your identifier.

You have 8 minutes to

1) put your preferred phone in the basket

2) take a print-screen shot of the basket with your preferred phone in it

3) and placing the print screen file into the folder renamed after your identifier

Use only Firefox, do not close the timer tab. If you do not comply with these instructions you run the risk of losing the reward. Do not close Firefox before the supervisor advises to do so after the task. You can get information about popular smartphones in Finland. You do this by placing the red paper outside the sidewall of your cabin, after which the supervisor comes and hands you a chart showing the 15 most popular smartphones sold in Finland in the last week of April (in the order of popularity). Buying this information costs you 3 euros. This cost will be deducted from your reward for this task. Please, place the green paper outside the wall if you do not want the information.

If you decide not to buy the information, the supervisor brings you a chart showing one smartphone randomly drawn among the 15 most popular smartphones sold in Finland in the last week of April. The same phone among the 15 will be shown to every participant who did not buy the information.

As a reward for this task, we will give you

a) the phone you place in the basket.

b) a gift voucher to an online telecom store with a value of 90% of the 1000 euros. However, the price of the phone and possible expenses (such as shipping cost) will be deducted from the 1000 euros. We apply the price and the expenses of the online store where you made the 'purchase'.

Furthermore, for each participant (in this task, there are 28 participants including you) whose choice in Task C differs from your choice in Task C an additional 4 euros will be deducted from the 1000 euros gift voucher, and if you bought the piece of information an additional 3 euros will be deducted.

In conclusion: if you are the winning participant drawn in the lottery on May 23rd and Task C is drawn as the relevant task, you will receive

a) the phone you chose
b) a gift voucher of value 90 % x (1000 euros â price â expenses â 4 euros x choices in C that differ from your choice in C â 3 euros if you bought the info) In this case your identifier will be announced as the winning code and you will be asked to collect the phone and the voucher as explained above.

Example: You choose Phone Z in Task C. The web store you decided to 'buy' the phone from sells it for 250 euros and there are 10 euros shipping costs. You decide to buy the information. On May 23rd your personal identifier wins the lottery and Task C is randomly chosen as the one to be rewarded. In Task C 19 participants did not choose Phone Z. You get the Phone Z and a 0.9 x (1000 - 250 - 10 - 3 - 4x19) = 594.90 euros gift voucher.

Questionnaire

After you have completed all three tasks, you go to "enter hyperlink", enter your personal identifier and fill in the short questionnaire. After you have submitted your answers, the experiment is over, and you can collect the 20 euros gift voucher from the supervisor. If you don't have time to complete the survey today, you can also do that later. However, if you decide to do so, you can collect your 20 euros gift voucher only after you have completed the survey but not before May 20th. You can collect the gift voucher at the reception desk in Hanken lobby by giving your 6-digit personal identifier."

Above excerpt describes one of the treatments used in tasks B and C, others follow the same methodology to calculate the final reward. List of treatments are included in Figure 2.11.

Table 2.7: Cognitive Reflection Test

CRT Questions System 1 and System 2	System 1	System 2
A pen and an eraser cost 1.10eur in total. The pen costs 1.00eur more than the eraser. How much does the eraser cost (in Euro cents)?	10 cents	5 cents
If it takes 10 machines 10 minutes to make 10 gadgets how long would it take 100 machines to make 100 gadgets (in min)?	100 min	10 min
In a pond there is a patch of pond-lilies. Every day the patch size is doubled. If it takes 48 days for the patch to cover the entire pond how long would it take for the patch to cover half of the pond (in days)?	24 days	47 days

System 1: Intuitive and spontaneous thinking
System 2: Rule-based and analytical thinking

Figure 2.10: Schwartz (2003) 21-item PVQ questionnaire

BENEVOLENCE
12. It's very important to him to help the people around him. He wants to care for other people. 18. It is important to him to be loyal to his friends. He wants to devote himself to people close to him.
UNIVERSALISM
3. He thinks it is important that every person in the world be treated equally. He wants justice for everybody, even for people he doesn't know. 8. It is important to him to listen to people who are different from him. Even when he disagrees with them, he still wants to understand them. 19. He strongly believes that people should care for nature. Looking after the environment is important to him.
SELF-DIRECTION
1. Thinking up new ideas and being creative is important to him. He likes to do things in his own original way. 11. It is important to him to make his own decisions about what he does. He likes to be free to plan and to choose his activities for himself.
STIMULATION
6. He likes surprises and is always looking for new things to do. He thinks it is important to do lots of different things in life. 15. He looks for adventures and likes to take risks. He wants to have an exciting life.
HEDONISM
10. Having a good time is important to him. He likes to "spoil" himself. 21. He seeks every chance he can to have fun. It is important to him to do things that give him pleasure.
ACHIEVEMENT
4. It is very important to him to show his abilities. He wants people to admire what he does. 13. Being very successful is important to him. He likes to impress other people.
POWER
2. It is important to him to be rich. He wants to have a lot of money and expensive things. 17. It is important to him to be in charge and tell others what to do. He wants people to do what he says.
SECURITY
5. It is important to him to live in secure surroundings. He avoids anything that might endanger his safety. 14. It is very important to him that his country be safe from threats from within and without. He is concerned that social order be protected.
CONFORMITY
7. He believes that people should do what they're told. He thinks people should follow rules at all times, even when no-one is watching. 16. It is important to him always to behave properly. He wants to avoid doing anything people would say is wrong.
TRADITION
9. He thinks it's important **not** to ask for more than what you have. He believes that people should be satisfied with what they have. 20. Religious belief is important to him. He tries hard to do what his religion requires.

Table 2.8: The AISS (Arnett Inventory of Sensation Seeking) Questionnaire

The AISS (Arnett Inventory of Sensation Seeking) Questionnaire
2. I can see how it would be interesting to marry someone from a foreign country.)
2. When the water is very cold I prefer not to swim even if it is a hot day. (-)
3. If I have to wait in a long line I’m usually patient about it. (-)
4. When I listen to music I like it to be loud.
5. When taking a trip I think it is best to make as few plans as possible and just take it as it comes.
6. I stay away from movies that are said to be frightening or highly suspenseful. (-)
7. I think it’s fun and exciting to perform or speak before a group.
8. If I were to go to an amusement park I would prefer to ride the roller coaster or other fast rides.
9. I would like to travel to places that are strange and far away.
10. I would never like to gamble with money even if I could afford it.(-)
11. I would have enjoyed being one of the first explorers of an unknown land.
12. I like a movie where there are a lot of explosions and car chases.
13. I don’t like extremely hot and spicy foods. (-)
14. In general I work better when I’m under pressure.
15. I often like to have the radio or TV on while I’m doing something else such as reading or cleaning up.
16. It would be interesting to see a car accident happen.
17. I think it’s best to order something familiar when eating in a restaurant. (-)
18. I like the feeling of standing next to the edge on a high place and looking down.
19. If it were possible to visit another planet or the moon for free I would be among the first in line to sign up.
20. I can see how it must be exciting to be in a battle during a war.

For each item, indicate which response best applies to you:
A) describes me very well
B) describes me somewhat
C) does not describe me very well
D) does not describe me at all

Table 2.9: AISS Novelty and Intensity subscale

AISS Novelty and Intensity subscale
Intensity subscale
1. I can see how it would be interesting to marry someone from a foreign country.
3. If I have to wait in a long line I'm usually patient about it.(-)
5. When taking a trip I think it is best to make as few plans as possible and just take it as it comes.
7. I think it's fun and exciting to perform or speak before a group.
9. I would like to travel to places that are strange and far away.
11. I would have enjoyed being one of the first explorers of an unknown land.
13. I don't like extremely hot and spicy foods. (-)
15. I often like to have the radio or TV on while I'm doing something else such as reading or cleaning up.
17. I think it's best to order something familiar when eating in a restaurant. (-)
19. If it were possible to visit another planet or the moon for free I would be among the first in line to sign up.
Intensity subscale
2. When the water is very cold I prefer not to swim even if it is a hot day. (-)
4. When I listen to music I like it to be loud.
6. I stay away from movies that are said to be frightening or highly suspenseful. (-)
8. If I were to go to an amusement park I would prefer to ride the rollercoaster or other fast rides.
10. I would never like to gamble with money even if I could afford it.(-)
12. I like a movie where there are a lot of explosions and car chases.
14. In general I work better when I'm under pressure.
16. It would be interesting to see a car accident happen.
18. I like the feeling of standing next to the edge on a high place and looking down.
20. I can see how it must be exciting to be in a battle during a war.

For each item, indicate which response best applies to you:
A) describes me very well
B) describes me somewhat
C) does not describe me very well
D) does not describe me at all

Scoring: Combine responses to items, with A = 4, B = 3, C = 2, D = 1, so that higher score = higher sensation seeking.
For items followed by (-), scoring should be reversed

Table 2.10: Post Questionnaire

Post Questionnaire
1. Please write your personal identifier in the field below.:(6 digits)
2. Please explain what made you choose the particular phone in Task A.
3. Please explain what made you choose the particular phone in Task B.
4. Please explain what made you choose the particular phone in Task C.
5. Which smartphone do you currently use? :(enter the model)
6. Was your current smartphone new or secondhand when you bought it? If new where did you buy it?
7. Was your current smartphone new or secondhand when you bought it? If new where did you buy it? Specify the place of purchase: Open text answers
8. Which telecommunications operator connection are you currently using on your smartphone?
9. Which telecommunications operator connection are you currently using on your smartphone? Other please specify: Open text answers
Is your current smartphone a company phone?
10. Do you know which smartphone you would buy next? If yes which one?
11. Do you know which smartphone you would buy next? If yes which one?:Yes: Open text answers
12. How much do you think your next smartphone will cost?
13. When do you plan to purchase a new smartphone?
14. What is your preferred way of shopping smartphones?
15. Select below the firms whose online store you have visited in the past year to explore smartphones tablets and computers.
16. Select below the firms whose online store you have visited in the past year to explore smartphones tablets and computers. Other please specify: Open text answers
17. What would be the most likely place for you to shop a new smartphone? Pick 1-3 alternatives.
18. What would be the most likely place for you to shop a new smartphone? Pick 1-3 alternatives. Other please specify: Open text answers
19. Which store/company do you think has the best selection of smartphones?
20. Which store/company do you think usually has the best offers on smartphones?

Table 2.11: Experimental variation in the incentive schemes

	Task B		**Task C**		
	1	2	3	4	
	Deduction:	Option to purchase information:	Deduction:	Option to purchase information:	
Treatments	EUR 2 deduction for each subject whose choices in task A differ from the subject's choice in B	EUR 3	EUR 4 for each subject whose choices in task C differed from the subject's choice in C	EUR 3	Number of subjects
Treatment 1	Yes	No	Yes	Yes	11
Treatment 2	Yes	No	Yes	No	9
Treatment 3	Yes	Yes	Yes	No	12
Treatment 4	Yes	Yes	Yes	Yes	14
Treatment 5	No	Yes	No	Yes	20

Table 2.12: OLS: Unique domains on personal characteristics

	Dependent variable:Number of domains			
	Task A	Task B	Task C	All tasks
	(1)	(2)	(3)	(4)
Conformity	0.420*	0.674**	0.402	0.499***
	(0.248)	(0.254)	(0.269)	(0.177)
Hedonism	0.663**	0.712**	0.492	0.622***
	(0.291)	(0.298)	(0.316)	(0.215)
Self Direction	0.705*	0.934**	0.779*	0.806**
	(0.378)	(0.387)	(0.410)	(0.357)
CRT	−0.675**	−0.555*	−0.534	−0.588***
	(0.305)	(0.313)	(0.331)	(0.214)
Patience	0.114	0.039	−0.309	−0.052
	(0.350)	(0.359)	(0.380)	(0.210)
Gender	0.312	−0.370	0.263	0.068
	(0.466)	(0.477)	(0.505)	(0.390)
Constant	3.255***	3.329***	3.044***	3.209***
	(0.457)	(0.468)	(0.496)	(0.460)
Observations	66	66	66	198
R^2	0.225	0.232	0.155	0.189
Adjusted R^2	0.146	0.154	0.069	0.163
Residual Std. Error (df = 59)	1.797	1.842	1.950	1.3816 (df = 191)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 2.13: OLS: Domain and Search queries including Huawei treatment interaction terms in Task C

	Dependent variable:					
	Search queries Task C			Domains Task C		
	(1)	(2)	(3)	(4)	(5)	(6)
Hedonism	0.756**	0.897***	−0.040	0.459	0.588**	0.003
	(0.314)	(0.331)	(0.637)	(0.291)	(0.264)	(0.616)
Conformity	0.740***	0.913***	0.406	0.437*	0.653***	0.352
	(0.277)	(0.291)	(0.406)	(0.256)	(0.232)	(0.393)
Self Direction	1.307***	1.494***	2.184***	0.782*	1.234***	2.049***
	(0.426)	(0.439)	(0.788)	(0.394)	(0.351)	(0.763)
CRT	−0.479	−0.343	−0.675	−0.552*	−0.238	−0.613
	(0.343)	(0.359)	(0.558)	(0.317)	(0.287)	(0.540)
Huawei Task C	0.581	0.031		1.411**	1.526	
	(0.695)	(1.156)		(0.644)	(0.923)	
Hedonism:Huawei Task C		−1.227			−1.096	
		(0.935)			(0.747)	
Conformity:Huawei Task C		−2.083**			−2.991***	
		(0.983)			(0.785)	
Self Direction:Huawei Task C		−3.022*			−5.806***	
		(1.693)			(1.352)	
CRT:Huawei Task C		−0.205			−1.115	
		(1.108)			(0.885)	
Treatment w/ option to purchase information in Task C			−0.416			−0.205
			(0.800)			(0.775)
Hedonism:Treatment w/ option to purchase information in Task C			1.019			0.594
			(0.729)			(0.706)
Conformity:Treatment w/ option to purchase information in Task C			0.378			−0.109
			(0.555)			(0.537)
Self Direction:Treatment w/ option to purchase information in Task C			−1.187			−1.895**
			(0.956)			(0.926)
CRT:Treatment w/ option to purchase information in Task C			0.477			0.283
			(0.707)			(0.685)
Constant	2.091***	2.063***	2.287***	3.015***	2.841***	3.156***
	(0.382)	(0.388)	(0.648)	(0.354)	(0.310)	(0.627)
Observations	66	66	65	66	66	65
R^2	0.254	0.328	0.335	0.206	0.468	0.235
Adjusted R^2	0.191	0.220	0.226	0.140	0.383	0.110
Residual Std. Error	2.025 (df = 60)	1.988 (df = 56)	1.979 (df = 55)	1.874 (df = 60)	1.588 (df = 56)	1.916 (df = 55)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2.14: OLS: Price on consumer characteristics

	Dependent variable: log(Price)
Gender	-0.502^{**}
	(0.206)
Search time	-0.029
	(0.042)
Patience	-0.309^{**}
	(0.149)
AISS	0.044^{***}
	(0.016)
CRT	-0.020
	(0.130)
Income	0.067
	(0.193)
Constant	6.708^{***}
	(0.281)
Observations	66
R^2	0.210
Adjusted R^2	0.130
Residual Std. Error	0.760 (df = 59)
F Statistic	2.615^{**} (df = 6; 59)
Note:	*p<0.1; **p<0.05; ***p<0.01

Chapter 3

Information search in the internet markets: experience versus search goods[1]

3.1 Introduction

Nelson's seminal work on economics of information and advertising (Nelson, 1970, 1974) sheds light on how information about product quality has significant impact on consumer demand and the market structure. Based on prior information about product characteristics, he categorised goods as *search* or *experience*. For search goods, a consumer has the opportunity to inspect an option before purchasing it, whereas for experience goods, quality evaluation is made possible only through purchase and subsequent use. In other words, search goods are dominated by attributes for which full information can be known prior to purchase, while experience goods are dominated by attributes for which full information cannot be known without direct experience. Typically, information search is more expensive for experience goods. This is because sources of product related information such as, word-of-mouth, advertising, user reviews etc. are less expensive than to sample via purchase. Furthermore, Nelson postulates that buyers are more likely to follow recommendations from others for experience goods as compared to search goods. However, with the emergence of electronic commerce and subsequent popularity of the online shopping phenomenon, it

[1]This chapter is based on an article published in Electronic Commerce Research and Applications (Basu, 2018)

has been argued by several researchers (Lisa, 1998; Lynch Jr. and Ariely, 2000; Alba et al., 1997) that, consumers are now exposed to a broader spectrum of information as compared to the traditional brick-and-mortar stores, which thereby bridges the gap between search and experience goods to a major extent.

This paper identifies relevant factors affecting associated search processes, as it studies how search and experience goods can be distinguished in the context of online search for information. The key questions that are addressed in this paper are: Has the availability of information enabled by the internet markets altered search intensities for experience goods versus search goods, compared to brick-and-mortar stores? Do multi-product online retailers enforce higher search intensities as a result of facilitating increased information acquisition? Do online recommendation agents have a stronger impact for experience goods as compared to search goods, as claimed by Nelson? These questions allow the reader to not only understand the implications of sequential search online in the context of the said types of goods, but also when a consumer stops searching for more information and makes a decision on whether to purchase a product, given her prior knowledge and information aids, such as recommendation agents. Furthermore, these questions have important implications for firms. Understanding the factors that affect information search, firms could potentially influence the extent of search. $Ex\ ante$, it is not evident if firms directly benefit when consumers spend more time searching for product information, however, the information obtained could either increase or decrease consumer valuation of the product. As a result, the extent of search is closely related with the optimal price and subsequently profit realisation for a firm.

Studying the same market shift and it's consequences, Klein (Lisa, 1998) suggests that by integrating features such as expert reviews, virtual demonstrations, information about history and personality of a brand and its product, internet markets can transform the search processes associated with experience and search goods. Nakayama et al. (Nakayama et al., 2010) examine past studies and empirical data on SEC (Search, Experience and Credence goods) ratings to check for evidence for the argument that the web has transformed experience goods into search goods, providing "experience without ownership." From the survey data on hypothetical online purchases they find that, while some products do exhibit ES (experience to search) shifts as a result of online presence of stores and more accessible information structure, others do not, depending entirely on counter-forces of internet advertising. The classification of search and experience goods laid down in the 1970s by Nelson and ever since is based on qualitative definitions. Most of the existing literature, to my knowledge, use qualitative data or empirical methods alone to study search and experience goods. This paper aims to provide a formal framework around the search processes associated with these two types of goods, stemming from their inherent definitions. To formalise a differentiating criterion between the types, a search theoretic model is developed that pins down optimal stopping rules showing that, in equilibrium, consumers search less for experience goods as compared to search goods, i.e., the optimal number of searches prior to purchase is lower for experience goods than search goods in the internet markets.

Commercial interactions online have drastically increased in the past few years, leading to the emergence of visible product networks that explicitly link related or complementary products to one another. Pre-purchase information acquisition is materialised in numerous ways in order to help potential consumers make the most efficient purchase decisions. One of these several mediums happen to be recommendation agents which provides potential buyers with product recommendations based on user-specified preferences, individual shopping history, or choices made by other consumers with similar search profiles or browsing histories. Online search for product information (based on web browsing) and choice in the absence of personalised recommendations detriments the validity of any study on consumer search behaviour in the internet markets. Virtual shopping has proved to reduce search costs for consumers and helped them make informed choices as they have access to a wide range of information without incurring the associated costs of visiting the physical stores (Kim et al., 2010; Xiao and Benbasat, 2007). However, there is mostly an enormous amount of information about a particular product from several different sources, which further leads to information explosion. The challenge which buyers commonly face in an online environment is to filter out the most relevant information that helps them make purchase decisions. Xiao and Banbasat (Laband, 1986) construct several propositions based on the effects of recommendation agents on consumer decision making and find that personalised recommendations generated via *collaborative filtering* or *content filtering* improves decision quality of shoppers. Chen et al.(Chen et al., 2004) collect data from Amazon on books categorising them into bestsellers, popular and less popular books with the aim of studying how recommendations impact firm revenue. Their main findings show that recommendations are mostly positively related to sales. An important issue with regard to recommendations is their perceived quality or usefulness from the consumer's point of view. Aggarwal and Vaidyanathan (Aggarwal and Vaidyanathan, 2005) conducted an experiment to gather information from the subjects about their reactions to recommendations which were generated based on their stated preferences. The experiment was designed for a search good (camera) and an experience good (music CD). The results of their study show that the perceived effectiveness and satisfaction from recommendations are higher for search goods as compared to experience goods. In this paper, we investigate the effect of recommendations on search intensity. The effect of the choice to follow a recommendation prior to purchase is studied in detail and it is shown that it has different implications when for different product types. The findings are aligned with Nelson's claim that consumers follow recommendations for experience goods more strongly than search goods and that also holds true even for the internet markets. However, it must be noted that recommendations in Nelson's claim meant third-party recommendations, while this study focuses on personalised recommendations that are auto-generated. Including this information not only enriches the data and significantly improves the model, but the findings also throw light on the importance of such informational aids in studying information search. The empirical specifications in this paper takes into account two variables indicating the effects of recommendation agents on search online; the first being the choice to follow a recommendation and the second represents the quality of a recommendation, which basically determines ex-post how similar a recommendation followed, is in its attributes to the good that was bought.

The underlying assumption of complete learning after one alternative is searched, is in line with that of search for the lowest price (Diamond, 1971; Stahl, 1989) or search for the best matched alternative (Weitzman, 1979; Wolinsky, 1986; Armstrong et al., 2009) in previous literature. According to the original search model developed by Stigler (Stigler, 1961), a consumer decides on the number of searches she is going to undertake prior to searching and then choose the best match from the set of alternatives she has examined. An alternative way of understanding consumer search was proposed by McCall (McCall, 1970), Burdett and Judd (Burdett and Judd, 1983) etc., namely, sequential search, wherein, a potential buyer examines the available alternatives sequentially, that is, she searches one more time if the marginal benefit from an additional search exceeds its marginal cost. In more recent times, Santos et al. test several restrictions that classical search models impose on search behaviour using browsing and purchase data from ComScore. They find that fixed sample search strategy explains observed search behaviour better than sequential search strategy, when shoppers are searching for the best price online (De Los Santos et al., 2012). However, in a more recent paper they explain that when consumers learn about the utility distribution by Bayesian updating, there is recall of the previously sampled options which explains why fixed sample search strategies outperforms sequential search strategies in online search behaviour (Alba et al., 1997). In the current dataset, over 68 percent of the consumers adopt the sequential search strategy which was concluded from the 'No recall' test consistent with (De Los Santos et al., 2012), hence we will assume that consumers engage in sequential search with no recall. [2] There is growing literature on the effects of uncertainty and learning on demand for experience goods. Erdem and Keane (Erdem and Keane, 1996) study brand choice by estimating a Bayesian learning model on household purchases of laundry detergent, where consumers are learning about a single unobservable product attribute. Learning is identified from the behaviour of repeat buyers with different purchase and advertising exposure histories which theoretically explains the formation of brand loyalties. Ackerberg (Ackerberg, 2003) builds on Erdem and Keane(Erdem and Keane, 1996), estimating a learning model on individual level panel data on consumers' choice to purchase a new brand of yogurt which can be considered as an experience good, since consumers have no prior information about this particular brand. On the supply side, Laband (Laband, 1986) found based on Yellow pages data, that, sellers provide more information for experience goods than search goods, at least in terms of advertisements. In a subsequent paper, Laband(Laband, 1991) found that information provided by the firms is positively related to product prices, based on newspaper advertisements. This paper contributes to the literature on information search, however, differing in framework and assumptions. Unlike Ackerberg(Ackerberg, 2003) or Erdem and Keane(Erdem and Keane, 1996), in this paper, consumers learn about entire product descriptions while sampling an alternative and decide if the utility they derive from the information is greater than their reservation utility. In case that is true, they buy the product without searching any further. Furthermore, it differs from current literature in empirically looking at a space of more than one product or a controlled set of *n* products; the data offers hundreds of products which are categorised into two broad groups. The scope of this data is broad, as it records search behaviour accurately in face of information explosion as customers may engage in directed or undirected search.

Apart from investigating optimal search patters in the context of experience and search goods, this paper makes methodological contributions towards developing a measure for "prior knowledge". In studying search and matching models, knowledge and information

[2]Under the null hypothesis of the standard sequential search model, recall of previously sampled alternatives should not be observed unless the consumer has sampled all of the stores she is aware of.

acquisition prior to purchase play a vital role, and product descriptions allow us to identify the information shocks to a major extent. Novel methodology from the field of information retrieval is used to construct similarity scores between each good searched and the bought good, taking all product information into account. Using *Term frequency-Inverse document frequency* (Tf-Idf) a new proxy is defined for prior information, adding all similarity scores until the session in which a good was finally purchased. This method has several applications, especially in scoring and ascertaining relevancy. Most of the above studies have focused on a single good across several stores, however, in this paper, consumers search among different products across several categories and brands in a single multi-product store, which makes it prudent to take into account all relevant product attributes into account when developing a measure for prior knowledge. Moreover, the information sets are partitioned into prior and posterior, where the latter is precisely the current information obtained by the agent given her prior knowledge. Novel data on browsing and purchase behaviour of a relatively large set of consumers on a Nordic multi-product retailer between March 2014 to March 2016 is used, containing diverse set of search and experience goods. The data described in Section 2 allows us to observe how consumers typically search, percentage of repeat buyers, average search time per session, implications of controlling for the existence of recommendation agents and their relationship with search variables. In the empirical part of the paper, certain products are focused on that are typically classified as search or experience goods in literature.

In Section 3, a model of optimal search is developed and it shows that the equilibrium number of searches prior to purchase is lower for experience goods as compared to search goods when consumers are searching for product information sequentially. The optimal stopping rules are based on the assumption that attribute search is continuous and the marginal utility from an additional search can be modelled as an *Itô process* (Dixit and Pindyck, 1994). The online search environment is interpreted in a way that at each search opportunity, consumers get an information shock, positive or negative, that takes them closer to identifying their ‘best-matched’ good. If the shock is positive then the current viewed product has similar attributes to the ideal product, while if the shock is negative, there is very little or no similarity.

3.2 Data

Click stream data records entire search paths of agents and is an extremely powerful source of information on consumer behaviour online (Bucklin and Sismeiro, 2003; Chatterjee et al., 2003). It not only monitors purchases, but also tracks consumer actions, such as their search behaviour over time. Using click stream data involves several challenges such as understanding and structuring the online shopping environment with the associated data preprocessing. Figure 3.1 looks at a typical example of clickstream data which shows the entire customer journey online, from the point she clicks on the first good to the point a purchase is made. At ever search (click), the products are viewed along with their descriptions and the customer updates her preference.

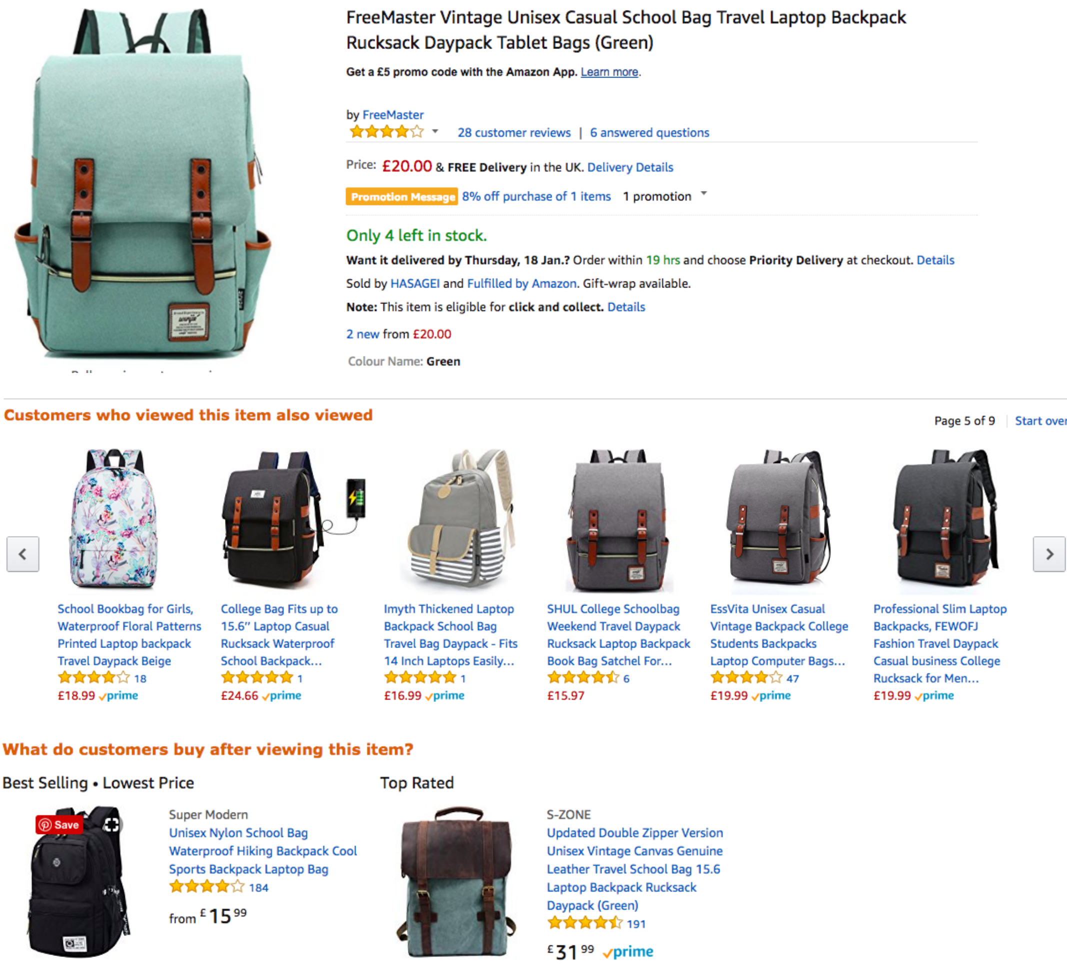
Figure 3.1: Clickstream data: Example

The dataset is constructed by monitoring purchases of anonymised Nordic shoppers online between November 2015 and March 2016. The data comes from a leading Nordic multi-product retail store that offers sufficiently many types of search and experience goods[3]. The main idea is to examine search patterns of the representative agent across search and experience goods, hence browsing and purchase data is collected only for a single large store, in order to avoid heterogeneity of preferences across stores offering substitutes. Only browsing sessions of users who have eventually made a purchase from the store have been considered. This is because often consumers visit online stores to obtain product related information and then make purchases from physical stores. The results in section 5 are representative of search paths when the consumer definitely buys at the end. Purchases during the period between November 2015 to March, 2016 are observed along with the related browsing behaviour of the unique users, since their first arrival at the store. This means that although all the non-purchase sessions were excluded as data points, but the whole search history was used to construct some of the control variables explained in section 4. Including entire user search histories, data was collected from March 12, 2014 to March 23, 2016. Furthermore, the dataset includes information when an agent clicks on a *recommended* product. Recommendation agents track each customer's search behaviour and shows her a set of relevant products, based on similar search profiles of other customers. Figure 3.2 gives an example of how recommendations generated via recommender systems typically

[3]The data source cannot be revealed due to non-disclosure agreements.

appear on online retail platforms. For each product viewed, appears one or more panels below with suggestions with statements such as **"Customers who viewed this item also viewed"** or **"What do customers buy after viewing this item"**.The effect of clicking on each such recommended good and of its similarity with the bought good on optimal search for information, is studied at length in the empirical model.

Figure 3.2: Online recommendations: Example



For accuracy, we look at the goods purchased and consider when in the browsing sessions those particular goods were carted immediately before purchase. This is because once a good is carted, a potential buyer may browse through other goods that he would like to buy, for example, a consumer carts good A in period t and is also interested in buying good B afterwards, such that he searches for B from t+1 through to t+4, and then carts B in period t+5 and finally purchases both goods A and B together in the following period. In this case, from period t+1 to t+4, the consumer does not search for any information related to good A, hence considering carted events rather than purchase events as data points is likely to yield more accurate results. The dataset contains duration of visit, number of brands and products, price of goods and related product descriptions of each good viewed. Table 3.1 gives a description of the data sample.[4]

Table 3.1: Descriptive statistics of the multi-product store sample

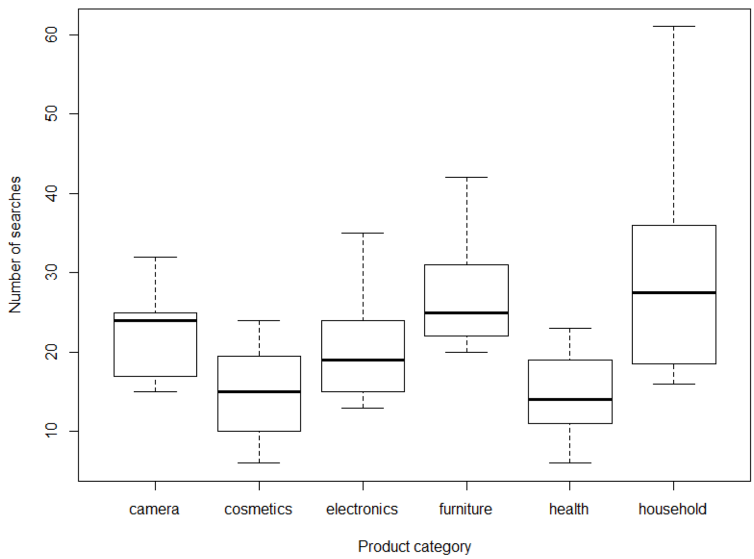
Search patterns	Mean	St. Dev.
Number of goods searched prior to purchase	18.34	16.63
Search time prior to purchase(in minutes)	26.48	26.35
Number of goods searched when no purchase is made	4.81	7.59
Search time when no purchase is made(in minutes)	6.87	14.50
Total umber of goods during November 2015 - March 2016		
Number of unique goods bought	10000	
Number of buyers	15000	
Percentage of repeat buyers	40.5	
Number of unique goods in each category in the sample		
Camera	97	
Cosmetics	278	
Electronics (Mobile phones, Laptops etc.)	521	
Furniture	452	
health	1795	
Household appliances	866	
Total	4009	

As is observed Table 3.1, approximately 40 percent of the shoppers in this particular sample are repeat buyers. However, for the baseline model, every purchase is considered unique. This means that if an agent buys two products on two different occasions, they are

[4]'Number of goods bought' and 'Number of buyers' in Table 3.1 represent approximate figures since the data sample is proprietary.

treated as two unique search paths. The difference in average time spent on search when it leads to a purchase and when it does not, is stark. This motivates the exclusion of non-purchase search paths when analysing differences in consumer behaviour towards search and experience goods. The total number of goods and consumers in the period of five months is much larger than what is used in the dataset for this paper. This is due to several restrictions that had to be imposed for modelling purposes and are mentioned in course of the paper, where necessary. The final sample consists of a total of 4009 observations, each of which is a consumer-product pair, that is, each point a consumer is on a unique product page.

Figure 3.3: Number of searches before purchase across different products for a super-store over a period of 5 months



In the empirical section of the paper, many of the variables of interest are derived or calculated from the data, however figures 3.3 and 3.4 exhibit some other key features that are more directly observable, namely, the extent of search and prices, respectively. It can be observed that, cosmetics and healthcare products are searched the least among all goods in the sample, the number of searches ranging between 8 and 20, and the median being approximately around 12-14. Cosmetics and healthcare products are classic examples of experience goods and consumers tend to behave similarly while searching for these goods. On the other hand, consumers engage in longer periods of searches for cameras, furniture and other electronics, for example, the maximum number of searches prior to purchase for a household appliance was 60, the median number of searches being approximately around 30. This implicit clustering behaviour of consumers' extent of search across the two categories of goods can be explained simply from their inherent definitions. As the experience good shopper is able to evaluate her match quality with a particular good only after buying it, there is not much incentive for her to search extensively prior to purchase, while search good shoppers have much higher marginal benefits of search.

Figure 3.4: Prices for different products for a super-store over a period of 5 months

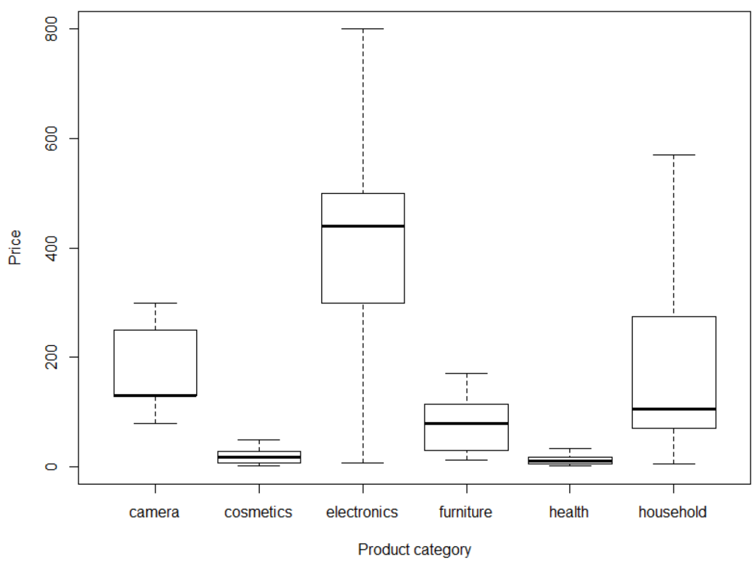


Figure 3.4 shows that all those under the broad umbrella of electrical and electronic goods have a broad range of prices, the upper limit being rather high. For example, in this sample, the most expensive personal electronic good (laptop/mobile) bought was 800 euros, camera was approximately 300 euros and household appliance, around 600 euros. However,

the median prices for each of these three goods vary quite a bit; while, the median price of personal electronics is closer to the third quartile, it is between the first and second quartile for cameras and household electronics. The experience goods on the other hand, namely, cosmetics and healthcare products are relatively a lot less expensive, the upper limit being approximately around 100 euros.

3.3 Model

Search and experience goods are categorised based on highly subjective characteristics and it is entirely dependent on individuals as to how much information they consider sufficient before making purchase decisions and the search costs they are willing to incur to obtain that information. Based on several observations, some goods are typically characterised as search and some as experience goods in literature. Electronics, furniture and household appliances are classic examples of search goods,while cosmetics and health products are typical examples of experience goods (Huang et al., 2009; Leahy, 2005; Nakayama et al., 2010; Nelson, 1970).

Consider a set up where consumers are gathering information about their best matched product, sequentially at the online platform of a multi-product retailer, with no recall. They consider multiple attributes of a product before a purchase a made, such as, colour, size, brand, prices, technical specifications etc. Each product has its own set of attributes and at each search (when a consumer clicks on a product and views it), she learns of these attributes. The realisation of each attribute either increases of decreases the total expected utility. Every product viewed gives incremental information about the potential attributes of her best matched product, assuming she does not know the exact location of the good she wants to buy. Let an agent derive an instantaneous utility, u_t from the information obtained at each search event, t. At every search t, $u(t)$ increases or decreases by q. *Ex ante*, the consumer does not know the value of q at any t, she pays a search cost of time and learns the realisation of q. If we imagine $u(t)$ as a *random walk* which increases or decreases at every t, and if the product attributes and time intervals are extremely small valued, then in the limit $u(t)$ is a *brownian motion*. As there are many alternatives to sample from and the process of browsing through products online is quick, the search process is assumed to be continuous (Branco et al., 2012).[5]

[5]Branco et al. study optimal search for product information, where consumers search across multiple attributes of one product. With an infinite number of attributes, each providing an infinitesimal amount of

Optimal stopping rules are defined for search and experience goods in a search theoretic model framework, consistent with Weitzman (Weitzman, 1979), where the optimal sequential search procedure is to start searching at the highest reservation utility and to terminate search whenever the maximum sampled utility exceeds the reservation utilities of all remaining unsampled alternatives. The consumer's utility function can be written as follows, given, $u(t)$ is the instantaneous utility at search event t :

$$U = E \int_0^\infty e^{-\rho t} u(t) dt$$

where ρ is the discount rate. u is considered here to be a state variable that changes stochastically with the number of searches, and hence, the change in utility from an additional search, du is modelled as an Itô process, a continuous time stochastic process, such that,

$$du = \mu dt + \sigma dz$$

An Itô process is a generalised brownian motion where parameters μ denote the drift rate and σ^2 denotes the instantaneous rate of variance or diffusion rate. The drift rate that exhibits the average rate of growth in the long run, must be positive in the current set-up. This is because with each marginal search, a potential buyer obtains information about her best-matched good, so it must be the case that utility from getting additional information is growing positively on an average, otherwise there exists no incentive to search over time. In the economic sense, σ may be interpreted as indicating the informativeness of search, such that a high σ would imply less informative search and vice versa. In a way it captures how the agent processes relevant information from the entire search path, as she updates her reservation utility after each search.

The central assumption of the model is that, informativeness of search which is represented as the inverse of σ, is lower for experience goods as compared to search goods. This is because, by definition, potential buyers are better informed about the attributes of search goods, hence it is reasonable to assume that they will search for the most important attributes (ones they are unaware of) first. Attribute search takes place in decreasing order of

information, one can assume that the expected valuation follows a Brownian motion while searching. They assume attributes to be sufficiently small-valued, hence search to be continuous, such that marginal utility from attribute search gets infinitely smaller as the number of searches go to infinity. They model the u process as a *brownian motion* in the limit.

importance for search goods, so the change in utility from each marginal search is changing consistently. However, for experience goods, the first time potential buyers have very little to no information about their best matched product, hence their search for information is undirected and the change in utility is likely to be inconsistent across periods. Apart from basing the assumptions on qualitative definitions of these product classes, we look into the data to verify any existing search patterns. Figure 3.5 shows two representative agents' search paths, one that buys an example search good (electronics) and one that buys and example experience good (clothing) at the end. The first row observes the customer journey of the latter, from the first search event until a purchase was made, and the send row observes the former. Evidently, the second customer engages in directed search (low variance), while the first customer browses through markedly different products from the one that she buys at the end. Performing several exposed checks in the current data set, combined with the intuitive classification of the product types, we assume that experience goods are related to lower search variances as compared to search goods.

Figure 3.5: Search paths: experience versus search good

If a consumer stops search at any point without buying the good, her instantaneous pay-off is zero. The stopping rule is to terminate search and buy the good, whenever utility derived from the good being 'sampled' is greater than or equal to a cut-off level, $u^\star$. In case utility is strictly less than $u^\star$, it is optimal for an agent to continue her search.

$$u^\star = \begin{cases} u \geq u^\star & \implies \text{stop search and buy} \\ u < u^\star & \implies \text{continue search} \end{cases}$$

Let $u' = u + du$ and $t' = t + dt$, where dt is a very small change in time. Then the value function can be written as,

$$V(u,t) = \max[udt + \frac{1}{1+\rho dt} EV(u',t')]$$

This is the Bellman equation that defines the consumer's optimal decision rule, given that the utility derived from having searched for t periods is udt and the present value of searching in future is $EV(u',t')$ times the discount rate, ρ. In the continuation region, the expected utility reduces to

$$\rho V(u)dt = udt + E(dV) \tag{3.1}$$

Now, since du follows an Itô process, we can apply the Itô's lemma,

$$dV = \left[\frac{\partial V}{\partial t} + \frac{\partial V}{\partial u}\mu + \frac{1}{2}\frac{\partial^2 V}{\partial u^2}\sigma^2\right] dt + \frac{\partial V}{\partial u}\sigma dz$$

$$\implies E(dV) = \left[\frac{\partial V}{\partial t} + \frac{\partial V}{\partial u}\mu + \frac{1}{2}\frac{\partial^2 V}{\partial u^2}\sigma^2\right] dt$$

since dz is a Wiener process. Plugging the above into (1), the reduced equation is obtained,

$$-\rho V + \mu V' + \frac{\sigma^2}{2} V'' = 0$$

whose solution can be written as,

$$-\rho e^{ru} + \mu r e^{ru} + \frac{\sigma^2}{2} r^2 e^{ru} = 0$$

$$r = \frac{-\mu \pm \sqrt{\mu^2 - 2\sigma^2\rho}}{\sigma^2}$$

The consumer's objective function can be expressed in the following way:

$$U = E\int_0^\infty e^{-\rho t} u(t)dt = \int_0^\infty e^{-\rho t}[\mu t]dt = \frac{\mu}{\rho^2}$$

Then, the particular solution obtained is of the following form,

$$V(u) = \frac{\mu}{\rho^2}$$

such that the general solution is:

$$V(u) = \frac{1}{\rho}(u + \frac{\mu}{\rho}) + c_1 e^{r_1 u} + c_2 e^{r_2 u}$$

At every t, the consumer must decide whether to search one more time or to stop search and purchase the good. $u^\star$ pins down the optimal value of the control variable, when shoppers derive utility from searching for product related information. In order to find a solution for the optimal $u^\star$, a set of boundary conditions are required. The following depicts the *value matching* condition which ensures continuity of the value function, $V(u)$. In economic terms, it implies that utility derived from search is large enough for the consumer to be indifferent between searching once more and buying the good. $u^\star$ is the upper bound such that when u reaches $u^\star$, the consumer buys the good with certainty, getting utility $u = u^\star$.

$$V(u^\star) = u^\star \tag{3.2}$$

The following depicts the *smooth pasting* condition which is the partial derivative of $V(u^\star)$ which ensures there are no kinks at the boundaries. The condition is implied by the fact that the consumer maximises $V(u)$ for all u (Branco et al., 2012) (Dixit and Pindyck, 1993).

$$V'(u^\star) = 1 \tag{3.3}$$

Substituting the boundary conditions into the general solution yields the following,

$$V(u^\star) = \frac{\mu}{\rho^2} + c_2 e^{r_2 u} = u^\star$$

$$V'(u^\star) = r_2 c_2 e^{r_2 u} = 1$$

(since, $\lim_{u \to \infty} V'(u) = \frac{1}{\rho}$, then from the boundary conditions, $c_1 = 0$). The optimal stopping rule is therefore represented by the following expression,

$$u^\star = \frac{\mu}{\rho^2} - \frac{\sigma^2}{\mu + \sqrt{\mu^2 + 2\sigma^2 \rho}} \tag{3.4}$$

$u^\star$ depicts the threshold value or the stopping rule itself; when $u > u^\star$, it is optimal to stop search and buy the good and when $u \leq u^\star$, search continues for a more suitable alternative. Now, consider two separate Itô processes for the two type of goods in question, such that the across t, for search goods is, ${\sigma_s}^2$ and that of experience goods is, ${\sigma_e}^2$

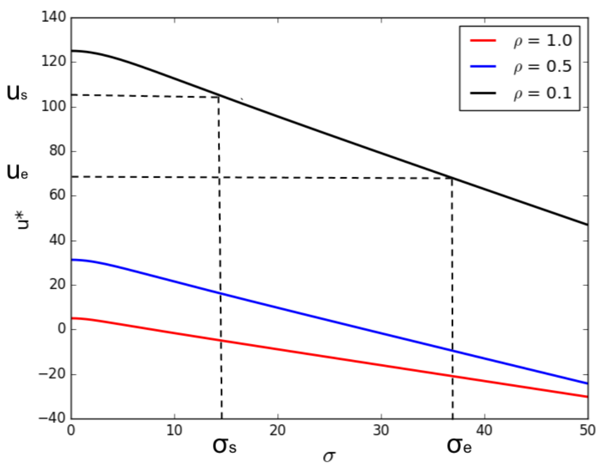
$$du = \mu dt + \sigma_s dz$$

$$du = \mu dt + \sigma_e dz$$

such that ${\sigma_e}^2 \geq {\sigma_s}^2$. Equation (3.4) along with this condition immediately gives the following proposition.

Proposition 1. *The optimal level of utility to stop search and buy the good being sampled is higher for search goods as compared to experience goods, i.e.,* $u_e^\star < {u_s}^\star$*, as long as* ${\sigma_e}^2 \geq {\sigma_s}^2$ *holds true.*

Figure 3.6: Optimal search paths for varying discount rates



This implies that consumers have higher search intensities for search goods as compared to experience goods which is explained by the difference in informativeness of each additional search. Therefore, the continuation range is much larger for former type and as a result, the time to buy is also higher on average compared to the latter type. Figure 3.6 exhibits changes in the path of $u^\star$ as σ changes for varying levels of ρ. It is evident that increasing the marginal benefits of search, by making it more informative, in turn decreases the level of utility at which the consumer is indifferent between buying the good and searching one more time. So, the main finding is that, the extent of search for experience goods is less than search goods, simply because, search for experience goods may be less informative as consumers arrive to the store with little to no prior information and can only evaluate match quality through consumption.

3.4 Empirical Analysis

Section 3 models consumer search behaviour across two product types and formally shows that the extent of search, prior to purchase, varies between search and experience goods when a consumer engages in information search online. In line with $Proposition$ 1, an empirical model is developed that aims to validate or disprove this finding, which further motivates the choice of the dependent variable, *Number of searches* in the model. The objective is to investigate thoroughly the factors that affect search, and how they vary across the two product groups, which leads to the following key hypothesis:

Hypothesis 1. *Search intensities for experience and search goods are the same when consumers engage in online search for information*

The above is in line with several studies in the past (mentioned in section 1) that claim that the internet markets or online retailers allow customers access to unconstrained information that enables them to search for experience and search goods equally intensively, such that it bridges the inherent gap between these two product types. Considering the key assumption that pins down our primary finding, it is evident that this hypothesis hinges largely on the role of informativeness of search in determining the optimal number of searches. In order to identify the variance of information shocks, let us start by defining the probability of agent n buying a good at search i, by P_{ni}. Let the utility derived by agent n from search i be denoted by, U_{ni}. Then, agent n chooses good i over j if $U_{ni} > U_{nj}$ and the probability is given by,

$$P_{ni} = Prob(U_{ni} > U_{nj}, \forall j \neq i)$$

U_{ni} is decomposed as $U_{ni} = V_{ni} + \epsilon_{ni}$, where, ϵ_{ni} captures all the factors that affect utility but are not observable by the researcher. It is assumed that ϵ_{ni} is i.i.d type- I extreme value distributed, which represents idiosyncratic taste of consumer n for good i,. The part of utility observed by the researcher or the representative utility is contained in V_{ni}, such that,

$$V_{ni} = \beta price_i + \theta H_{ni} + \lambda S_{nii^*}$$

where, H_i is a vector of environmental factors which includes dummies for holiday season or ongoing dicounts/sale periods and S_{nii^*} represents the similarity score between each good being sampled and the the bought good (Details on how S_{nii^*} is calculated has been provided in section 3.4.1).

$$P_{ni} = Prob(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \forall j \neq i)$$

$$\implies P_{ni} = \frac{e^V_{ni}}{\sum_j e^V_{ni}}$$

So, agent n buys good i with probability P_{ni} when the decision to buy depends on information derived from attribute search. The difference in choice probabilities across search, $P_{n(i+1)} - P_{ni}$ identifies the marginal utility of every additional search opportunity. Furthermore, the variance of $P_{(i+1)} - P_i$ for each customer n describes how informative search has been in identifying one's optimal match, thereby affecting the optimal number of searches prior to purchase. The informativeness of search is calculated by taking the inverse of search variance for the full journey of each customer. Different choice models can be applied to evaluate choice probabilities depending on the specifications of the density of unobserved factors. Assuming that unobserved factors are uncorrelated over choices and for the sake of consistency with existing literature, a logit model is used to estimate the probability of purchase at every search event for each customer. The weights of the control variables obtained from a standard logistic regression are used to calculate P for customer n for the good being sampled at search event i. As prices and similarity scores (key control variables), among other factors, vary across products, so do their weights, which contributes towards varying choice probabilities. The variance of these marginal probabilities $(P_{n(i+1)} - P_{ni})$ signals the informativeness of a customer's journey of information search. [6]

Additionally, the model aims to investigate if being more knowledgeable of product characteristics, owing to ease of information search online, makes search more efficient. In other words, the model aims to test if there is an inverse relationship between prior information and search intensity, which leads to the following proposition:

[6] The sample size was restricted only to those agents who engaged in at least four searches in order to get a reasonable measure for variance.

Hypothesis 2. *Increased product knowledge (through prior search) leads to increased search intensity, consistently across all product types*

One of the most prominent factors that classifies goods as search or experience, is clearly the amount of prior information, hence a reasonable assumption is that prior information impacts time to buy. In order to calculate prior information from the data, first we observe the browsing sessions of each unique buyer from the time of her first arrival to the store, until she carts the good (or goods) that is finally purchased. In each search until purchase of the said good, there is transfer of information, only some of which is relevant to the buyer. So, in order to understand how information acquisition influences purchase decisions, search events that could potentially inform consumers about the particular good that she has purchased, must be filtered. The idea is to collect all such search events that help a consumer make his final choice. To this end, a measure of similarity between products that have been viewed before the final purchase and the bought good is defined, based on the related descriptions. The method used to calculate the similarity scores are described in detail in the next section.

3.4.1 Calculating information based similarity scores

Each unique product has a description of the set of attributes associated to it in the data. These attributes typically describe the category or the type of every product, along with its key features. However, entity-related search on web data is non-trivial, in the sense that the attributes are described in different ways for different products and collecting the most important ones out of them is not straight-forward. The objective is to find similar products based on their intrinsic attributes, hence its essential to find commonality between every word of the attribute vectors (combination of several keywords). For example, an agent buys camera B in the period, t. In period t-2, say, she views camera A and in period t-4, she views a mobile phone. Now suppose, product descriptions for each of the three goods are described as in Table 3.2.

Table 3.2: Product attribute example

Product	**Attribute**
Camera B	’camera B with new features’
Camera A	’camera A’
Mobile phone	’mobile phone with new features ’

It is evident that similarity measures constructed based on the keywords in the attribute vectors is rather misleading. This is because, sampling camera A before purchasing camera B must give a buyer more information about her final purchase, rather then sampling a mobile phone. However, comparing the keywords in each of the vectors shows that, based on common keywords,

$$similarity\ score_{\{Camera\ A,\ Camera\ B\}} = 1$$

$$similarity\ score_{\{mobile\ phone,\ Camera\ B\}} = 3$$

Therefore, constructing similarity measures from the product descriptions, as is, yields highly inaccurate similarity measures between goods and hence constructing a statistic for prior information from these measures can lead to distorted results. In order to mine the relevant product attributes from the entire vocabulary of attribute elements, several methods from the field of text analysis and information retrieval can be used. In information retrieval, the task is to find important or relevant words from a large body of text or information set. *Vector space models* are most commonly used in computational linguistics and cognitive science, in the context of information retrieval.

A product description (document) is represented as a vector of it attributes (terms). The term frequency inverse document frequency (tf-idf) model is the most classical vector space model, wherein tf-idf is a measure used to evaluate how important a word/term is in a document amongst the entire collection of words/terms. It is often used as a weighting factor, as the importance of a word increases proportionally to the number of times it appears in the document, but is offset by the frequency of the word in the entire collection. Let t_y be the term frequency or the raw frequency of the term t in a document y, and y'_t be the document frequency or the number of documents in which term t appears. The inverse document frequency means how important a term is in a given collection, that is, how common it is across all the documents. Finally, let n be the total number of documents (product descriptions) in the data. Then the term frequency inverse document frequency weight, ω is calculated by,

$$\omega = t_y \times (y'_t)^{-1}$$

where,

$$(y'_t)^{-1} = log\frac{n}{y'_t}$$

Once the relevant words from the descriptions of the products have been retrieved, a measure of similarity between the attributes must be constructed. The *cosine similarity* between two vectors or two documents in the vector space is a measure that calculates the cosine of the angle between them. This is a comparison metric between two documents on a normalised space as it not only takes the magnitude of each word count (tf-idf) of each document into consideration, but also the angle between the documents. Let us consider vectors a and b, then the cosine of these two vectors is given by the Euclidean dot product:

$$a \cdot b = || a |||| b || \, cos \, \theta$$

$$S(a, b) = \frac{a \cdot b}{|| a |||| b ||}$$

The above is the metric for cosine similarity, (S(a,b)) shows how correlated two goods are based on their product descriptions. The cosine similarity of each product in the search history with the bought good is then calculated; expectedly, similar products have higher degrees of cosine similarity compared to products differing in attributes. Observing similarity scores between products in several random samples of goods, only those products are collected from the browsing history that have a similarity coefficient of 0.4 or above. This seems to be a reasonable threshold, as browsing through products that have markedly different attributes compared to the purchased good, do not essentially give potential buyers relevant information. Hence, towards the aim of constructing a variable for prior information related to the bought product, only those goods are considered that are evidently similar to the final purchase. For each purchase, the entire search history is observed in order to construct similarity scores for each sampled good with respect to the final purchase. Finally, summing over all the similarity scores greater than 0.4 until the purchase session finally gives us a proxy for the level of prior information of the agents.

Online search behaviour cannot be fully captured without understanding the effects of personalised recommendations on consumer choices in the event such a feature exists on an e-commerce platform. It is important to study the outcomes when a consumer's purchase decision has been in some way influenced by the recommendation agents compared to when it has not. The effect although is difficult to isolate, as recommendations might prolong the search process by exposing more of the product catalogue on one hand, on the other hand, it may suggest a potential buyer a well-matched product that one buys immediately, thereby reducing search time and effort. In order to capture this particular effect, a categorical variable, indicating the presence of a recommendation in the search path is introduced, which takes the value 1 if a purchase has been induced by a click on a recommended product during the final browsing session (that has ended in a purchase), or the value 0 if the buyer has rejected all recommended goods and sequentially searched until he has found his best-matched product. Recommendations aid the search and purchasing process for online shoppers, hence it is likely to have an effect on the time to buy, however according to prevalent claims in literature, experience good shoppers are more affected by these recommendations comparatively. The firm's incentives to recommend must be noted here: not only does a firm increase the likelihood of buy to making it an almost sure event by recommending the best matched product, but also keeps a potential buyer engaged in searching more by recommending a product that is matched just sufficiently well, such that a shopper believes that he is in his continuation region, or at least moving towards it. Therefore arises the need to study the effects of recommendations generated online and whether they can be viewed as a source of information from the perspective of potential buyers.

Furthermore, the 'quality' of a recommendation or its degree of match with each consumer's preferences, influences her search path and in turn often decides whether or not she is in the continuation region. In order to account for this effect, a variable is constructed that sufficiently indicates the quality of the recommendation clicked on at any point during the final browsing session that induces the final purchase. This is done in the same way as the attribute-based similarity scores are calculated for the prior information. This means that if in the final session, an agent clicks on a recommended good, the attributes of that particular good is matched against the set of attributes of the bought good and depending on the match, a measure for the quality of the recommendation is arrived at. This motivates the final hypothesis:

Hypothesis 3. *Recommendations agents exhibit stronger association with search for experience goods than search goods*

The following equation models the relationship between number of dependent variable, $Y = \textit{search intensity}$ and the vector of independent variables, $X = [P_j, I, K, (P_j * E), (I * E), (K * E)]$, where, P_j represents price for alternative j, I represents informativeness of search which is calculated from the choice probabilities, $I = (\sigma^2)^{-1}$, and K represents prior information that agent n possesses when she samples alternative j,

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \tag{3.5}$$

The indicator variable, E takes value 1 if a good is of type experience and 0 if it is of type search. Furthermore, interaction terms are added to observe the difference in the effect of the controls on number of searches across the two groups of consumer goods.

3.5 Results

The model is estimated using three classic examples of search goods (personal electronics such as mobile phones, laptops etc., furniture and household appliances) and two classic examples of experience goods (health and beauty products) for n shoppers, who have in total purchased 4009 products over a period of five months.

Table 3.3 (1) shows OLS estimates of search intensity on several important constructs across search and experience goods, while in (2), variables representing the choice to follow a recommendation and its quality are included in the baseline model. It is evident that including the effects of recommendation agents significantly improves the model, which shows the importance of these shopping aids in search and purchasing behaviour online. An increase in price has a significant positive relation with search intensity for search goods which implies, that more expensive the goods are, the more thoroughly they are searched for. On the other hand, an increase in prices for experience goods reduces search, although nominally.

An increase in prior information is positively related to the number of searches prior to purchase for search goods, but no significant association with the extent of search for experience goods. As product knowledge increases, search intensity increases by approximately 12 percent for search goods. This result allows us to reject *Hypothesis 2*: increased product knowledge that is facilitated by the unconstrained availability of information online, enables higher search intensity only for search goods but has no impact on experience good. This is because, in case of search goods, consumers direct their search towards important product attributes they might be unaware of, because they are able to identify the order of importance of these attributes. As more information only exposes more of the available options, they are likely to search more. As a result, more relevant information keeps customers engaged in search which increases the likelihood of a purchase per customer. This also sheds light on the firm's incentive to employ recommender systems on their retail platforms: by providing relevant product information, firms are able to increase customer engagement and improve chances of a successful transaction, at least for search goods. On the other hand, information search does not enable consumers determine match quality for experience goods. So, even for an increase in prior product knowledge, consumers have no effect on their search path, as prolonging their search does not enable them to learn about their optimal match. Browsing through more product characteristics simply does not pose as an incentive to search longer as it does not guarantee improved match quality for experience goods.

The coefficient for variance of search is highly significant and is inversely proportional to the amount of search, shoppers engage in. The inverse of variance represents informativeness of search, in identifying an agent's optimal match. This implicitly affects the marginal benefits of searching for product information. Uninformative search reduces the extent of search across both types of goods, which is consistent with the optimal stopping rule in section 3. However, findings from the empirical model suggest that, the magnitude of association is larger for experience goods. The extent of search associated with experience goods decreases three times more than search goods for an unit increase in variance. Given the assumption that experience goods are associated with higher search variance, even if search was equally as informative for both product types, a typical consumer would optimally search at least three times less for experience goods than search goods, prior to purchase. This result allows us to reject *Hypothesis 1* and deviate from prevalent claims in literature that the internet markets negates the difference between search and experience goods. As experience good shoppers have little to no information about true match quality at the beginning of search, the variance of the information shocks they face over time is rather high. This implies that their search is not informative of the optimal match and they have less incentive to continue searching. Due to limited knowledge, they may browse through goods that may have no similar attributes to the purchased good at all. This further reinforces the assumption in Section 3 that the variance of the information shocks from an additional search is higher for experience goods than search goods. In other words, a typical experience good shopper does not know of the most important attributes of his preferred product, hence has less incentive to search more, as opposed to search good shoppers, who are well informed of their preferences, hence search efficiently when they choose to buy.

Inclusion of explanatory variables that capture the effects of personalised recommendations on the extent of online search for information, improves the model considerably, which sheds light on the importance of recommendation agents on consumer search in the internet markets. The choice of following a recommendation significantly affects both search and experience goods; when consumers choose to follow a certain recommendation, it increases the optimal number of searches by an estimated 16.8 percent for search goods and decreases the number of searches almost by 9 percent for experience goods. The former is indeed a surprising result, since recommendations are expected to inform shoppers of their best-matched product, thereby reducing search. However, if the recommendations are such that the consumer is being informed of close substitutes to her expected match, she is likely to investigate more options before deciding on a particular good. This is indeed true for search goods, as investigating substitutes enables a buyer to determine her optimal match. Online recommendation agents have significantly opposing effects on search for the two products groups, impacting search goods more that experience goods, which rejects *Hypothesis 3*. In previous literature, recommendations have only been perceived, studied and analysed as decision aids that make the shopper more efficient, or reduce her search time. Contrarily, this result provides reason to investigate the behavioural mechanisms that firms use to employ in designing the recommendation algorithms. Keeping shoppers engaged on a website is beneficial for firms, although, it is often the case that when a potential buyer has spent considerable amount of time viewing a particular (type of) product, the recommendations thereafter are unrelated. This is contrary to the principle of recommendation agents per se, but firms may design them in such a way only to maximise the likelihood of the purchase at every search, due to lack of other relevant visible alternatives. For example, while searching for a coffee machine, if a buyer is shown recommendations of coffee filters, it is suggestive of the fact that there exists no better options than what she is viewing at present, which in turn leads her to confirm the purchase without getting distracted by other options. The challenge for firms is to design the cut-off point between when recommendations are useful sources of product related information and when they might distract the buyer by prolonging search.

Table 3.3: Estimates of search

	Dependent variable: Number of searches	
	(1)	(2)
Price	0.021***	0.018***
	(0.003)	(0.002)
Prior information	0.118***	0.131***
	(0.012)	(0.009)
Variance of search	−1.753***	−1.940***
	(0.087)	(0.095)
Recommendation quality		−0.101***
		(0.007)
Recommendation (1=used;0=not used)		0.168***
		(0.028)
Price*E	−0.032***	−0.102***
	(0.005)	(0.016)
Prior information*E	−0.019	−0.012
	(0.026)	(0.014)
Variance*E	−4.092***	−4.470***
	(0.633)	(0.324)
Recommendation quality*E		−0.151***
		(0.012)
Recommendation*E		−0.254***
		(0.038)
Constant	2.778***	2.696***
	(0.024)	(0.027)
Observations	4,009	4,009
Adjusted R^2	0.363	0.404

Note:
*p<0.1; **p<0.05; ***p<0.01
(1) exhibits the estimates of search;
(2) exhibits the estimates of search with recommendations
E=1 for experience goods; 0 for search goods

Table 3.4: Summary of results

Hypothesis	**Result**
1. Search intensities for experience and search goods are the same when consumers engage in online search for information	Reject
2. Increased product knowledge (through prior search) leads to increased search intensity, consistently across all product types	Reject
3. Presence of recommendations is associated with higher search intensities for experience goods than search goods	Reject

3.6 Discussion

Search intensities associated with experience goods is at least three times lower than search goods

This paper investigates the extent of online search that shoppers engage in as they search for product information for specific type of consumer goods. They obtain information by searching through webpages featuring different goods and their associated descriptions. By using detailed data on browsing and purchase behaviour of shoppers, as well as product descriptions of every good searched, optimal search paths of potential buyers across search and experience goods are studied. One of the primary findings is that search good shoppers search at least three times more as compared to experience good shoppers in equilibrium when they engage in continuous sequential search. This result further validates the finding in Proposition 1 empirically. The theoretical model in Section 3 aims to formally pin down a criterion that discriminates between experience and search goods in the context of information search in the internet markets. Optimal stopping rules that are derived from imposing boundary conditions allows the reader to understand search for these types of products in question beyond only qualitative definitions that are currently the benchmark. The framework of the model is chosen to reflect the nuances of attribute search online and suitable assumptions are made by internalising definitions of the two product groups, further vetted by the data used in this study. The empirical results reject Hypothesis 1 which claims that both types of goods have equivalent search intensities in the internet markets due to ease of information search, as we find that experience goods are searched for at least three times more as compared search goods. The theoretical and the empirical result combined contributes to the primary finding of this study. The paper starts with the popular claim from literature, that search good shoppers search more in brick-and-mortar stores than experience good shoppers and the counter-claims, that with the advent of e-commerce, both types of shoppers will search equally, as online search allows "experience without ownership", thereby bridging the gap between these two product groups. The theoretical model shows formally that the latter claim does not hold true, which is further validated by the empirical result later in the paper.

Combining the theoretical and empirical result, this study provides a formal criterion that distinguishes between search and experience goods, besides the amount of product knowledge of consumers prior to their arrival at the store, as postulated by Nelson. Furthermore, the extent of search depends highly on how informative search is, in the sense that it enables potential buyers identify their optimal match. Predicting the likelihood of a purchase at every click or pageview remains one of the more challenging problems in literature, however by looking at purchase likelihood in retrospect we are able to identify the variance of search or it's informativeness. It is interesting to observe that higher the variance of search, lower is the intensity across both product groups. Given that experience good shoppers systematically exhibit high search variance, their search paths are three times longer than search good shoppers.

Presence of recommendations have opposing effects on search for the two product groups, showing significantly higher association with search goods than experience goods.

Personalised recommendations are shown to affect online search and purchase decisions significantly as it takes into account user intent based on their search histories. The supply side value of deploying recommendation agents on an online shop has been studied in length as it creates massive cross-sell and up-sell opportunities for firms. On the demand side, however, there are fewer studies pertaining to this subject. The consensus is that online recommendations reduce search costs ,as shoppers can find information without actual search. This study contrarily points out that, how recommendations affect search intensities largely depends on shoppers being able to determine match quality for every product they sample. While, a recommendation reduces search for an experience good, it increases that for a search good. It is observed that online recommendations play a major role in consumers' purchase decisions. Following recommendations do not seem to affect the extent of search for search goods, however, in the occasion that a search good shopper does follow a recommendation, its quality matters: closer the match of the recommended good to her true preference, more she searches. This makes intuitive sense: having the opportunity to improve match quality, a typical shopper will search longer as he gets more information about the available assortment. On the other hand, recommendations agents reduce search time for experience good shoppers.

Novel methodology has been used to quantify prior consumer knowledge in the internet markets, which positively affects only search goods

In understanding online shopping behaviour, a measure has been developed in this study for prior product knowledge. The underlying assumption is that, browsing histories on a particular store sufficiently influences current purchase decisions, hence a metric is developed indicating the varying levels of product knowledge consumers begin their search with, in the session whose purchases are being considered. Vector space models such as Tf-idf is used to measure the relevancy of each good sampled to the good bought, which allows us to calculate prior information directly from search. Partitioning the information sets accurately to discriminate current and prior knowledge, enables successful mapping of search intensity to product knowledge. This leads to one of the key findings that increased prior knowledge is associated with a 12% increase in search for search goods, but no significant impact on experience goods.

The results may be generalised only to an extent primarily due to two *limiting factors*. Firstly, the data set does not contain any demographic or behavioural information about the shoppers. It is obvious that consumer search and purchase behaviour should vary across factors such as age, sex, location, psychological factors etc. The focus of this study is extensively on product characteristics as most of the explanatory variables are constructed based on those. Inclusion of consumer characteristics in the analysis would improve our understanding of the behaviour of online shoppers. Secondly, the metric for prior information may not be a true measure of the prior knowledge. Prior information in this paper, is calculated from the point an unique agent visits this particular store in question, for the first time. It may be restrictive to assume that if the agents have bought a good from this store, they obtain information only from this particular store. This assumption may be violated in reality as there may be other sources of information, including the physical counterparts of the online store. However, the goal of this paper is primarily to demonstrate the role that information search plays in purchase decisions online, and that these decisions are not simply based on the price distributions. The idea is to focus on the direction of the effects as well as the magnitude in order to get an idea of how the discussed factors determine the extent of search based on the type of good.

Acknowledgements

This work was funded by a personal grant to the author from the Yrjo Jahnsson Foundation.

References

Daniel A. Ackerberg. Advertising, learning, and consumer choice in experience good markets: an empirical examination*. *International Economic Review*, 44(3):1007–1040, 2003.

Praveen Aggarwal and Rajiv Vaidyanathan. Perceived effectiveness of recommendation agent routines: search vs. experience goods. *International Journal of Internet Marketing and Advertising*, 2(1/2), 2005.

Joseph Alba, John Lynch, Barlton Weitz, Chris Janiszewski, Richard Lutz, Alan Sawyer, and Stacy Wood. Interactive home shopping: Consumer, retailer, and manufacturer incentives to participate in electronic marketplaces. *Journal of Marketing*, 61(3):38–53, 1997.

Mark Armstrong, John Vickers, and Jidong Zhou. Prominence and consumer search. *The RAND Journal of Economics*, 40(2):209–233, 2009.

Shreya Basu. Information search in the internet markets: Experience versus search goods. *Electronic Commerce Research and Applications*, 30, 05 2018.

Fernando Branco, Monic Sun, and J. Miguel Villas-Boas. Optimal search for product information. *Management Science*, 58(11):2037–2056, 2012.

Randolph E. Bucklin and Catarina Sismeiro. A model of web site browsing behaviour estimated on clickstream data. *Journal of Marketing Research*, 40(3):249–267, 2003.

Kennet Burdett and Kenneth L. Judd. Equilibrium price dispersion. *Econometrica*, 51(4): 955–969, 1983.

Patrali Chatterjee, Donna L. Hoffman, and Thomas P. Novak. Modeling the clickstream: Implications for web-based advertising efforts. *Marketing Science*, 22(4):520–541, 2003.

Pei-Yu Chen, Shin-yi Wu, and Jungsun Yoon. The impact of online recommendations and consumer feedback on sales. *ICIS 2004 Proceedings*, (58):520–541, 2004.

Babur De Los Santos, Ali Hortaçsu, and Matthijs R. Wildenbeest. Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, 102(6):2955–80, 2012.

Peter A Diamond. A model of price adjustment. *Journal of Economic Theory*, 3(2):156 – 168, 1971.

Avinash K Dixit and Robert S Pindyck. Investment under uncertainty. *(Princeton University Press, Princeton, NJ*, 1993.

Avinash K. Dixit and Robert S. Pindyck. *Investment Under Inverstment*. Princeton University Press, Princeton, New Jersey, 1994.

Tulin Erdem and Michael Keane. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science*, 15(1):1 – 20, 1996.

Peng Huang, Nicholas H. Lurie, and Sabyasachi Mitra. Searching for experience on the web: An empirical examination of consumer behavior for search and experience goods. *Journal of Marketing*, 73:55–69, 2009.

Jun B. Kim, Paulo Albuquerque, and Bart J. Bronnenberg. Online demand under limited consumer search. *Marketing Science*, 29(6):1001–1023, 2010.

David N Laband. Advertising as Information: An Empirical Note. *The Review of Economics and Statistics*, 68(3):517–21, August 1986.

David N Laband. An objective measure of search versus experience goods. *Economic Inquiry*, 29(3):497–509, 1991.

Arthur S. Leahy. Search and experience goods : evidence from the 1960's and 70's. *The Journal of Applied Business Research*, 21(1):45–51, 2005.

Klein Lisa. Evaluating the potential of interactive media through a new lens: Search versus experience goods. *Journal of Business Research*, 41(3):195–203, 1998.

John G. Lynch Jr. and Dan Ariely. Wine online: Search costs affect competition on price, quality, and distribution. *Marketing Science*, 19(1):83–103, 2000.

John J. McCall. Economics of information and job search. *The Quarterly Journal of Economics*, 84(1):113–126, 1970.

Makoto Nakayama, Norma Sutcliffe, and Yun Wan. Has the web transformed experience goods into search goods. *Electronic Markets*, 20(3):251–262, 2010.

Philip Nelson. Advertising as information. *Journal of Political Economy*, 82(4):729–54, 1974.

Phillp Nelson. Information and consumer behaviour. *Journal of Political Economy*, 78(2): 311–329, 1970.

Dale O Stahl. Oligopolistic Pricing with Sequential Consumer Search. *American Economic Review*, 79(4):700–712, 1989.

George Stigler. The economics of information. *Journal of Political Economy*, 69, 1961.

Martin Weitzman. Optimal search for the best alternative. *Econometrica*, May, 1979.

Asher Wolinsky. True monopolistic competition as a result of imperfect information. *The Quarterly Journal of Economics*, 101(3):493–511, 1986.

Bo Xiao and Izak Benbasat. E-commerce product recommendation agents: Use, characteristics, and impact. *MIS Q.*, 31(1):137–209, March 2007.

Chapter 4

Personalized product recommendations and firm performance[1]

4.1 Introduction

4.1.1 Motivation

Personalized recommendations have evolved to be an integral part of the online shopping experience. While economics and marketing literature have largely focused on the impact of recommender systems on consumer choices online, information systems and computer science has focused on their algorithmic design and improving efficiency (Adomavicius and Tuzhilin, 2005). However, little research has been done to capture firm incentives that motivate use of recommendation agents in electronic markets. Online platforms today, leverage various sales support tools to improve conversion rate. Two of the most popular are recommender systems (generated by the firm) and online review systems (generated by shoppers). Recommender systems are widely used to inform and persuade shoppers to consider alternative products as they evaluate competing offerings, thereby converting browsers into buyers, increasing loyalty and improving consumer retention (Adomavicius and Tuzhilin, 2005; Adomavicius et al., 2018; Chen et al., 2004; Fleder and Hosanagar, 2009; Haubl and Murray, 2003; Hennig-Thurau et al., 2012; Senecal and Nantel, 2004; Xiao and Benbasat,

[1]This chapter is based on an article published in Electronic Commerce Research and Applications (Basu, 2021)

2007). While increasing the relevance of a recommendation has been the focus of academic research, it is not clear that maximizing predictive accuracy is the eventual goal of all recommender systems (Hosanagar et al., 2008). Several studies have provided evidence on how online recommendations significantly pull consumers' willingness to pay in the direction of the recommendation (Adomavicius et al., 2018; Senecal and Nantel, 2004; Amatriain and Basilico, 2016; Benlian et al., 2012; Kumar and Benbasat, 2006).

The importance of recommender systems in today's digital world is manifested in the firms' willingness to invest in acquiring and improving these systems, for example, the much publicized Neflix prize that "sought to substantially improve the accuracy of predictions about how much someone is going to enjoy a movie based on their movie preferences".[2]. Firms deploy recommender systems online that are either provided by third-parties or developed in house. In this paper, we focus on the former, where a firm incurs a direct cost as a function of the purchases via recommendations. Recommender systems predict with varying degrees of accuracy what is most relevant for potential buyers and display these predictions as shoppers browse, so as to help them make more informed choices or even to influence consumer preferences (Haubl and Murray, 2003; Hennig-Thurau et al., 2012).

This paper investigates the incentives of online platforms to show such recommendations and formally outline conditions under which they might improve performance. Furthermore, the trade-off between relevance of a recommendation and price of the recommended good is examined in detail in the context of firm's reputation. Given the state of the consumer, we examine optimal policies of the firm based on current profits and transition probabilities of shoppers switching between states. State {H,L} is determined by the reputation of firm generated recommendations, where in state H, consumer purchases via recommendations and in state L, via search alone. Deploying recommendations online or allowing consumers to only search is the decision variable in our model, which in equilibrium, is a trade-off between price of the recommended good and relevance of the recommendation to the consumer. The key questions addressed in this paper are: How do personalized recommendations impact firm revenue? Does the relevance of recommendations have an impact on firm revenue? How does reputation of recommendations determine optimal decision policy for online platforms?

---

[2]The million dollar Netflix prize (http://www.netflixprize.com/) was awarded to the team "Bellkor Pragmatic Chaos" in 2009 for improving Netflix's recommendation accuracy by 10%

4.1.2 Literature Review

As stated by Stiglitz (1989), in the absence of information, products may be viewed as perfect substitutes, but as information becomes available, they may become imperfect substitutes, giving rise to search costs in identifying one's match quality (Stiglitz, 1989). Online markets are often associated with enlarged consideration sets, simply due to larger set of alternatives compared to physical stores (Punj and Moore, 2009; Close and Kukar-Kinney, 2010; Hauser, 2014). This raises consumer search costs as shoppers aim to find the product most relevant to them. Firms recognize the existence of increasing search costs incurred by the consumer with an increase in available information. Therefore, to maximize purchase likelihood and performance, firms adopt numerous strategies with the aim of reducing such costs, such as providing product information on prices, availability, ease of use, to name a few, via online recommendations. As personalized recommendations help guide potential buyers on their search path and lead them to their 'best-matched' good, it builds a reputation of such firm-generated instruments (Fleder and Hosanagar, 2009; Chen et al., 2004; Hosanagar et al., 2008). Relevance of firms' recommendations in the future are dependent on its past performance. Assuming consumers can judge quality with complete accuracy after they have purchased a good, irrelevant recommendations will have a lifetime of only one period. Economic literature has studied extensively the relationship between firm pricing strategy and asymmetric market information. Typically in multi-period games, firms may need to practice predatory or limit pricing in the short-run, so as to maximize future pay-offs under incomplete information (Milgrom and Roberts, 1982b,a; Kreps and Wilson, 1982). It has been established that firms with high quality offerings aim to maintain their reputation, as it is rewarded with high prices and high profits in the long-run (Klein and Leffler, 1981; Shapiro, 1982; Allen, 1984; Houser and Wooders, 2006). A large class of goods do not satisfy the assumption of perfect consumer accuracy (Klein and Leffler, 1979; Dybvig and Spatt, 1980), therefore firms need to take reputation effects into account while choosing optimal policy (Mailath and Samuelson, 2001; Hörner, 2002; Cabral and Hortacsu, 2010). The findings in this study resonate with literature to the extent that showing recommendations to shoppers who strictly buy via search may be costly for the firm, however, may be optimal from the long-term policy design perspective.

Impact of firm reputation is two-fold. Firstly, relevant recommendations are more likely to get repeat buyers, especially ones with relatively high search costs. Secondly, they are also more likely to receive new consumers who are dissatisfied with their current options. Therefore, firms with more relevant recommendations, will in the long run have a larger consumer base and in turn higher revenue. Rogerson (1983) shows via number of theory-driven empirical tests that differences in feedback histories lead to differences in the prices of substitutes, across sellers with different feedback aggregates (Rogerson, 1983). One of the key findings of this study is that, in order to switch shoppers from buying via search to recommendations, relevance of a recommendation must be sufficiently high. Therefore, incentivizing those who buy via recommendations is more profitable for the firm in the long run. The empirical analysis in this study points to evidence on consumer state being endogenous to the model which stems from unobserved reputation effects. We show that reputation of firm generated recommendations have a significant positive association with revenue, which translates to clear incentives for firms to improve relevance of product recommendations.

The rest of the paper is arranged as follows. A theoretical framework is developed in section 2 to formalize conditions under which it is optimal for firms to recommend, given that consumers either purchase via recommendations or search alone. Section 3 provides a description of search and purchase data from a multi-product online platform during 2014-16, which is further used in the empirical analysis in Section 4.

4.2 Model

In this paper, we study a multi-product online platform's optimal decision policy to show product recommendations at a cost or let consumers obtain product information via search alone. Consider a set-up where the firm offers n products and has the option to show a recommendation at each period for every consumer. Based on the number of 'success events' the firm incurs a cost, C, to a third party that designs the recommendation engine. In other words, for each successful purchase via recommendation, firm pays to the third party a percentage, x of the total price, P_R of recommended product, R. Firms choose to deploy recommendation engines on their platforms either to improve purchase likelihood by showing personalized, relevant product suggestions, or to increase sales diversity by increasing consideration sets of consumers, as they engage in online search. As the firm hosts a number of sellers/brands on its platform, the recommendation policy is a binary decision variable: to recommend or not, given prices are provided by suppliers individually. For the sake of simplicity, it is assumed that ex-ante, the firm knows the value of good i, V_i.

V_i is made of two components. One is the baseline utility derived from good i, which is constant for the entire mass of consumers. The other component varies as it is dependent on individual match quality that consumers ascertain through search and learning. This paper investigates specifically under which circumstances it is optimal for the firm to recommend, given the state of the consumer, $s \in \{H, L\}$. Consumers typically have varying degrees of trust on product recommendations offered by different platforms. The model assumes that in state H, consumer completes a purchase directly following a recommendation, while in state L consumer buys strictly via search. The state implicitly reflects the reputation of the firm recommender system to the consumer. Intuitively, if a recommender system consistently recommends the most relevant products to the consumer, then it would have a good reputation. In that case, recommendations would influence choice to a greater extent than would be the case if it often recommends irrelevant products. One of the central assumptions of the model based on Hosanager et al. (2008) (Hosanagar et al., 2008) is that, a consumer's satisfaction with the recommender system on a retailer website is assumed to be reflected in her purchase behavior. In other words, if the retailer offers relevant recommendations to consumers, they are more likely to click on them and ultimately purchase the recommended good, thereby developing a sense of trust in such recommendations. This, in turn, improves the reputation of the underlying recommender system over a period of time. Therefore, purchases through recommendations signal a higher reputation of the firm's recommender system as compared to purchases via search.

Purchase likelihood and firm's performance heavily depend on the state of the consumer, therefore, in designing optimal policy firms have to determine when it is profitable to incentivize consumers moving from one state to another. The current state of a consumer is determined by her action in the previous period: if she is in state H in period t and purchases the recommended good, in period $t + 1$ she remains in state H, however, if she purchases a good that was not a recommendation, with probability q, she transitions to state L. Similarly, if consumer is in state L in period t, purchases a good that was not recommended, she remains in state L and if she does purchase a recommended good, with probability q' she transitions to state H. q, q' are transition probabilities that represent the persistence of reputation of recommendations on the said platform. Without loss of generality, it is assumed $q = q' = 1$ which studies the extreme case when reputation effects are extremely persistent.

This means that when a consumer in state L views a product recommendation with a high perceived match quality, she immediately transitions to state H, while a consumer in state H immediately transitions to state L when presented with an irrelevant recommendation.

Recommendations are assumed to have a *salience effect*, δ_S (Hosanagar et al., 2008) which can be interpreted as a temporary boost in perceived match quality of a good. Now, consumer states are largely indicative of the relevance of recommendations, hence their reputation. Intuitively, if a recommendation engine structurally shows relevant recommendations, it will have a good reputation and in turn impact consumer choice to a larger extent than if it shows irrelevant recommendations. This leads to a key assumption $\delta_H > \delta_L$.

The utility of the consumer purchasing good i in period t is given as:

$$U_{it} = v_{i,t} + \mathbb{1}\delta_{s,t} + \epsilon_{i,t} \tag{4.1}$$

where $\mathbb{1} = 1$ if i is a recommended good and $\epsilon_{i,t}$ captures all the factors that affect utility, but are not observable to the firm.

It is assumed that $\epsilon_{i,t}$ is i.i.d type-I extreme value distributed, which represents the idiosyncratic tastes of consumers for good i observable just before purchase. Then following a multinomial logit specification, the probability of buying good i in period t can be expressed as:

$$Pr\{i, t|S\} = \frac{e^{v_{i,t}+\mathbb{1}\delta_{s,t}}}{1 + \sum e^{v_{i,t}+\mathbb{1}\delta_{s,t}}} \tag{4.2}$$

The firm's objective, as below, is to maximize the expected discounted future stream of profits with a decision policy that affects consumer choice per period:

$$\max_D E\Big(\prod [D]\Big) = E\sum_{t=1}^{\infty} \beta_t\Big(p_t - \mathbb{1}C_t\Big) \tag{4.3}$$

where $\beta \in [0, 1]$ is the discount factor, p the price and C the cost incurred by the firm to show recommendations on its platform, when it is following policy D.

Let $P(D)_{t,s,s'}$ be the probability of next period state being s' when current period state is s and $\pi(D)_{t,s}$ denote the current period profit from following policy D_t. Then lifetime profit functions in state $\in \{H, L\}$can be expressed as the following Bellman equations:

$$\prod_H(D) = \pi_H + \beta\Big[P_{HH}(D)\prod_H(D) + P_{HL}\prod_L(D)\Big] \tag{4.4}$$

$$\prod_L(D) = \pi_L + \beta\Big[P_{LL}(D)\prod_H(D) + P_{LH}\prod_L(D)\Big] \tag{4.5}$$

Recommendation policies have varying considerations depending on both present and future state of the consumer. The trade-off between firm revenue and relevance to buyer must be taken into account, while recommending in each state. In this framework, the firm has four exhaustive policies to consider:

Policy Specifications	
D1	Recommend in H Recommend in L
D2	Recommend in H Search option in L
D3	Search option in H Recommend in L
D4	Search option in H Search option in L

To determine equations (4.4) and (4.5), per period profits and transition probabilities are derived as follows:

Policy **D1**

$$\begin{aligned}
P_{HH} &= \frac{e^{V_R+\delta_H}}{e^{V_R+\delta_H} + e^{V}_{S} + 1}\\
P_{LL} &= \frac{e^{V_S} + 1}{e^{V_R+\delta_L} + e^{V_S} + 1}\\
\pi_H &= \frac{e^{V_R+\delta_H}(P_R - C_R) + e^{v_S}P_S}{e^{V_R+\delta_H} + e^{V_S} + 1}\\
\pi_S &= \frac{e^{V_R+\delta_L}(P_R - C_R) + e^{v_S}P_S}{e^{V_R+\delta_L} + e^{V_S} + 1}
\end{aligned} \tag{4.6}$$

Policy **D2**

$$\begin{aligned}
P_{HH} &= \frac{e^{V_R+\delta_H}}{e^{V_R+\delta_H} + e^{V_S} + 1} \\
P_{LL} &= \frac{e^{V_S} + 1}{e^{V_R} + e^{V_S} + 1} \\
\pi_H &= \frac{e^{V_R+\delta_H}(P_R - C_R) + e^{v_S} P_S}{e^{V_R+\delta_H} + e^{V_S} + 1} \\
\pi_S &= \frac{e^{V_R}(P_R - C_R) + e^{v_S} P_S}{e^{V_R} + e^{V_S} + 1}
\end{aligned} \tag{4.7}$$

Policy **D3**

$$\begin{aligned}
P_{HH} &= \frac{e^{V_S}}{e^{V_R} + e^{V_S} + 1} \\
P_{LL} &= \frac{e^{V_S} + 1}{e^{V_R+\delta_L} + e^{V_S} + 1} \\
\pi_H &= \frac{e^{V_R}(P_R - C_R) + e^{v_S} P_S}{e^{V_R} + e^{V_S} + 1} \\
\pi_S &= \frac{e^{V_R+\delta_L}(P_R - C_R) + e^{v_S} P_S}{e^{V_R+\delta_L} + e^{V_S} + 1}
\end{aligned} \tag{4.8}$$

Policy **D4**

$$\begin{aligned}
P_{HH} &= \frac{e^{V_S}}{e^{V_R} + e^{V_S} + 1} \\
P_{LL} &= \frac{e^{V_R} + 1}{e^{V_R} + e^{V_S} + 1} \\
\pi_H &= \frac{e^{V_R}(P_R - C_R) + e^{v_S} P_S}{e^{V_R} + e^{V_S} + 1} \\
\pi_S &= \frac{e^{V_R}(P_R - C_R) + e^{v_S} P_S}{e^{V_R} + e^{V_S} + 1}
\end{aligned} \tag{4.9}$$

Given a specific policy choice in state i, pair-wise comparisons are made between policy combinations in state j. In other words, we compare policies D1 and D3 in state H, when the firm definitely chooses to recommend in state L and D2 and D4 in state H, when the firm definitely chooses to not recommend or only have the search option in state L. This leads to the following conditions:

Condition 1: Given, firm always recommends in state L, it will recommend also in state H, if and only if $\prod_H(D3) - \prod_H(D1) < 0$

Condition 2: Given, firm never recommends in state L, it will recommend in state H, if and only if $\prod_H(D4) - \prod_H(D2) < 0$

From equation (4.4), we can express Condition 1 as:

$$\frac{\pi_H{}^{D3}(1-\beta P_{LL}{}^{D3}) + \pi_L^{D3}\beta P_{HL}{}^{D3}}{1-\beta(P_{HH}{}^{D3}+P_{LL}{}^{D3})+\beta^2(P_{HH}{}^{D3}P_{LL}{}^{D3}-P_{LH}{}^{D3}P_{HL}{}^{D3})} - \frac{\pi_H{}^{D1}(1-\beta P_{LL}{}^{D1}) + \pi_L^{D1}\beta P_{HL}{}^{D1}}{1-\beta(P_{HH}{}^{D1}+P_{LL}{}^{D1})+\beta^2(P_{HH}{}^{D1}P_{LL}{}^{D1}-P_{LH}{}^{D1}P_{HL}{}^{D1})} < 0$$

and Condition 2 as:

$$\frac{\pi_H{}^{D4}(1-\beta P_{LL}{}^{D4}) + \pi_L^{D4}\beta P_{HL}{}^{D4}}{1-\beta(P_{HH}{}^{D4}+P_{LL}{}^{D4})+\beta^2(P_{HH}{}^{D4}P_{LL}{}^{D4}-P_{LH}{}^{D4}P_{HL}{}^{D4})} - \frac{\pi_H{}^{D2}(1-\beta P_{LL}{}^{D2}) + \pi_L^{D2}\beta P_{HL}{}^{D2}}{1-\beta(P_{HH}{}^{D2}+P_{LL}{}^{D2})+\beta^2(P_{HH}{}^{D2}P_{LL}{}^{D2}-P_{LH}{}^{D2}P_{HL}{}^{D2})} < 0$$

The above conditions must be satisfied such that there exists a $[P_R{}^*, x^*]$ (where, $C_R = xP_R$), for which it is globally optimal for the firm to recommend in state H, irrespective of the choice in state L. For simplicity, it is assumed in this set-up that the firm knows consumers' initial state. As described so far, the payoff function of the retailer is dependent on three key variables in each state, namely, value of the good being sampled via search or recommendation (V_S, V_R), price of the good being sampled via search or recommendation (P_S, P_R) and the salience effect state (δ_s). In order to outline the optimal policy decision of the firm, the relationships between these variables are studied in detail, based on the pair-wise comparison of the policy combinations as per conditions (1) and (2). In that, few notable cases are examined to pin down the optimal price, value and salience thresholds computationally, both for state H and L

Case 1: $\mathrm{V}_S > V_R$

This is a trivial case. Firms in this scenario have no incentive to show recommendations for one of two reasons: 1) consumer derives lower utility from recommendations. This may be due to lower costs in time spent on search, or greater awareness of the firm's product offering; 2) there are costs associated with deploying a recommendation agent on the firm's retail platform. Typically every online platform invests in personalized recommendations either developing the algorithms in-house or commits to a subscription with external service providers. So, if consumers get higher utility from searching themselves and the firm needs to bear a cost of recommending, then it is not profitable to show recommendations.

Case 2a: $V_R > V_S; P_R > P_S$

When consumers derive higher utility from purchasing a recommended good and the corresponding price is sufficiently high, it is profitable for the firm to recommend. In this case, not only is the purchase likelihood high, but also the willingness to pay. This leads to the following proposition:

Proposition 1: When $V_R > V_S$, it is optimal for seller to recommend when P_R is sufficiently high, such that $\frac{P_R}{P_S} > \bar{P}$

where, $\bar{P}$= max[$\prod_H(D3) - \prod_H(D1)$, $\prod_H(D4) - \prod_H(D2)$], where D1, D2, D3 and D4 represent the set of exhaustive policy specifications for the firm. In other words, $\bar{P}$ is determined via the inequality conditions specified in **Condition 1** and **Condition 2**, wherein pair-wise comparisons are made between firm's payoff functions in state j, given specific policy choice in state i.

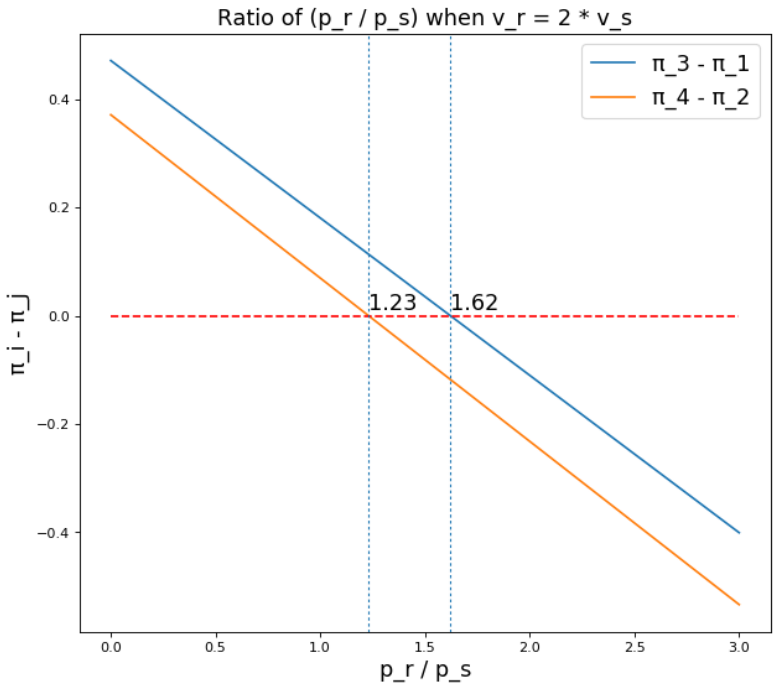
Figure 4.1 depicts **Condition 1** and **Condition 2** for Case 2a. It is evident that as P_R increases, the resultant expressions derived from the above conditions keep decreasing. By simulation, it is possible to derive the optimal value of $\bar{P} = 1.62$, that would make recommending profitable for the firm in either state, {H,L}. As firms improve the relevance of product recommendations, their value to potential shoppers increase. Relevant recommendations help convert potential buyers to paying consumers. However, firm's performance depends on the trade-off between value of the recommendation itself and its price. This trade-off is essentially quantified according to Proposition 1.

Case 2b: $V_R > V_S; P_S > P_R$

Proposition 2: When $P_S > P_R$, it is optimal for seller to recommend if and only if V_R is sufficiently large, such that $\frac{V_R}{V_S} > \bar{V}$

where, $\bar{V}$ can be obtained from **Condition 1**: $\prod_H(D3) - \prod_H(D1) < 0$ and **Condition 2**: $\prod_H(D4) - \prod_H(D2) < 0$. The result may seem counter-intuitive as recommending a lower priced good over a high priced good should never be profitable for the firm. However, a fraction of potential buyers value product recommendations only if they are able to get a better deal in terms of price

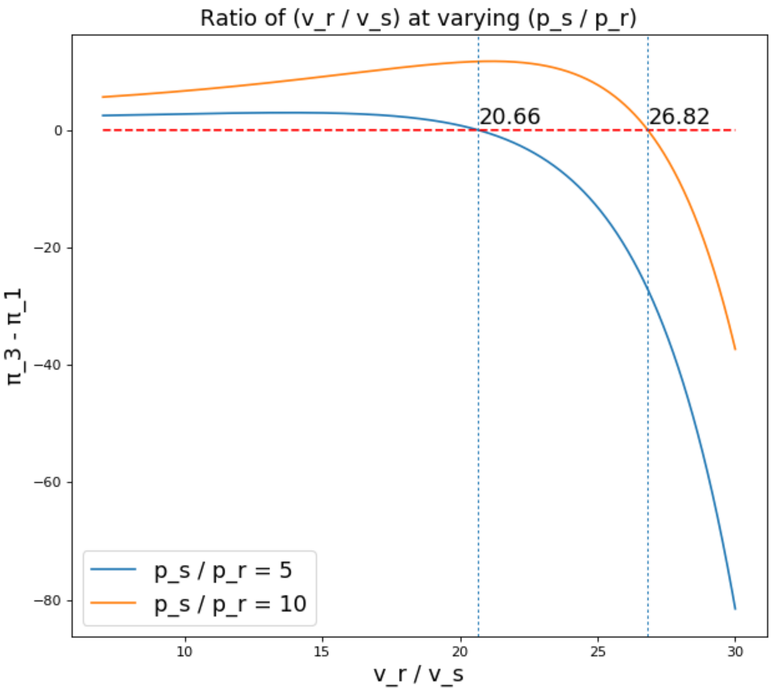
Figure 4.1: Case 2a



through a recommendation. Given that firms are incurring costs to show recommendations in order to boost sales and retain consumers who are highly likely to buy via recommendations, they may find it profitable to suggest a lower priced product if it guarantees a successful purchase. This is relevant especially if firms are in earlier stages of the maturity curve in the electronic commerce environment. During this phase improving consumer engagement on the platform is crucial and if value of recommendations is high enough for the shopper, ($\bar{V}$) that it increases likelihood of purchase significantly, then it would still be profitable for the firm.

Figure 4.2 and 4.3 respectively depict **Condition 1** and **Condition 2** for Case 2b. It is evident that as V_R increases, the resultant expressions derived from the above conditions keep decreasing. Interestingly, as the gap between the high priced option that the consumer samples via search and the low priced recommendation widens, $\bar{V}$ decreases at a slightly slower rate. For a very low P_R and a very high P_S, the threshold at which it starts becoming profitable for the firm to recommend is comparatively much higher. Simulating several scenarios, it is possible to pin down the optimal value of $\bar{V}$ = 20.66, that would make recommending profitable for the firm in either state when P_S/P_R =

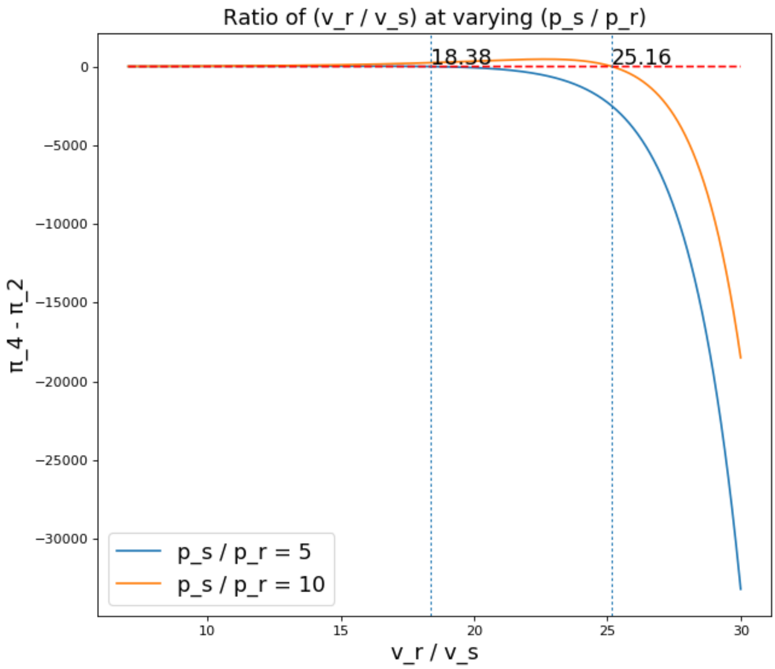
Figure 4.2: Case 2b: Comparing D3 and D1



5. In summation, when price of the recommended good is sufficiently high, recommending is always profitable. But even when the recommended option is lower priced compared to the searched option, it would still be profitable to recommended as long as value of the recommendation to the shopper is sufficiently high.

Given the model structure, it is arguably easier for the firm to make a choice in state H than L, as consumers will buy the recommended product in H, by definition. However, in state L, firm has no incentive to show recommendations if considering short-term gains only. In other words, irrespective of the price, recommending may be less profitable in terms of per-period profits. In state H, the per-period performance objective is dominant when deciding optimal firm policy, while in state L firm policy should ideally aim at restoring its reputation so as to optimize long-term profits, even if it is sub-optimal in the short run. This implies that if $\delta_H > \delta_L$, then firms might have incentive to potentially switch consumers from state L to H in the next period. So, the objective of the firm is two-fold: not only does it aim to improve performance via the recommendation policy, but also improve reputation of the firm in terms of relevance of its offering to consumers. As shown already,

Figure 4.3: Case 2b: Comparing D4 and D2



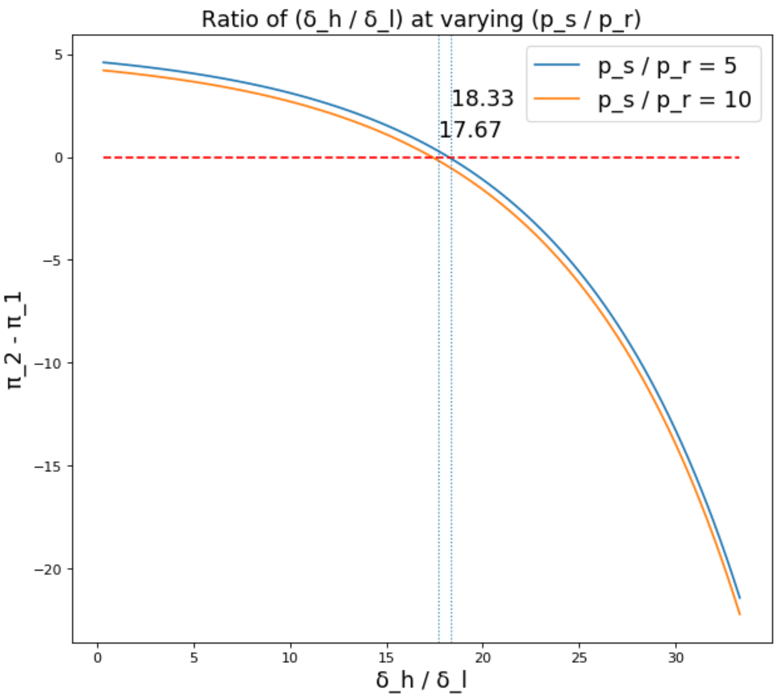
policies D1 and D2 are profitable in state H as compared to D3 and D4. Using similar methodology that leads to propositions 1 and 2, we now compare policies D1 and D2 in state L, when the firm definitely chooses to recommend in state H. This leads to the following condition:

Condition 3: Given, firm always recommends in state H, it will recommend also in state L, if and only if $\prod_H(D2) - \prod_H(D1) < 0$. Simulating several scenarios as per the following cases leads to Proposition 3.

Case 3a: $\mathrm{V}_R > V_S; P_R > P_S$

This case is trivial as the firm will always recommend if potential buyers derive higher utility from a product recommendation compared to searching themselves. Moreover if the recommended alternative is also priced higher, it is the obvious choice. A typical case when a shopper would opt for this alternative in the low state is if she becomes aware of a new product via recommendation. Firms often tend to increase sales diversity by recommending new products to potential buyers. In this case the firm not only has a higher current profit since $P_R > P_S$, but also higher future returns as $V_R > V_S$ would imply $\delta_H > \delta_L$.

Figure 4.4: Case 3b



Case 3b: $V_R > V_S; P_S > P_R$

Proposition 3: When $P_S > P_R$, it is optimal for seller to recommend if and only if δ_H is sufficiently large, such that $\frac{\delta_H}{\delta_L} > \bar{\delta}$.

where $\bar{\delta}$ can be derived from $\prod_H(D2) - \prod_H(D1) < 0$ as specified in **Condition 3**. The result may seem surprising at the outset. If the search option is higher priced and a shopper is in state L, then ideally the firm should not recommend. However, as objective of the firm is to maximize expected discounted future stream of per-period profits, therefore in some cases it may the optimal choice. By definition, $\delta_H > \delta_L$ as $V_R > V_S$, so there must be a δ_H such that if a shopper switches state form L to H, all future profits are high enough to compensate for the lower current profit. Figure 4.4 depicts exactly this phenomenon. As δ_H/δ_L increases, it becomes more and more profitable for the firm to recommend. Similar to the other cases, we are able to derive computationally the optimal value of $\bar{\delta}$ that would make recommending profitable for the firm in state L. However, as P_S/P_R increases, $\bar{\delta}$, or the level at above which recommending is profitable in state L decreases gradually. This is not surprising, as the limiting case of this would lead us to Case 3a where it is always optimal to recommend.

4.3 Data

Click stream data records complete search paths of online shoppers and is a powerful source of information on consumer behaviour in the internet markets. The dataset used for this study contains purchases made by shoppers, whose identity have been anonymized, between November 2015 and March 2016 on a leading Finnish multi-product electronic commerce platform. It is an "Amazon"-like platform for the Finnish market with a diverse assortment from several brands spanning across product categories such as electronics, home, clothing, cosmetics and stationary (Figure 4.7 provides the revenue distributions across categories in states H and L). The individual brands determine product pricing, however, it is the platform that controls the algorithm designing relevant product recommendations for shoppers based on their search histories. In other words, if a high priced product is recommended more frequently over its low priced alternative, it is likely to be an outcome of the platform's recommendation strategy as opposed to the individual brand's[3]. As a result, the dataset contains associated search histories leading up to purchases of unique shoppers, duration of visit, number of brands and products, price of goods and related product descriptions of each good viewed. Figure 4.5 looks at a typical example of the entire consumer journey in the data, from the point she clicks on the first good to the point of purchase(Basu, 2018). A key objective of this study is to investigate the reputation effects of platform generated recommendations, therefore shoppers with no visible purchase history are excluded from the analysis.

Figure 4.5: Clickstream data example: The point of purchase is highlighted in red

For the sake of relevance, browsing sessions of only those users who have made a purchase from the store have been considered. This is because consumers often engage in online search to obtain product prior to visiting the physical stores, where they eventually buy from. All purchases are observed along with the related browsing behaviour of the unique users, since their first arrival at the store. This means, that although all the non-purchase sessions were excluded from the sample, the entire search history was used for this study. So, the final sample including search histories mapped to purchases goes back to March 12, 2014. The richness of the data lies not only in that we are able to map unique agents' search behaviour to purchases, but also to derive insights on consumer preferences

[3]The data source cannot be revealed due to non-disclosure agreements

on brands, how that evolves over time, to what end they use information aids such as recommendations online, and their impact on buying behaviour. It allows us to examine search and purchase patterns of the representative agent across a diverse set of goods in a multi-product platform, while avoiding heterogeneity of preferences across stores offering substitutes. For each session that culminates into

Table 4.1: Descriptive statistics I

Data sample overview		
Number of H states in sample	14498	
Number of L states in sample	27853	
Total number of repeat buyers	11697	
% of repeat buyers	40.5%	
Number of unique products	68782	
Average search time	Mean	St. Dev.
Search time prior to purchase (in minutes)	11.31	19.69

a successful purchase, the data shows if the purchase was via search or a platform generated product recommendation. Recommendation agents cluster consumers based on similar preferences, search behaviour, demographics and several such factors to show a set of relevant products to inform and persuade consumers to buy. Figure 4.6 illustrates how recommendations typically appear on online platforms. For each product viewed, there are suggestions based on similar purchases:"Customers who viewed this item also viewed" or "What do customers buy after viewing this item".

The goal of this study is to investigate firm incentives to deploy recommendations on their online platforms when a fraction of shoppers have a clear preference to buy via search (state L) and the other via recommendations (state H). As shown in Table 4.2, the sample contains 1000 shoppers that are strictly in state H, 4500 strictly in state L and 6000 in both H and L. Interestingly, on average shoppers spend 9 minutes less in the store when in state L. This is not surprising as one of the objectives of recommendation engines is to enlarge consideration sets by informing consumers of relevant options available, thereby increasing time spent in store and in turn, improving likelihood of purchase. As supported by theoretical findings in Section 2, when costs are below a certain threshold it is optimal to recommend a relevant good even in state L. Improving reputation of the recommendations may implicitly improve performance in the long run, so it is in the firm's interest to show recommendations of high enough relevance that consumers switch from state L to H. In this particular sample, 65% of repeat buyers have switched between states at least once (Table 4.2).

Figure 4.6: Online recommendations example (source:amazon.co.uk): Top panel shows viewed book; bottom panels show relevant book recommendations

The data allows observing purchases made via clicks on recommendations. Table 4.3 summarizes the product distribution across states in the data sample and it can be observed that a higher percentage of the total revenue is generated from purchases via search (68%), however, purchases via recommendations on average are higher.[4] This implies although shoppers, may not follow recommendations as much but if a recommendation is highly relevant, they buy more on average. This clearly points to the challenge firms face today in realizing the value of allowing recommendations on online retail platforms.

---

[4]Retailer revenue is calculated as price multiplied by quantity of each product

Table 4.2: Descriptive statistics 2

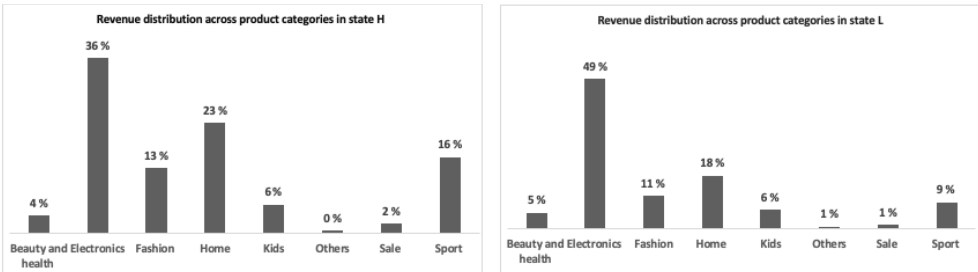
Distribution of repeat buyers in various states		
Number of repeat buyers exclusively in state H	977	
Number of repeat buyers exclusively in state L	4543	
Number of repeat buyers in both states H & L	6177	
% of repeat buyers who switch between states only once	65%	
% of repeat buyers who switch between states more than once	35%	
Average search times in various states	Mean	St. Dev.
Search time prior to purchase(in minutes) in state H	32.76	30.02
Search time prior to purchase(in minutes) in state L	23.32	20.43

Table 4.3: Descriptive statistics III

Product distribution in various states	
% of non-unique products purchased in state H	40%
% of non-unique products purchased in state L	60%
% Retailer revenue of non-unique products purchased in state H	32%
% Retailer revenue of non-unique products purchased in state L	68%
Average number of products purchased per session in state H	4.1
Average number of products purchased per session in state L	3.2

Figure 4.7 shows the revenue distribution across categories in both states. Generally speaking, reputation of platform generated recommendations do not have varying effects on revenue generated across product categories. As is observed from the data, electronic goods generate the highest revenue both via recommendations and search, followed by home goods. This is expected as these categories are likely to have higher prices on average compared to the rest. Interestingly, sport goods exhibit twice the revenue via recommendations than search, which is driven by volume, i.e., shoppers simply purchased more in this particular category via recommendations. Although, recommendations are positively associated with search for search goods, (such as, electronics) and negatively for experience goods (such as, beauty and health, there is no clear correlation with revenue. (Basu, 2018).

Figure 4.7: Revenue distribution by product categories



4.4 Empirical Analysis

In this section, several sources of empirical evidence on the effects of personalized recommendations on firm revenue are presented. Consider the following linear model in the population:

$$y = \alpha + \beta \mathbf{x} + \gamma s + \epsilon \tag{4.10}$$

where, for each purchase, y represents firm revenue, s, a binary variable that takes the value 1 if consumer was in state H at the purchase event, 0 if in state L and x, a k-dimensional vector of covariates. Among the three population parameters (α, β and γ), we focus primarily on γ which can be interpreted as the effect of consumer state, as defined in the theoretical model, on firm revenue.

As a first step in understanding the relationship between consumer states and firm revenue, we study the OLS estimates of a set of regressors $\mathbf{x} = [R, T, Q, I]$, where R = relevance of a recommendation, T = time spent on search, Q = number of searches and I = prior product knowledge. These variables are constructed from the browsing and purchase data described in Section 3. At each purchase session, T is expressed in seconds spent on the platform prior to purchase, while Q is the total number of unique page views during the session that leads to a successful purchase. R and I are derived from the data for the empirical model. In order to calculate I or prior product information from the data, we observe search paths of each shopper from their first arrival to the store till the penultimate session. At each search (product page view) there is transfer of information, only some of which is relevant to the buyer. So, in order to understand how information search influences purchase decisions, search events that do not potentially inform consumers must be filtered out. The idea is to collect only search events that help shoppers make the final choice. To this end, a measure of similarity is defined between products that have been viewed prior to the session of purchase and the purchased good, based on product descriptions (Basu, 2018).

4.4.1 Calculating measures for *Relevance* and *Prior product knowledge*

Each product has a description based on associated attributes or features. Entity-related search on web data is non-trivial, in the sense that the attributes are described in different ways for different products. The objective is to find similar products based on such attributes, hence its essential to find commonality between every word of the attribute vectors (combination of several keywords). A common method in the field of information retrieval, namely Tf-idf (term frequency-inverse document frequency) model is used here to mine relevant words from a large body of text or information set. It is often used as a weighting factor such that importance of a word increases proportionally to the number of times it appears in the document, but is offset by the frequency of the word in the entire collection. Let t_d be the frequency of the term t in a document d, and d'_t be the document frequency or the number of documents in which term t appears. The inverse document frequency means how important a term is in a collection, that is, how common it is across all the documents. Finally, let n be the total number of documents (product descriptions) in the data. Then the term frequency inverse document frequency weight, ω is calculated by,

$$\omega = t_d \times (d'_t)^{-1}$$

where,

$$(d'_t)^{-1} = log \frac{n}{d'_t}$$

Once the relevant words from the descriptions have been retrieved, a measure of similarity between the attributes is derived. The *cosine similarity* between two vectors or documents in the vector space is a measure that calculates the cosine of the angle between them. This is a comparison metric between two documents or product descriptions on a normalised space as it not only takes the magnitude of each word count into consideration, but also the angle between the documents. Let us consider vectors x and y, then the cosine of these two vectors is given by the Euclidean dot product:

$$x \cdot y = || x |||| y || \, cos \, \theta$$

$$S(x, y) = \frac{x \cdot y}{|| x |||| y ||}$$

Cosine similarity, (S(x,y)) gives how similar goods are based on their product descriptions. Observing similarity scores between products in several random samples of goods, only those products are collected from the browsing history that have a similarity coefficient of 0.4 or above (Basu, 2018). For each purchase session, the entire search history is observed in order to construct similarity scores for each sampled good with respect to the final purchase. Summing over all the similarity scores greater than 0.4 until the final purchase session gives a measure of prior product information, I. R or relevance of recommended products to a shopper is calculated with the same principle, but only restricted to the purchase session (not search history leading up to the session). Summing pairwise comparisons of each recommendation with the purchased item(s) during this session gives a measure of product relevance to shopper's preferences.

4.4.2 Consumer state and firm revenue: OLS estimates

The OLS estimates of firm revenue are reported in Table 4.4 Columns (1) and (2), along with robust standard errors. Column (2) includes the interaction terms of all control variables with the consumer state in order to study their combined effects on firm revenue. Evidently, this improves accuracy of the model. All coefficients except for prior product knowledge, I, are statistically significant at 1% or 5% level.

Recommendations play a critical role in the world of digital commerce and remains a key lever for firms to boost their revenue. The variable of interest in this study is, therefore, the consumer state, where state H indicates purchase via recommendations and L indicates purchase via search. As observed, consumer state being H or purchases via personalized recommendations is associated with a 29% increase in firm revenue. The positive relationship between revenue and personalized recommendations on the retail platform appears to be fairly robust. This effect on revenue clearly points to the incentives of the firm to show personalized recommendations to online shoppers.

Furthermore, as the theoretical model in Section 2 shows, net revenue from recommendations are dependent on the value shoppers derive from them and their willingness to pay a premium (purchasing a recommended good that is high priced). This is reflected in the OLS estimates, as relevance has a high positive correlation with revenue when shoppers are in state H (30%). These results clearly indicate the importance of personalization in recommendations, which is rather intuitive, as the likelihood of purchase depends on the shopper's match quality with the recommended product. Therefore, recommendations are profitable provided they are above a certain relevance threshold. As firms invest more in improving the relevance of recommendations, it implicitly increases perceived value, therefore consumers' willingness to pay for the right offering. This allows platforms to optimize cross-sell and upsell opportunities as well as highlight products with higher margins.

Additionally, a positive association between time spent on search and firm revenue is observed. More time spent on search likely affects purchase probability positively, hence online platforms implement several strategies to keep shoppers engaged with the goal of improving conversion rate. Surprisingly, the number of searches or pages (urls) viewed have the exact opposite relationship with revenue. With every additional product page viewed, revenue is likely to go down by approximately 7% in this model. As potential shoppers sift through product pages without spending much time on any page, it signals either recreational browsing or low buying intent. This implies that simply increasing consumer search time online does not necessarily imply higher firm revenue. The quality of search matters. Firms have to design user interfaces and provide relevant information in ways that engage potential buyers truly, it is not sufficient to simply increase traffic.

The OLS estimate of key variable as discussed above, measures the magnitude of association between consumer state and firm revenue, however not necessarily causation (Cameron and Trivedi, 2005). The goal is to estimate revenue impact to any exogenous changes in consumer state. However, whether consumers purchase via recommendations or only search is typically dependent on the reputation of firm-generated recommendations. As discussed extensively in both academic literature and media, firms often invest heavily into improving relevance of recommendations in order to show more and more relevant products to potential buyers. Over time it is likely that they develop a reputation on the relevance of their recommendations. Consumers' willingness to follow such recommendations will be influenced by their past experiences. If recommendations on a platform have a high reputation and trust, it will stimulate a higher conversion rate. As consumer engagement increases, likelihood of purchase also rises, in turn boosting firm revenue. As correlation between a firm's reputation and revenue is expected to be positive, so the estimate would likely be biased upwards. Therefore, the OLS estimate of consumer state alone will overstate the effect of showing recommendations, on firm revenue.

Let $Cov(x_i, \epsilon) = 0, \forall i = 1...k$ and $Cov(s, \epsilon) \neq 0$, such that we have an endogenous dummy variable model (Heckman, 1978). As the state variable is endogenous, OLS will inconsistently estimate parameter of interest, γ. There is both a direct effect via γ and an indirect effect via ϵ, which in turn impacts y. Instead of estimating the first effect only, in this case the OLS estimate combines both effects such that, $\hat{\gamma} > \gamma$, when both effects are positive. The estimate will be biased upwards if

Table 4.4: OLS estimates

	Dependent variable: Firm revenue	
	(1)	(2)
State (H=1)	0.228***	0.295***
	(0.020)	(0.011)
Relevance	0.195	0.227***
	(0.026)	(0.018)
Search time	0.139***	0.191***
	(0.061)	(0.056)
Number of searches	−0.067***	−0.069***
	(0.008)	(0.009)
Prior information	0.081	0.079
	(0.011)	(0.012)
Relevance-State		0.305***
		(0.023)
Search time-State		−0.198***
		(0.066)
Number of searches-State		0.042***
		(0.010)
Prior information-State		−0.088
		(0.015)
Constant	0.065***	0.073***
	(0.028)	(0.031)
Observations	40,544	40,544
R^2	0.335	0.378
Adjusted R^2	0.283	0.362
Residual Std. Error	0.017	0.015
Method	OLS	OLS with interaction terms
Time period (search and purchases)	2 years	2 years

*p<0.1; **p<0.05; ***p<0.01

consumers in high state are generating higher revenue and $Cov(s, \epsilon) > 0$. To gain further insight into the nature of the relationship between recommendations and firm performance, we use instrumental variables to isolate the effects of consumer state on firm revenue from other sources of variation.

4.4.3 Consumer state and firm revenue: Three-step IV estimates

The possible endogeneity of consumer state is dealt by means of the instrumental variables method. If there exists a valid set of instruments $\mathbf{z}$ for consumer state, it will ensure consistent estimation of γ by following the steps below (Renee Adams and Ferreira, 2009; Wooldridge, 2001):

1. Estimate a probit of the determinants of consumer state and obtain the fitted values $\hat{s_{i,t}}$

2. Regress $s_{i,t}$ on $\hat{s_{i,t}}$ and $x_{i,t}$, but not $z_{i,t}$ and $x_{i,t}$

3. Regress $y_{i,t}$ on $x_{i,t}$ and the fitted values of the second step

This procedure is applied to avoid any possibility of a *forbidden regression* which involves direct application of 2SLS to a non-linear model (Angrist and Pischke, 2009). Furthermore, it does not require the binary response model to be correctly specified in step one of the three step procedure(Renee Adams and Ferreira, 2009; Wooldridge, 2001). The above is a three step procedure which enables treating a binary endogenous variable with IV regression (Renee Adams and Ferreira, 2009).

Next, we discuss the economic arguments supporting validity of the instrument used, namely, *preferred state*. This variable is derived from historical transactional data of each unique shopper to the penultimate purchase session. It represents the ratio of the total number of purchases via recommendations and the number of purchases via search. This is believed to satisfy the conditions necessary for a valid instrument. The state of the representative agent in the current period is quite significantly influenced by her reputation of the firm's recommendation engine. So, if she has received a number of relevant product recommendations in the past that lead to successful purchases, then the likelihood of her being in state H at time t is higher than being in state L. Higher the reputation of firm generated recommendations, higher is the likelihood that consumers purchase via recommendations at the current time period. The motivation of choosing such a variable is two-fold: firstly, the probability that a consumer is going to be in state H at time t increases as the value of the variable, preferred state keeps increasing. The theoretical findings in Section 2 show that the price threshold for a firm to recommend, with the aim of switching potential buyers from L to H is significantly lower. In other words, it is easier, hence more attractive for firms to incentivize high state shoppers to remain in that state every period. Secondly, the instrument should be exogenous in the current framework as consumer's preferred state is unlikely to have a direct impact on revenue.[5]. Given that, first step of the IV estimation of the endogenous dummy variable model is as follows:

$$Pr(f = 1|\mathbf{x}, \mathbf{z}) = F(\sigma + \sigma' z + \sigma'' \mathbf{x}) \tag{4.11}$$

where F is the CDF for a standardized normal random variable, $\mathbf{z}$ is the instrument and $\mathbf{x}$ the vector of exogenous regressors. As discussed, the three-step IV procedure does not necessarily require this specification to be correct, however the instrument must be correlated to the probability of the consumer being in a certain state.

[5]Correlation between revenue and preferred state = 0.00046. Furthermore, this variable was included in the list of control variables for the OLS estimation and was found to be statistically insignificant.

From the probit estimates in Table 4.5 it is evident that the proposed instrument is highly significant and correlated with consumer state. Also, sign of the coefficient is consistent with our intuition, preferred state of a shopper which is the measure of reputation of firm generated recommendations, is positively related to current state of the consumer.

Table 4.6 reports the results of the instrumental variables regression. The model uses data on purchases between November 2015 and March 2016. Associated browsing or transaction history of unique shoppers go back to 2014 and the resultant sample has approximately 40,544 observations. As the main first step coefficients are statistically significant at 1% and 5 % level and the F-statistic is 12 (p-value=0.0005) we conclude that the instrument is strong (Stock et al., 2002). The coefficient for state is directionally comparable with the corresponding OLS estimate, however, it is much smaller in magnitude. This may suggest that there exists an upward bias in OLS estimates stemming from endogeneity of the consumer state. Intuitively, if the reputation of a platform's recommendations is improving, shoppers are more likely to purchase recommended products, as it reduces search cost of time. With more relevant recommendations, there is an increase in likelihood of shoppers switching from low state to high state. Consistent with proposition 2 in section 2, when recommendations are valued sufficiently by potential buyers, it is always optimal for a firm to recommend above a certain price threshold. Therefore, as reputation of recommendations improve, it has a positive impact on firm revenue.

Summarizing the results of the empirical analysis in sub-sections 4.2 and 4.3 as follows. Firstly, we find clear evidence that recommendations are significantly positively associated with firm revenue, as improving relevance of recommendations increases buyer's willingness to pay, thereby allowing firms to position high margin products as recommendations. This finding is consistent with the second key result that improving relevance of recommendations will have a positive impact on firm revenue. As consumers face alternatives recommended by the platform with improved match quality, additional utility derived over the functional value of the product is two-fold: one through reduced search cost of time and two higher salience effect which impacts purchase decisions. Thirdly, we find evidence on a positive relationship between search time and revenue which explains why firms use several tools in order to enhance shopper engagement online. The impact of search time on revenue can be explained via increase in aggregated conversion rates. We also find strong evidence that consumer state is endogenous in firm revenue regressions, and once the direct effect of consumer state on firm revenue is factored out, the remaining correlation between revenue and reputation of firm generated recommendations is significantly positive.

Table 4.5: Probit model

	Dependent variable:
	Consumer state
Preferred state	3.376**
	(0.027)
Relevance	1.931***
	(0.125)
Search time	2.016***
	(0.098)
Number of searches	-0.117
	(0.066)
Prior information	-0.298
	(0.115)
Constant	1.223***
	(0.021)
Observations	40,544
Method	Probit (Step 1 of 3-step IV)
Time period (search and purchases)	2 years
	*p<0.1; **p<0.05; ***p<0.01

4.5 Discussion

4.5.1 Summary of findings

This paper sheds light on several factors that influence the firm's optimal choice to invest in personalized product recommendations, with the aim of higher sales or increased performance. Additionally, increased relevance of personalized recommendations is shown to be positively associated with revenue. One of the key findings of this study is the unobserved reputation effects of firm-generated recommendations on consumers' willingness to pay. A three-step IV treatment allows isolating this effect from the set of control variables, which shows a significantly positive relationship between the firm's reputation and revenue. Furthermore, both OLS and the IV regressions show that if consumers prefer to purchase via recommendations (state H), it increases the firm's revenue at a higher rate than purchases via search alone (state L). This result explicitly points to the firm incentives not only to convert potential buyers from low to high state, but also to keep them in the high state.

The empirical results are supported by theoretical findings, where in state H it is always optimal for firms to recommend goods above a price threshold. Even if the recommended good is low-

Table 4.6: Three step IV estimates

	Dependent variable:
	Revenue
State	0.122***
	(0.010)
Relevance	0.202***
	(0.023)
Search time	0.258***
	(0.022)
Number of searches	−0.117***
	(0.037)
Prior information	0.139
	(0.034)
Constant	0.046**
	(0.029)
Observations	40,544
Weak instruments	57.87***
Wu-Hausman	2.44**
R^2	0.346
Adjusted R^2	0.319
Residual Std. Error	0.015 (df = 40538)
Method	Three-step IV regression
Time period (search and purchases)	2 years
	*p<0.1; **p<0.05; ***p<0.01

priced, is optimal to recommend this, as long as the consumer derives certain value from such a recommendation. Identifying this value or relevance threshold enables firms to maximize their long-term revenue, as increased relevance of the recommended good is shown to be significantly positively associated with revenue. It may seem unnecessary for firms to recommend when a potential buyer is in state L, as it generates lower per-period profits. However, the model shows that for a sufficiently high salience effect, it is always optimal to show recommendations. Based on the assumption that high state consumers experience a higher salience effect, it allows for the expected stream of future revenue to be high enough to compensate for the lower current period profits.

Finally, it is shown that the quality of online search is correlated with the firm's revenue. On the one hand, increased search time positively affects revenue, but, on the other hand, as the number of pages viewed increases, revenue diminishes. These findings clearly imply the type of traffic firms aim to generate on their retail platforms. Having a large number of page views does not necessarily boost revenue; however, improving the quality of information available for consumers, to accurately sample their match quality at every search, will improve the likelihood of a purchase and, in turn, increase the firm's revenue.

4.5.2 Future research directions

This paper presents a theoretical framework pinning down optimal price and value thresholds that allow firms to profitably show personalized recommendations. Furthermore, the empirical relationships between these variables and firm performance are studied in detail, accounting for unobserved heterogeneity via reputation effects. The aim is to provide a baseline study for future research to derive optimal price and relevance thresholds, empirically, which should be valuable for e-commerce platforms in designing long-term recommendation algorithms. It is of further interest to characterize these thresholds while controlling for the consumer state, which make recommending the optimal choice for the firm especially when reputation effects are as significant.

Additionally, consumers may exhibit contrasting behaviour when exposed to personalized recommendations for different product categories. Controlling for product groups such as, experience versus search goods and studying the varying impact of reputation effects would enable a deeper understanding of consumer motivations and how firms may differentiate strategies across different product groups.

4.5.3 Managerial Implications

As retailing has moved towards digital commerce, persuasive advertising via recommendation agents has become rather commonplace. Online platforms with large and diverse assortments additionally invest in long-run recommendation policy designs to maximize their revenue. The key considerations in that have been studied here. Firstly, is there a trade-off between the relevance of a recommendation and the price of the recommended good? Online retailers must determine if personalized recommendations are of sufficiently high value to the consumer to justify the associated price point.

Secondly, the trade-off between short-term profitability and the discounted stream of future revenue is studied in detail. This is especially relevant for platforms that have a low reputation in the current period. Their objective is two-fold: first, to build a reputation with consumers by showing high value or relevant product recommendations. Second, to recommend following a long-run profit maximizing strategy which, as shown in this paper, may lead to a seemingly sub-optimal outcome in the short run. The empirical findings point to the positive impact of reputation effects on the firm's performance, which is likely driven by a higher willingness to pay.

Thirdly, it can be observed from the theoretical model that the price thresholds in state L are much lower as compared to state H. This is because in state L firms have the additional task of building a reputation with shoppers, as it is positively related to revenue. However, improving the reputation of recommendations will improve the firm's performance to a degree, beyond which it may not yield sufficiently high marginal value to compensate for the low price. It is this threshold that firms must consider in order to effectively guide consumer search and choice, that maximizes profitability in the long-run.

References

Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005. ISSN 1041-4347.

Gediminas Adomavicius, Jesse C. Bockstedt, Shawn P. Curley, and Jingjing Zhang. Effects of online recommendations on consumersâ willingness to pay. *Info. Sys. Research*, 29(1):84â102, March 2018.

Franklin Allen. Reputation and product quality. *RAND Journal of Economics*, 15(3):311–327, 1984.

Xavier Amatriain and Justin Basilico. Past, present, and future of recommender systems: An industry perspective. In *Proceedings of the 10th ACM Conference on Recommender Systems*, RecSys â16, page 211â214. Association for Computing Machinery, 2016.

Joshua Angrist and JÃ¶rn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. 01 2009. ISBN 9780691120348 (hardcover : alk. paper).

Shreya Basu. Information search in the internet markets: Experience versus search goods. *Electronic Commerce Research and Applications*, 30, 05 2018.

Shreya Basu. Personalized product recommendations and firm performance. *Electronic Commerce Research and Applications*, 48, 2021.

Alexander Benlian, Ryad Titah, and Thomas Hess. Differential effects of provider recommendations and consumer reviews in e-commerce transactions: An experimental study. *Journal of Management Information Systems*, 29(1):237–272, 2012.

Luis Cabral and Ali Hortacsu. The dynamics of seller reputation: Evidence from ebay. *The Journal of Industrial Economics*, 58(1):54–78, 2010.

A.C. Cameron and P.K. Trivedi. *Microeconometrics: Methods and Applications*. Cambridge University Press, 2005. ISBN 9781139444866.

Pei-Yu Sharon Chen, Shin-yi Wu, and Jungsun Yoon. The impact of online recommendations and consumer feedback on sales. In *International Conference on Information Systems (ICIS)*, pages 711–724, 2004.

Angeline G Close and Monika Kukar-Kinney. Beyond buying: Motivations behind consumers' online shopping cart use. *Journal of Business Research*, 63(9-10):986–992, 2010.

P.H. Dybvig and C.S. Spatt. *Does it Pay to Maintain a Reputation?* Financial Research Center memorandum. Princeton University, Department of Economics, 1980.

Daniel Fleder and Kartik Hosanagar. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science*, 55(5):697–712, 2009.

Gerald Haubl and Kyle B. Murray. Preference construction and persistence in digital marketplaces: The role of electronic recommendation agents. *Journal of Consumer Psychology*, 13(1-2):75–91, 2003.

John R Hauser. Consideration-set heuristics. *Journal of Business Research*, 67(8):1688–1699, 2014.

James J. Heckman. Dummy endogenous variables in a simultaneous equation system. *Econometrica*, 46(4):931–959, 1978.

Thorsten Hennig-Thurau, AndrÃÂ© Marchand, and Paul Marx. Can automated group recommender systems help consumers make better choices? *Journal of Marketing*, 76(5):89–109, 2012.

Johannes Hörner. Reputation and competition. *American economic review*, 92(3):644–663, 2002.

Kartik Hosanagar, Ramayya Krishnan, and Liye Ma. Recommended for you: The impact of profit incentives on the relevance of online recommendations. volume 31. AIS, 2008.

Daniel Houser and John Wooders. Reputation in auctions: Theory, and evidence from ebay. *Journal of Economics & Management Strategy*, 15(2):353–369, 2006.

Benjamin Klein and Keith B. Leffler. The role of price in guaranteeing quality. UCLA Economics Working Papers 149, UCLA Department of Economics, 1979.

Benjamin Klein and Keith B. Leffler. The role of market forces in assuring contractual performance. *Journal of Political Economy*, 89(4):615–641, 1981.

David M Kreps and Robert Wilson. Reputation and imperfect information. *Journal of Economic Theory*, 27(2):253 – 279, 1982.

Nanda Kumar and Izak Benbasat. Research note: The influence of recommendations and consumer reviews on evaluations of websites. *Information Systems Research*, 17(4):425–439, 2006.

George J Mailath and Larry Samuelson. Who wants a good reputation? *The Review of Economic Studies*, 68(2):415–441, 2001.

Paul Milgrom and John Roberts. Predation, reputation, and entry deterrence. *Journal of Economic Theory*, 27(2):280 – 312, 1982a.

Paul Milgrom and John Roberts. Limit pricing and entry under incomplete information: An equilibrium analysis. *Econometrica*, 50(2):443–59, 1982b.

Girish Punj and Robert Moore. Information search and consideration set formation in a web-based store environment. *Journal of Business Research*, 62(6):644–650, 2009.

Heitor Almeida Renee Adams and Daniel Ferreira. Understanding the relationship between founder ceos and firm performance. *Journal of Empirical Finance*, 16(1):136 – 150, 2009.

William P. Rogerson. Reputation and product quality. *Bell Journal of Economics*, 14:508–515, 1983.

Sylvain Senecal and Jacques Nantel. The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80(2):159 – 169, 2004.

Carl Shapiro. Consumer information, product quality, and seller reputation. *Bell Journal of Economics*, 13(1):20–35, 1982.

Joseph E. Stiglitz. Imperfect information in the product market. volume 1 of *Handbook of Industrial Organization*, pages 769 – 847. Elsevier, 1989.

James Stock, Jonathan Wright, and Motohiro Yogo. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business Economic Statistics*, 20(4): 518–529, 2002.

Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*, volume 1 of *MIT Press Books*. The MIT Press, February 2001.

Bo Xiao and Izak Benbasat. E-commerce product recommendation agents: Use, characteristics, and impact. *MIS Quarterly*, 31(1):137–209, March 2007.

4.6 Appendix

Figure 4.8: Posterior information across different products over the sample period

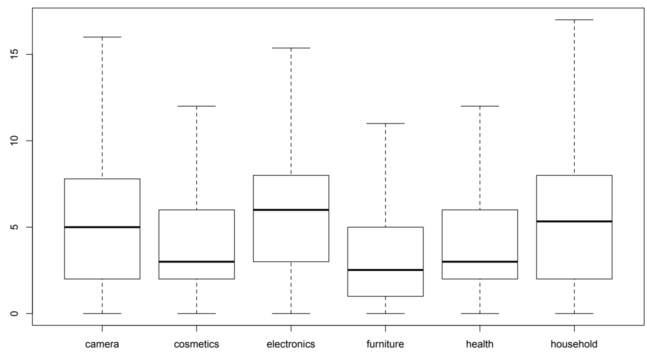


Figure 4.9: Distribution of prior knowledge

Table 4.7: Estimates of the number of searches prior to purchase for product categories: OLS

	Dependent variable: Number of searches					
	(Camera)	(Cosmetics)	(Furniture)	(Healthcare)	(Household appliances)	(Personal electronics)
Price	−0.021	−0.021***	0.053*	−0.015**	0.005***	0.031***
	(0.004)	(0.001)	(0.001)	(0.004)	(0.001)	(0.003)
Prior information	0.124**	−0.207	0.184***	−0.162	−0.096*	0.213***
	(0.017)	(0.013)	(0.012)	(0.015)	(0.009)	(0.014)
Variance	−1.162***	−4.334***	0.245***	−3.038***	−2.006***	−2.727***
	(0.149)	(0.903)	(1.633)	(0.849)	(0.337)	(0.410)
Recommendation quality	−0.134	−0.168***	0.024	−0.026**	−0.076	−0.014
	(0.045)	(0.020)	(0.011)	(0.016)	(0.012)	(0.011)
Recommendation (1=used:;0=not used)	0.274***	−0.342***	0.147***	− 0.341***	0.238***	0.230***
	(0.016)	(0.084)	(0.049)	(0.059)	(0.027)	(0.042)
Constant	2.571***	3.122***	2.425***	3.079***	3.174***	2.640***
	(0.134)	(0.089)	(0.050)	(0.060)	(0.026)	(0.040)
Observations	97	278	452	1,795	866	521
Adjusted R^2	0.341	0.356	0.315	0.306	0.384	0.312

Note: *p<0.1; **p<0.05; ***p<0.01