Convincing online consumers to purchase

Empirical studies on online advertising, mobile advertising, user generated content and social shopping tools

Sascha Leweling

Convincing online consumers to purchase: Empirical studies on online advertising, mobile advertising, user generated content and social shopping tools

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften durch die Wirtschaftswissenschaftliche Fakultät der Westfälischen Wilhelms-Universität Münster

vorgelegt von

Sascha Leweling, M.Sc.

Dekanin:	Prof. Dr. Theresia Theurl
Erstgutachter:	Prof. Dr. Thorsten Wiesel Institut für Wertbasiertes Marketing Westfälische Wilhelms-Universität Münster
Zweitgutachter:	Prof. Dr. Kay Peters SVI-Stiftungsprofessur für Dialogmarketing Universität Hamburg
Tag der mündlichen Prüfung:	20.04.2017

Sascha Leweling

Convincing online consumers to purchase

Wissenschaftliche Schriften der WWU Münster

Reihe IV

Band 15

Sascha Leweling

Convincing online consumers to purchase

Empirical studies on online advertising, mobile advertising, user generated content and social shopping tools

Wissenschaftliche Schriften der WWU Münster
herausgegeben von der Universitäts- und Landesbibliothek Münster
http://www.ulb.uni-muenster.de

Die Publikation ist Teil der Dissertationsreihe des Instituts für Wertbasiertes Marketing.

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;
detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Dieses Buch steht gleichzeitig in einer elektronischen Version über den Publikations- und
Archivierungsserver der WWU Münster zur Verfügung.
http://www.ulb.uni-muenster.de/wissenschaftliche-schriften

Sascha Leweling
„Convincing online consumers to purchase. Empirical studies on online advertising, mobile advertising,
user generated content and social shopping tools "
Wissenschaftliche Schriften der WWU Münster, Reihe IV, Band 15
Verlag readbox publishing GmbH – readbox unipress, Münster
http://unipress.readbox.net

Zugl.: Diss. Universität Münster, 2017

ISBN 978-3-8405-0184-5 (Druckausgabe)
URN urn:nbn:de:hbz:6-68109734122 (elektronische Version)

direkt zur Online-Version:

Satz: Sascha Leweling
Titelbild: Sascha Leweling
Umschlag: ULB Münster

Foreword

Online shopping is common for us nowadays. While we as consumers mostly favor our more or less unlimited shopping opportunities, those unlimited opportunities mean several challenges for marketers. One of the biggest challenge is how to stand out and generate attention for the firm’s product or the service in this cluttered digital environment. Display advertising is a common way to overcome this challenge. Display ads are served in the traditional online channel (using PCs and laptops) and the mobile channel (using smartphones and tablets). These channels differ in usage situation and consumer behavior and those differences should be taken into consideration when doing display advertising. Besides these ways of getting customer to visit one’s website, marketers also need to be aware of the fact that consumers like a more social shopping environment (e.g., by relying on other consumers’ opinions) when making purchase decisions.

This dissertation, entitled “Convincing online consumers to purchase: Empirical studies on online advertising, mobile advertising, user generated content and social shopping tools”, by Sascha Leweling focusses on three core research questions that are extremely relevant for both academia and practice: (1) Are display ads more effective if they match the consumer's interest in the product? (2) Do consumers react differently to display ads presented on mobile devices compared to conventional devices (e.g., desktop computers)? (3) Do consumers spend more money when they use purchase aids such as user generated content or social shopping tools? After introducing and presenting the overall framework of the dissertation, the following three chapters cover three empirical studies aimed at answering the above three research questions.

Thereby, Study 1 addresses the first research question. Advertisers face the problem that they mostly cannot “observe” how interested the consumer is in the advertised product. A good display ad should match the communication goal of the advertiser but also match the interest of the consumer. Sascha Leweling proposes that based on the unobservable interest of a consumer different ad characteristics (i.e., ad message and ad format) should be served and different targeting options should be used. To estimate this unobservable interest, Sascha developed a Hidden Markov Model and examines the transitions between certain interest states and how ad characteristics and targeting options influence these transitions based on a dataset of an American travel and tourism company. Based on the results, the implications for managers are that advertisements that correspond to the consumer's state of interest lead to improved online advertising performance.

Study 2 addresses the second research question. Mobile marketing is on the rise and consumers visit the internet more often with mobile devices. Research in the mobile channel is sparse. But researchers agree that the mobile channel is conceptually different from the traditional online channel. But, what do those differences actually mean for your mobile marketing campaigns? Therefore, Sascha conducted a quasi-field experiment in order to examine whether or not the behavioral outcome of consumers and consumers' reaction to display advertising is different when a consumer visits the website with a mobile or a conventional device (e.g., desktop computer). In doing so, he shows randomized display ads to the visitors of the website and analyzes the differences in customer reactions. He finds that click through rates for mobile ads are lower than for desktop devices. However, the location of the advertisement on the website affects this result – advertisements further down the website get more frequently clicked on mobile devices opposed to desktop devices. There is no significant difference between ads at the top position between these two channels.

The final research question is addressed with Study 3. Using the online channel as a distribution channel faces retailers and manufacturers with two challenges: Consumers using the online channel cannot evaluate the products before-hand and, moreover, shopping lacks the social experience of offline stores. Retailers try to overcome these limitations by offering user generated content and social shopping tools in their online shops. But, what is the effect of the usage of these two tools on customer revenue and customers' return behavior? Sascha finds that customers who use user generated content produce higher gross revenues for the retailer. Surprisingly, these customers also have higher return rates. However, they still generate significantly higher net revenues for the retailer. He also finds that social shopping tools do not affect the customer's revenue on average. Therefore, retailers should stimulate the use of user generated content because active users seem to be more loyal to the company and buy more. Furthermore, active consumers provide important information for passive consumers that they can use for their purchase decision.

Overall, the dissertation by Sascha tackles a very interesting, relevant, and timely topic. Thereby, Sascha offers many useful answers to the above mentioned main research questions and uses highly sophisticated methods to derive rigor managerial implications. This dissertation is a great piece of academic work that is very relevant to the business word. I can only applaud Sascha for his great achievement and wish all readers a lot of fun and insights reading the dissertation.

Finally, I want to thank Sascha for his hard work at the Institute for Value-Based Marketing (IWM) at the Marketing Center Münster (MCM) at the University of Münster. Sascha was one of the first two PhD students we supervised in Münster. Together with Simon, the two assisted us a lot in making the transition from the Netherland to Germany and building the institute. We together developed our "Start-Up" and I truly believe it is and will continue to be a huge success – not possible without the help, engagement, and time of Sascha. While Simon was a "take-away" from Groningen, Sascha studied in Münster and hence was a very valuable source for us in order to get acquainted with Münster, the student body, and the new colleagues – lots of interesting insights! All of us learned a lot during that time at Münster, whereas sometimes also the hard way. During that time, we developed a kind of theme that there are two things we would love students to take away with them from Münster: (1) *Roots* in terms of an excellent academic and practically relevant education and the being together at the IWM and (2) *Wings* in terms of free and critical thinking in an environment that allows making mistakes and where own creative ideas are more than welcome. This then leads hopefully to a long lasting relationship, true to the motto: "PhD Student for a few years, Alumnus for a life!"

Sascha, thank you very, very much for the time together in Münster! I hope that you got your roots and use your wings now – I am very optimistic that you do so.

Prof. Dr. Thorsten Wiesel

Münster / July 11, 2018

Danksagung

Die vorliegende Arbeit entstand während meiner Tätigkeit am Institut für Wertbasiertes Marketing. Als ich im Oktober 2012, nach meinem Masterstudium, am Institut von Thorsten Wiesel anfing, waren wir ein kleines Team von fünf Personen. Als ich im Dezember 2016 meine Arbeit am Institut beendete, um die finalen Züge der Dissertation fertigzustellen, haben mich viele Leute auf dem Weg begleitet, bei denen ich mich im Folgenden bedanken möchte.

Zu allererst möchte ich mich bei meinem Doktorvater Thorsten Wiesel bedanken. Durch intensive Diskussionen, vor allem in den Doktorandenseminaren, wurde die Arbeit zu dem was sie heute ist. Im gleichen Zuge danke ich Sonja Gensler, die immer neue Ideen hatte und immer für Fragen zur Verfügung stand. Ich danke euch beiden für das ständige Hinterfragen meiner Ideen und Ansätze. Ich habe viel in der Zeit am IWM gelernt. Ferner danke ich auch meinem Zweitgutachter Kay Peters, welcher das Ende des Promotionsprozesses mitbegleitet hat. Durch ein frisches Paar Augen wurden die letzten Inkonsistenzen aufgedeckt, welches mir noch einmal geholfen hat, die Arbeit zu verbessern. Zu guter Letzt danke ich Theresia Theurl, welche meiner Disputation beigesessen hat und mich mit vielen Fragen herausgefordert hat.

Neben den Gutachtern der Dissertation möchte ich mich auch bei den weiteren Mitarbeitern des Instituts bedanken. Als erstes ist natürlich Simon Monske zu nennen, mit dem ich zusammen angefangen und zusammen aufgehört habe. Man kann gar nicht stark genug hervorheben, wie wertvoll ein „Leidensgenosse", gerade in der Endphase der Dissertation ist. Nicht weniger Dank gebühren Charlotte Hufnagel, Lisette de Vries, Jane Johne und Fabian Kraut. Durch euch ist das Institut erst das geworden, was es ist. Durch die vielen Nebenaktivitäten, vor allem mit unseren HiWis, wurde die Arbeit am Institut erst das was es ein soll: eine familiäre Atmosphäre. Auch ihr wart es, die immer ein offenes Ohr hatten, wenn es mit dem Schreiben mal wieder nicht so lief, wie es sollte. Weiterhin möchte ich auch Christian Werner, Birte Geue und Maris Hartmanis danken, welche frischen Wind in das Institut brachten. Zu guter Letzt danke ich natürlich Gabriele Rüter, die es immer geschafft hat Ordnung in unsere administrative Arbeit zu bringen.

Auch allen weiteren Mitarbeitern des Marketing Center Münster möchte ich danken. Hervorzuheben sind hier Maria Bartschat, Kim Kötterheinrich und Michael Zylla mit denen eine Kaffeerunde immer lustiger wurde.

Da aber Arbeit nicht alles ist, danke ich natürlich auch meinen Freunden, die mich immer gut im Dissertationsstress abgelenkt haben. Vor allem Fabian Ehring, Julian Böhnert, Max Rickert, Theo Georgiu, Lena Strukamp, David Rüschenschmidt und Tim Ebner sind hier besonders hervorzuheben.

Natürlich will ich mich auch bei meiner Familie bedanken, welche mich im gesamten Prozess begleitet und gefördert hat. Danke an meine Mutter Angelika, meiner Tante Bettina, meinem Onkel Heinz und natürlich meinen beiden Omas Hildegard und Gertrud. Natürlich danke ich auch meinem Vater Robert, der viel zu früh verstorben ist.

Nicht zu vergessen ist aber die Unterstützung von Laura Purschke, welche mich vor allem in der Endphase der Dissertation ertragen musste und das hervorragend gemeistert hat. Ohne deine Unterstützung und aufbauenden Worte wäre der Weg viel schwieriger gewesen.

Danke!

Köln, 10.07.2018

Sascha Leweling

Table of contents

List of figures

List of tables

1 Introduction

Today, consumers spend more and more time on the internet browsing, communicating with others or searching for information (e.g., Ofcom 2015). Given that trend, firms are wondering how to use the internet to acquire customers. Generating attention for and interest in a firm's offerings is, however, difficult for marketers because the internet is cluttered and consumers have limited cognitive resources (Pieters et al. 2002, 2007). Nevertheless, marketers have invested heavily in online display advertising in the last years and continue to do so (Statista.com 2017). Especially, targeting approaches for display advertising have received a lot of attention recently (Statista.com 2014). Yet, previous research suggests that targeting approaches might only be effective when they consider a consumer's interest in the advertised product or service. This insight is valuable but little surprising. However, to consider a consumer's interest in a firm's offering when displaying online ads is challenging. Marketers cannot observe a consumer's interest in a firm's offering; rather a consumer's interest is latent (i.e., not observable). Yet, if marketers were able to derive a consumer's interest and consider this interest when displaying online ads, they might be able to improve the effectiveness of online display ads. Such a more fine-tuned display of online ads might also address current criticism from consumers as well as from marketers with respect to display advertising (Hoskins 2013).

When using display advertising, marketers also face the challenge that consumers use various devices to go online such as desktop computers, laptops, tablets and mobile phones (Ratcliff 2014). These devices differ, among other dimensions, with respect to the usage situation (e.g., in public vs. private, crowded vs. uncrowded environments, rather silent vs. noisy environments) and the screen size. Moreover, previous research suggests that consumers behave differently when using different devices (Ghose, Goldfarb, et al. 2013). When marketers use online display advertising to attract customers, the device consumers are using might moderate the effectiveness of such ads. However, previous research provides little insights into the influence of the device on consumers' reactions to online display ads. Having such knowledge would enable marketers to improve the effectiveness of online display ads on various devices.

If marketers succeed in attracting consumers to their website and they also offer the opportunity to buy online, they face the challenge to stimulate purchases. Only because the consumer is interested in an offering and attracted to the website, does not ensure that he or she will actually purchase and keep the product. Nowadays, firms offer user generated content (e.g., product reviews) and social shopping tools to facilitate the purchase decision (Zhu et al. 2010). However, we have little knowledge about the effectiveness of such information and tools. Previous research focuses on the effects of user generated content on aggregate product sales and the design of social shopping tools rather than on the individual-level effects on customers' purchasing behavior and customer revenue.

To sum up, marketers face three critical challenges today:

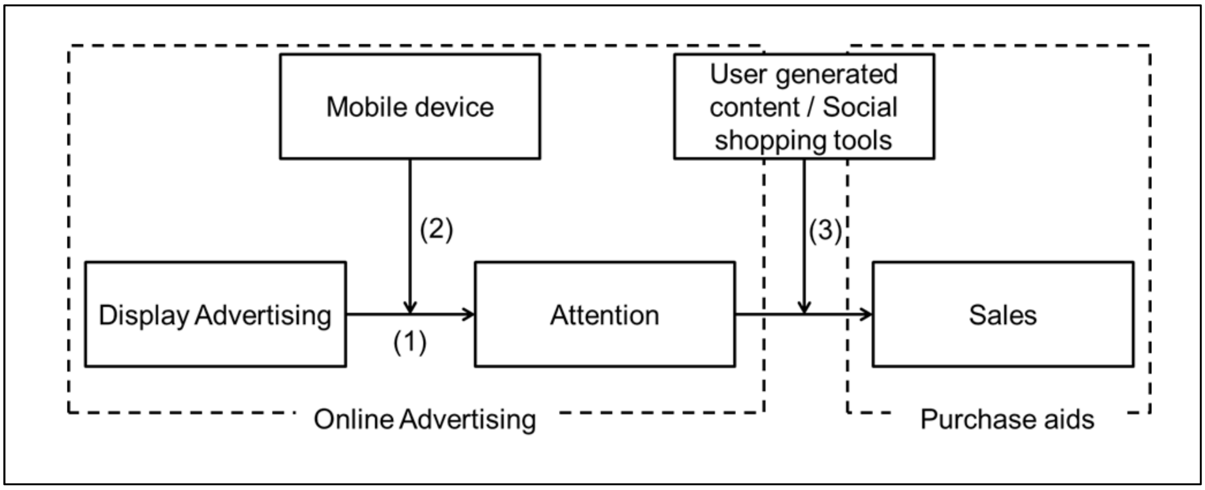
1. Improving the effectiveness of online display advertising by serving ads that stimulate a consumer's latent interest in the firm's offering.

2. Improving the effectiveness of online display ads by serving ads that match the device a consumer is using.

3. Offering information such as user generated content and social shopping tools that facilitate purchase decisions and affect the value generated from customers positively.

These challenges are the focus of this dissertation and are depicted in Figure 1-1.

Figure 1-1: Outline of dissertation



The first challenge deals with serving relevant advertising impressions with the goal to generate interest. This means that an ad should stimulate a consumer's interest and affects a consumer's behavioral response to an ad positively. But websites are cluttered and it is challenging to create consumer interest through online display advertising (Cho and Cheon 2004; Drèze and Hussherr 2003). A problem for marketers is that they do not know how interested a consumer is in the advertised product or service. To create interest and a favorable behavioral response (e.g., click, visit), many marketers use obtrusive and highly targeted ads that stand out on a webpage (Moses 2014). Targeting uses information about consumers' profiles or browsing behavior (Bleier and Eisenbeiss 2015; Bruce et al. 2016; Goldfarb and Tucker 2011; Lambrecht and Tucker 2013; Trusov et al. 2016). But ultimately, consumers might be annoyed by those kinds of ads. Marketers need a good understanding what ad to serve to a specific consumer at a certain point in time to effectively create interest and a favorable behavioral response (Urban et al. 2014). This challenge leads to the first research question, which will be answered in Chapter 2 of this dissertation:

RQ1: Can marketers improve the effectiveness of online display advertising by serving ads that match a consumer's latent interest in the firm's offering?

To address this research question, we develop a Hidden Markov Model to estimate a consumer's latent interest in the firm's offering. We use a dataset consisting of clickstream data from an American travel and tourism company. In total, we have information about approximately 80 million cookies with over 300 million impressions. We draw a sample out the dataset to make the estimation feasible. We find that different ad messages, ad formats and targeting approaches vary in their influence on consumers' latent interest and behavioral response. We find that the used targeting approach has the biggest influence on consumers' interest state and behavioral response. The ad format also has differential effects but interestingly the ad message only has an effect when the consumer has little interest in the firm's offering, but not when he/she has a high interest. These insights offer valuable managerial implications which will be outlined in Chapter 2 of this dissertation.

Yet, consumers not only go online with desktop computers but also with mobile devices (Ratcliff 2014). Studies show that consumers browsing behavior is affected by the device they are using (Ghose, Goldfarb, et al. 2013). For example, search costs are higher when using mobile devices, implicating that consumers process more information that is placed at the beginning of the website. Furthermore, mobile devices imply unique features such as portability (i.e., screen size) and they are used in different situations than desktop computers (Grewal et al. 2016). Marketers need to be aware of the differences and design their advertising in a way that it fits to the device a consumer is using.

However, marketers often serve the same ads on the devices (Patel et al. 2013). That is, there is little to no differentiation between serving display ads on desktop computers compared to mobile devices. We also have little academic knowledge about the effectiveness of ads displayed on desktop/laptop computers and mobile devices. On a product level, products that feature high involvement and are of utilitarian nature seem to be more suitable for mobile display advertising (Bart et al. 2014). On a campaign level, marketers should serve both channels simultaneously because the probability that a consumer will click on an ad is higher (Ghose, Han, et al. 2013). But so far, there is no study addressing the question whether the device used by the consumer affects the effectiveness of display ads. To the best of our knowledge, there is no study that compares the behavioral outcome of display ads comparing the mobile channel and the traditional channel leading to the second research question of this dissertation:

RQ2: Can marketers improve the effectiveness of online display ads by serving ads that match the device a consumer is using?

To address this research question, we design a quasi-experiment and investigate differences in consumers' click behavior considering the device used by the consumer. We find that in general consumers click less on display ads when using a mobile device. However, a display ad further down the website increases the probability to be clicked up to a certain amount of impression for the mobile device compared to the traditional device. There is no significant difference between ads at the top of the page between these two types of devices.

Display advertising aims to create awareness and to attract consumers to a firm's website. Once they have the attention and the consumer visits the website the question is what makes them purchase the product. To help consumers to make a purchase decision, most online retailers offer additional information next to product information. Such additional information may be product reviews and Q&A boards (i.e., user generated content) but also social shopping tools like collaborative shopping and sharing opportunities. These information and tools are supposed to address the lack of information about products and the lack of social experiences in an online environment. As such, it is critical for marketers to know whether these information and tools actually affect individual consumers' purchasing behavior. Previous literature on user generated content shows that products reviews affects product sales (Babić Rosario et al. 2016; Floyd et al. 2014). However, these studies implicitly assume that consumers consider user generated content if it is available on a webpage. This appears to be a rather strong assumption and we, thus, have no knowledge whether the actual consideration of user generated content affects consumers' purchasing behavior. Moreover, the literature on social shopping tools focusses on the design of such tools and does not study the impact of such tools on consumers' purchasing behavior (Kim et al. 2013; Zhu et al. 2010). These two gaps in current knowledge lead to the third research question of this dissertation:

RQ3: Do user generated content and social shopping tools facilitate purchase decisions and affect customer revenue positively?

To address this research question, we use survey data and transaction information from a large Dutch online retailer for a sample of more than 2,000 customers. We consider potential self-selection effects and find that customers who consider user generated content are more profitable for the retailer although these customers also return more. Customers who contribute content generate higher net revenues than customer who just consume user generated content. This result highlights the value of customers who are actually contributing content on a firm's website for the firm. These customers improve the value of the website for other consumers and are the better customers because they generate high revenues. We further find that social shopping tools have little effect on customers' shopping behavior. We summarize the studies of this dissertation in Table 1-1.

Table 1-1: Overview of the studies in this dissertation

Study title	Aim	Data	Method	Results
Effect of ad characteristics and targeting options on display ad effectiveness	Assessing the effect of display ad characteristics and targeting options on consumers' behavioral response considering the consumers' unobservable interest in the product	Clickstream data of an American travel and tourism company	Bayesian Hidden Markov Model	• Two different interest states • Ad characteristics and targeting options work different based on the interest state • Taking the interest state into account leads to higher click rates and website visits
User device and ad response: The moderating role of ad position	Addressing the differences in consumers' reaction to ads served on traditional and mobile devices considering ad position	Quasi experiment on a German nutrition website using clickstream data	Endogeneity corrected probit model	• Different click behavior between mobile and online channel • For mobile websites, advertisers should also take the ad slot further down a website into account
Value of user generated content and social shopping tools	Examining the effect of customers' use of user generated content and social shopping tools on customer revenue and return behavior.	Transaction and survey data from a Dutch online retailer	Propensity score matching	• User generated content has a positive effect on both, gross and net sales • Social shopping tools have no influence alone but work as a complementary tool to user generated content in the fashion category

2 Effect of ad characteristics and targeting options on display ad effectiveness

2.1 Introduction

Marketers who use online display advertising to create interest in their offerings have to consider what ad is most relevant for a specific consumer to create interest at a certain point in time. The ad message (e.g., brand-related, sales-related messages), ad format (e.g., obtrusive vs. non-obtrusive ads) and also the targeting approach (e.g., content integration, retargeting) influence an ad's relevance for a consumer. However, an ad's relevance also depends on a consumer's actual interest in the firm's offerings which is affected by previous encounters with the firm. However, a consumer's interest in the firm's offerings is not observable and, thus, latent.

Clickstream data can provide some information of a consumer's interest in a firm's offerings. Moreover, the ad message, ad format and the targeting approach might enhance a consumer's interest. However, certain ad messages, ad formats and targeting approaches might even decrease a consumer's interest in the firm's offering. Yet, our knowledge about the influence of ad messages, ad formats and targeting approaches on consumer's interest and behavioral response (i.e., clicks and visits) is limited. Previous research has either focused on ad messages (Braun and Moe 2013; Drèze and Hussherr 2003), ad formats (Goldfarb and Tucker 2011) or targeting approaches (e.g., Bleier and Eisenbeiss 2015; Ghose and Todri 2016; Hoban and Bucklin 2015; Lambrecht and Tucker 2013; Trusov et al. 2016). The only exception is Bruce et al. (2016) who consider ad message, ad format, and the targeting approach when studying the effectiveness of display advertising. Moreover, previous research has not paid much attention to the role of a consumer's interest in the firm's offerings so far. The main approach for including consumer's interest is using observable metrics, like previous website visits. For example, advertisers use retargeting when the consumer visited an online shop (Bleier and Eisenbeiss 2015; Lambrecht and Tucker 2013). Another approach is defining consumers interest by using past observable behavior metrics like previous website visits or registering at the website (Hoban and Bucklin 2015). But so far no study included consumer’s interest as latent. Moreover, no study let the interest be a variable that can change based on different encounters with the firm.

It is the aim of this study to assess the effect of ad messages, ad format and targeting approaches on consumers’ latent interest in the firm’s offerings and behavioral response. We use a unique dataset from an American travel and tourism company. The data contain information about individual ad exposures. Therefore, we have records on each impression, each click (and therefore a resulting website visit) and website visits that were not directly caused by a click on the respective ad. Moreover, we have information about the targeting approach (i.e., content integration, content amplification, behavioral targeting, retargeting), the type of ad message (i.e., brand-, product- or sales-related message) and whether the ad was obtrusive or not. Furthermore, we have click information on other digital advertising media like search and email. These forms of digital advertising are usually conducted by different agencies and it is common to only have click information and no information about the impressions (Abhishek et al. 2012; Ghose and Todri 2016).

Using this dataset, we develop a Hidden Markov Model (HMM) with the interest of the consumer as latent states. We are the first study that combines the effect of ad messages, ad format and targeting approaches considering a consumer’s latent interest state. Furthermore, we add encounters with other advertising formats like search and email as covariates. We contribute to research by modeling the consumer’s interest state as unobservable and use only clickstream data to estimate the interest state. Our contribution to practice is giving guidelines with respect to the influence of ad messages, ad format and targeting approaches on consumers’ interest and behavioral response. Therefore, we are able to provide implications what ad to serve to a certain consumer at a certain point in time.

The remainder of this chapter is as follows. The next section provides a literature review of online display advertising. Afterwards the HMM is introduced. After the description of the available data, we show the results and provide managerial implications. This chapter closes with limitations of the model and future research possibilities.

2.2 Literature review

This literature review considers current knowledge about the influence of different ad messages, ad formats and targeting approaches on consumer's latent interest in the firm's offerings and, ultimately, consumers' behavioral reactions to online display advertising (i.e., clicks and visits; Abhishek et al. 2012; Braun and Moe 2013; Li and Kannan 2014; Xu et al. 2014).

Effects of *ad messages* are well researched in traditional advertising like TV advertising (e.g., MacInnis et al. 2002). Research about online advertising focusses on different types of ad messages such as brand-related (Drèze and Hussherr 2003), product-related (Bruce et al. 2016) and sales-related messages (Bruce et al. 2016) and their effects on clicks or brand recall. Yet, the studies do not take into account that an ad's relevance depends on a consumers' interest in the firm's offerings. A first indication that different messages can have different effects can be found in Braun and Moe (2013). They find that it could be more effective to serve different ads to consumers to increase the click probability. However, they do not differentiate between the messages. Thus, we have little knowledge about the effectiveness of different ad messages.

The *format* of a display ad can be divided into obtrusive ad formats and non-obtrusive (standard) ad formats. Since 2003, the Interactive Advertising Bureau (IAB) publishes display advertising creative format guidelines to facilitate the serving of display ads (IAB 2015). Obtrusive formats usually have a higher viewability (e.g., interstitials which interrupt the browsing experience). Non-obtrusive ad formats have been investigated by previous research (Bruce et al. 2016; Goldfarb and Tucker 2015; Kuisma et al. 2010). The effect of obtrusive ads is rather sparse with the notable exception of Goldfarb and Tucker (2011). The authors used survey data to evaluate the effect of obtrusive display ads on purchase intention. However, they did not consider the effect of obtrusive ads on consumers' actual response.

The targeting option determines *where* the consumer sees the ad and, more importantly, *what* consumers see the ad. Whereas ad message and ad format are part of the design process, the targeting decision determines the audience. Early research shows that a precise targeting can have a positive effect for advertisers (Chatterjee et al. 2003). Later studies support this finding (Hoban and Bucklin 2015; Trusov et al. 2016). Among other targeting options, four strategies are commonly used: content integration, content amplification, retargeting and behavioral targeting.

With content integration, the content of the ad fits the content of the website (Goldfarb and Tucker 2011). The underlying idea is that the audience of the website fits the audience that the firm wants to see the ad. A previous field experiment shows that this strategy increases purchase intention. Consumers that are browsing the website are more prone to the ads because their mindset is already triggered for certain information. Hence, the firm's ads are less seen as annoying advertising but more as useful information. However, the ads should not be too obtrusive since this reverses the effect on purchase intention (Goldfarb and Tucker 2011).

Content integration targets a whole audience of a website. A similar approach is to target websites where the desired audience most likely spends their leisure time. This strategy is called content amplification. It is different from content integration in a way that it does not serve the ad that it matches the content of the website but matches the interest of the desired target group (e.g., the targeted audience is more likely to spend their time on the same news websites). This kind of targeting option was not part of previous research; hence this is the first study to include this option.

The previous mentioned targeting options, content integration and content amplification, show similar characteristics to classic TV advertising. Advertisers define a target audience and want to serve the ads in places where the consumer is most likely to be. However, online technologies allow more detailed targeting on a personal level, called micro-targeting. A common used micro-targeting option is retargeting (Bleier and Eisenbeiss 2015; Lambrecht and Tucker 2013). When a consumer visited a website or browsed a product in an online shop, the advertiser has the information that the consumer has been in contact with the product before. They assume that the consumer is interested in the product and show him or her related ads. Advertisers expect retargeting to increase the effectiveness of display ads. However, current research shows that surprisingly such micro-targeted ads are ineffective on average (Lambrecht and Tucker 2013). They become more effective when the consumer has a clear idea of what to buy, which is determined by the consumer's activity on review websites (Lambrecht and Tucker 2013). Moreover, retargeting varies by the degree of personalization which also affects consumer’s response to those ads (Bleier and Eisenbeiss 2015).

In recent years, more advanced micro-targeting strategies have evolved. Another approach is behavioral targeting where the goal is to predict the consumer profile based on the previous browsing history that is stored in a cookie of the computer. Based on the predicted profile, the ad with the highest click probability will be served. In contrast to retargeting, the consumer does not need to have previous encounters with the brand. Very few research is conducted with respect to this targeting option with the exception of Trusov et al. (2016). They predict user profiles using topic modeling and third-party data (which serves as a proxy for consumer’s interest). Furthermore, they show that the click-through rate increases when a company uses such profiles for targeting. However, the use of third party data is not always feasible for a company (e.g., because of legal concerns), therefore another approach is needed for estimating the consumer’s interest.

In general, previous research has largely ignored the mediating role of consumer’s interest in the advertised product. However, Abhishek et al. (2012) show that a consumer’s interest state has an impact on his/her behavior, meaning that the behavioral response to display advertising depends on the interest state. Other research includes consumer’s interest with observable metrics. This means they observe consumer behavior and based on the observation they specify how interested the consumer might be in the product. Hoban and Bucklin (2015) build a purchase funnel and differentiate between non-visitors, visitors, authenticated users and converted customers. They find display ad cause positive behavioral outcomes in most stages, but not for visitors who did not create an account. Ghose and Todri (2016) also include a funnel-like specification in their study that focuses on observables. They connect the touchpoints (advertising exposure, website visit etc.) for each consumer and construct a purchase funnel. Their main goal is to get insights about the attribution of different advertising media rather than optimize the serving of display advertising. However, they find that exposure to display advertising might increase the interest in the advertised product. But they did not split the ad characteristics in ad message and ad format and did not specify interest as a latent construct. Xu et al. (2014) model the path to purchase from consumers and how different ad formats (e.g., display, search) influence the consumer. Again, the focus is not on display advertising, but more on attribution to various digital media.

All these studies drop the fact that interest can also diminish over time. These studies use a classical purchase funnel that narrows down over time. In contrast, we focus on evolution of interest that can dynamically change after every impression. Moreover, the advertiser does not know how interested the consumer is when sitting in front of the computer. Previous observable metrics like browsing on related websites or related search queries might suggest that interest is present, but this can have changed over time. We are the first study that uses interest as a latent construct that can be influenced by ad message, ad format and the way the consumer is targeted.

Table 2-1 shows the current literature in display advertising, and demonstrates that the effect of different ad messages, ad formats and targeting approaches on consumers' interest and behavioral response is not well researched yet. We, thus, have little knowledge what the influence of ad messages, ad format and targeting approaches are on consumer's interest and, ultimately, their behavioral response to display ads. This chapter aims to close this gap.

Table 2-1: Literature review of display advertising

Authors	Ad message	Obtrusive ads	Targeting	Considering consumer interest
Abhishek, Fader and Hosanagar (2012)	-	-	-	Conversion funnel modeled with hidden Markov model. Influenced by display and search ads
Bleier and Eisenbeiss (2015)	-	-	Retargeting	-
Braun and Moe (2013)	15 unique banner creatives, but to clear focus on ad message	-	-	-
Bruce, Murthi and Rao (2016)	Product and price related messages	Flash and GIF ads in three formats (but no obtrusive formats)	Retargeting	-
Chatterjee, Hoffman, and Novak (2003)	-	-	Consumers must have been exposed to ads from two advertisers in the experiment	-
Drèze and Hussherr (2003)	Brand related messages	-	-	-
Ghose and Todri (2016)	-	-	Retargeting, prospecting, affiliate advertising	Observable: constructing a consumer specific funnel based on observable clickstream data

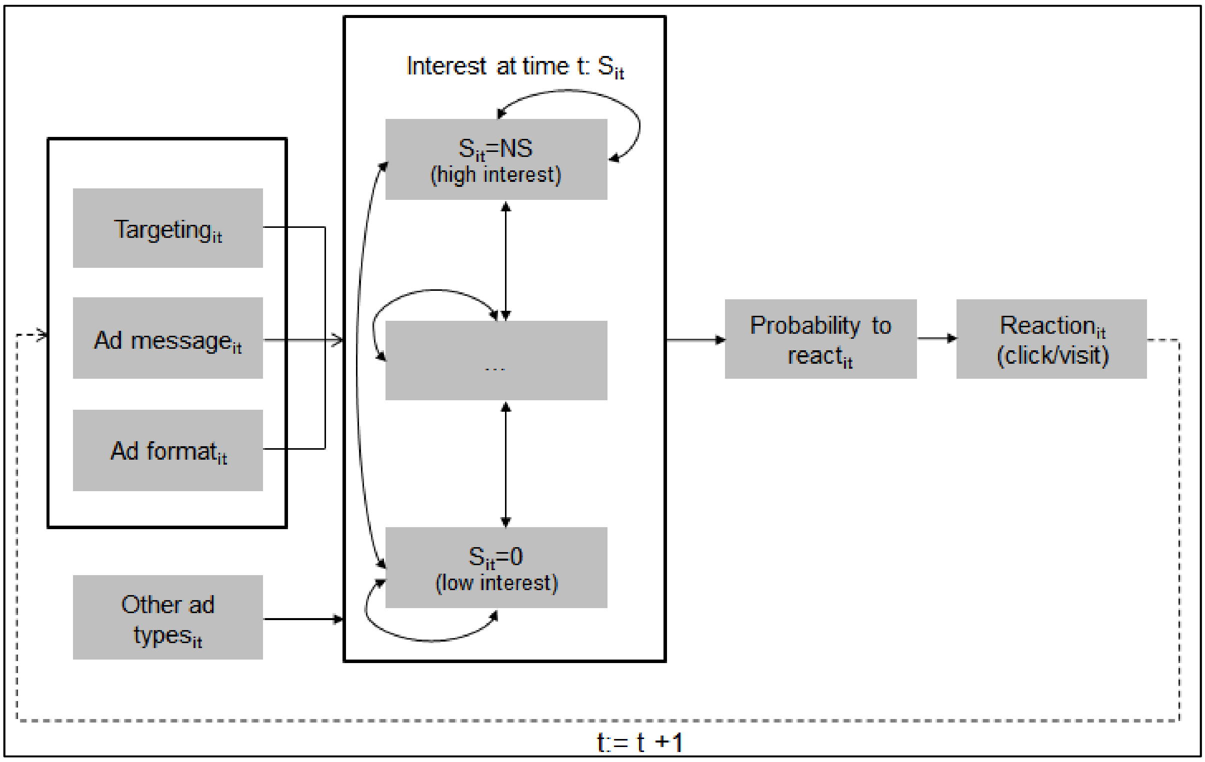
Table 2-1 (continued): Literature review of display advertising

Authors	Ad message	Obtrusive ads	Targeting	Considering consumer interest
Goldfarb and Tucker (2011)	-	Obtrusive formats (e.g. Pop-ups, Interstitials etc.)	Content integration	-
Hoban and Bucklin (2015)	-	-	Retargeting	Observable: effects on non-visitors, visitors and authenticated users
Lambrecht and Tucker (2013)	-	-	Retargeting	-
Trusov, Ma and Jamal (2016)	-	-	Behavioral targeting	Third party data: estimating user profiles via topic modeling
Xu, Duan and Whinston (2014)	-	-	-	Observable: constructing a consumer specific funnel based on observable clickstream data
This chapter	Brand related, product related, and sales related messages	Obtrusive formats (e.g. Pop-ups, Interstitials etc.)	Content integration, content amplification, retargeting, behavioral targeting	Unobservable: Latent consumer interest modeled with hidden Markov model. Interest states are influenced by ad message, ad format and targeting options.

2.3 Conceptual model and modeling approach

In this study, we consider a consumer's latent interest in a firm's offering and the influence of ad messages, ad format and targeting approaches on this latent interest. A consumer's interest then affects his/her behavioral response to a display ad. A behavioral response can be a click on the ad (which results in a website visit) or a later direct visit on the website. Figure 2-1 shows our conceptual research model.

Figure 2-1: Conceptual model



At any given time *t* when an ad is served to consumer *i*, a consumer updates his/her interest in the firm's offerings. Since every consumer is different, the updating process varies across consumers. Three outcomes are possible for the updating process. First, an ad might increase a consumer's interest in the firm's offering; a transition to a higher interest state occurs which, finally, results in a higher probability to react to the ad. Second, the ad does not affect a consumer's interest state and, thus, probability to react to the ad. Finally, the ad results in a lower interest in the firm's offerings and the probability to react to the ad decreases. At a new timestamp *t+1* this process repeats until the advertiser is not showing a related ad to the focal consumer. The update process is influenced by the ad message, the ad format and the targeting approach the advertiser chooses. The targeting option also influences where (i.e., on which website) the consumer sees the ad. To control for other ad types, such as search and email, we add covariates that indicate whether the consumer was in contact with the brand before display exposure. As mentioned, when serving display ads, the advertiser (if at all) has only information about search and email clicks, not about the number of impressions (Abhishek et al. 2012; Ghose and Todri 2016). The effect of previous ads is implicitly stored in the current interest state. For example, if previous ads had a negative impact on consumer's interest, the processing of a new ad at any given time *t* is different compared to a different, more positive interest at that time. With this approach, we close the gap in the literature: An ad is evaluated by the consumer, given his/her interest. The interest is influenced by ad format, ad message and the targeting approach. The targeting determines which consumer sees the ad and where the consumer sees the ad. The interest is determined on clickstream data alone and no additional third-party information is needed.

The latent interest states can either be continuous or discrete. Continuous states have the disadvantage that we cannot observe clear rules when a consumer switches the interest state. A discrete number of states is, thus, preferred, and we can use a Hidden Markov Model (HMM) to estimate the latent interest states. HMMs are emerging in marketing (Kumar et al. 2011; Luo and Kumar 2013; Montoya et al. 2010; Netzer et al. 2008; Schweidel et al. 2011). They are used to identify hidden states based on observable data. In our case, we see whether a consumer interacted with the brand by either clicking on an ad or by visiting the website.

An HMM is determined by three components: the initial state membership, the transition matrix and the probability to react. The initial state membership represents the starting point in the analysis. The transition matrix determines the switching probabilities between the hidden states. In our case, ad messages, ad formats and targeting approaches influence the transition probabilities. The probability to react determines how likely it is that the consumer shows a favorable behavioral response (i.e., click or visit). The next chapters explain the modeling of the three components in more detail.

2.3.1 Initial state probabilities

The initial state probabilities are needed to identify the model. Since the states are latent, there are two approaches to solve this issue. Let π_{is} denote the initial state membership, where i identifies the individual and s represents the hidden state. Since π_{is} are probabilities the following two conditions hold: $\pi_{is} > 0$ and $\sum_{s=1}^{NS} \pi_{is} = 1$, where NS represents the number of states. There are several procedures to identify the initial state probabilities. A common approach is to use the stationary distribution of the transition matrix (MacDonald and Zucchini 1997, p. 79). The stationary distribution can be calculated by solving $\boldsymbol{\pi_i} = \boldsymbol{\pi_i} \boldsymbol{\bar{Q}_i}$, where $\boldsymbol{\bar{Q}_i}$ is the transition matrix with all the covariates set to zero. Since we already know that $\sum_{s=1}^{NS} \pi_{is} = 1$, we can replace the last equation by this expression and use a standard procedure for solving linear equations.

The easiest way is to set π_{i1} to 1 and the remaining π_{is} to 0. This means that all individuals start at state 1 which represents a low interest in the firm's offering. This approach also helps us to model an interest path where we start at a low interest state and observe transition behavior over time. The modeling of the transition matrix follows next.

2.3.2 Transition matrix

The transition matrix determines the probability to switch the latent interest state. The general form of the transition matrix is:

$$\boldsymbol{Q_{i,t\text{-}1\to t}} = \begin{bmatrix} q_{it11} & q_{it12} & q_{it13} & \cdots & q_{it1NS-1} & q_{it1NS} \\ q_{it21} & q_{it22} & q_{it23} & \cdots & q_{it2NS-1} & q_{it2NS} \\ q_{it31} & q_{it32} & q_{it33} & \cdots & q_{it3NS-1} & q_{it3NS} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ q_{itNS1} & q_{itNS2} & q_{itNS3} & \cdots & q_{itNSNS-1} & q_{itNSNS} \end{bmatrix} \tag{1}$$

with $q_{itss'}$ representing the transition probabilities of state s to state s' of individual i in time t and NS represents the number of latent interest states. Similar to recent research (e.g., Netzer et al. 2008), we let our variables of interest (i.e., ad messages, ad format, and targeting approaches) influence the transition probabilities.

Additionally, we also control for the exposure to other ad types (e.g., search, email). Marketers have only information whether a consumer has clicked on other ad types or not previously, but lack information on the number of impressions. We, therefore, only consider dummy variables whether a click on another ad type has occurred or not. Thus, the probability to switch from state *s* to state *s*' can be written as

$$q_{itss'} = P(X_{it} = s' \mid X_{it-1} = s, \boldsymbol{a_{it}}, \boldsymbol{b_{it}}, \boldsymbol{c_{it}}), \tag{2}$$

where $\boldsymbol{a_{it}}$ are the ad characteristics (i.e. ad message and ad format), $\boldsymbol{b_{it}}$ the targeting options and $\boldsymbol{c_{it}}$ the covariates controlling for other ad exposures. Since $q_{itss'}$ are probabilities, it has to be fulfilled that $q_{itss'} > 0$ and $\sum_{s'=1}^{NS} q_{itss'} = 1$.
Defining T_s as the set of states that can be the subsequent state s' given state s the transition probabilities follow an ordinal logit model (Montoya et al. 2010; Netzer et al. 2008):

$$q_{its1} = \frac{\exp(\tau_{is1} - \boldsymbol{a}'_{it}\boldsymbol{\rho}_s - \boldsymbol{b}'_{it}\boldsymbol{v}_s - \boldsymbol{c}'_{it}\boldsymbol{\gamma}_s)}{1 + \exp(\tau_{is1} - \boldsymbol{a}'_{it}\boldsymbol{\rho}_s - \boldsymbol{b}'_{it}\boldsymbol{v}_s - \boldsymbol{c}'_{it}\boldsymbol{\gamma}_s)} \tag{3}$$

$$\begin{aligned} q_{its2} = {} & \frac{\exp(\tau_{is2} - \boldsymbol{a}'_{it}\boldsymbol{\rho}_s - \boldsymbol{b}'_{it}\boldsymbol{v}_s - \boldsymbol{c}'_{it}\boldsymbol{\gamma}_s)}{1 + \exp(\tau_{is2} - \boldsymbol{a}'_{it}\boldsymbol{\rho}_s - \boldsymbol{b}'_{it}\boldsymbol{v}_s - \boldsymbol{c}'_{it}\boldsymbol{\gamma}_s)} \\ & - \frac{\exp(\tau_{is1} - \boldsymbol{a}'_{it}\boldsymbol{\rho}_s - \boldsymbol{b}'_{it}\boldsymbol{v}_s - \boldsymbol{c}'_{it}\boldsymbol{\gamma}_s)}{1 + \exp(\tau_{is1} - \boldsymbol{a}'_{it}\boldsymbol{\rho}_s - \boldsymbol{b}'_{it}\boldsymbol{v}_s - \boldsymbol{c}'_{it}\boldsymbol{\gamma}_s)} \end{aligned} \tag{4}$$

$$\vdots$$

$$q_{itsNS} = 1 - \frac{\exp(\tau_{isNS-1} - \boldsymbol{a}'_{it}\boldsymbol{\rho}_s - \boldsymbol{b}'_{it}\boldsymbol{v}_s - \boldsymbol{c}'_{it}\boldsymbol{\gamma}_s)}{1 + \exp(\tau_{isNS-1} - \boldsymbol{a}'_{it}\boldsymbol{\rho}_s - \boldsymbol{b}'_{it}\boldsymbol{v}_s - \boldsymbol{c}'_{it}\boldsymbol{\gamma}_s)}, \tag{5}$$

where τ_{isNS} are the state specific intercept terms for each individual i, $\boldsymbol{\rho_s}$ is the vector of parameters capturing the effect of the ad characteristics (i.e., ad message and ad format), $\boldsymbol{v_s}$ is the vector of parameters capturing the effect of the targeting options and $\boldsymbol{\gamma_s}$ is the vector of parameters controlling for other ad types. To avoid having too many parameters to estimate, we let the parameter vectors $\boldsymbol{\rho_s}$, $\boldsymbol{v_s}$ and $\boldsymbol{\gamma_s}$ to be common across individuals. The variable vectors $\boldsymbol{a_{it}}$, $\boldsymbol{b_{it}}$ and $\boldsymbol{c_{it}}$ on the other hand change for every time step t and are different for each individual i. An alternative is to use a multinomial logit model for the transition matrix (Abhishek et al. 2012). But it uses more parameters and therefore we choose the ordered logit model.

Since the threshold parameters $\tau_{iss'}$ vary across individuals, the specification of the transition matrix also accounts for consumer heterogeneity. This is important to distinguish time dynamic effects from consumer specific effects (Heckman 1981; Netzer et al. 2008). While letting $\tau_{iss'}$ vary across consumers, the baseline probability to migrate between the states is different for every consumer. In some cases, consumers might be convinced easily by the ads and the migration to the next state happens fast. More critical consumers are harder to convince and the threshold is more difficult to cross. Therefore, every consumer has a different threshold that needs to be crossed in order to migrate to the next state.

2.3.3 Probability to react

At each time *t* the individual consumer *i* can choose whether to click on an ad or not or to visit the website of the firm or not, respectively. The probability of this choice follows a binary logit model:

$$m_{it|s} = \frac{\exp(\tilde{\beta}_{0s})}{1+\exp(\tilde{\beta}_{0s})};\ s=1,\dots NS, \tag{6}$$

with $\tilde{\beta}_{0s}$ as the intrinsic probability to convert. For identification of the different states the intrinsic probabilities have to be monotonically increasing. Technically speaking that is $\tilde{\beta}_{01} \le \tilde{\beta}_{02} \le \cdots \tilde{\beta}_{0NS}$. To ensure the restriction it is convenient to set $\tilde{\beta}_{01} = \beta_{01};\ \tilde{\beta}_{02} = \tilde{\beta}_{01} + \exp(\beta_{02});\ \tilde{\beta}_{0NS} = \tilde{\beta}_{0NS-1} + \exp(\beta_{0NS})$. Notice that there are no additional variables influencing the conditional choice.[1]

[1] One can let additional variables influence the conditional choice (Netzer et al. 2008). We tested several variations of the model and found that when all variables are included in the transition matrix, the model fits statistics are best. This approach is also recommended by Netzer et al. (2008).

2.3.4 Likelihood function

The HMM generates a sequence of observations $y_{i1}, y_{i2}, \ldots, y_{iT}$ for each individual i. The joint probability of such a sequence, according to MacDonald and Zucchini (1997, p. 77) is

$$L_{iT} = P_i(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \ldots, Y_{iT} = y_{it}). \tag{7}$$

The likelihood of observing this sequence is determined by all the possible routes an individual can take (Netzer et al. 2008). Therefore we can specify equation (7) to

$$\begin{aligned} L_{iT} &= P_i(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \ldots, Y_{iT} = y_{it}) \\ &= \sum_{s_1=1}^{NS} \sum_{s_2=1}^{NS} \cdots \sum_{s_t=1}^{NS} \left[P(S_{i1} = s_1) \prod_{t=2}^{T} P\left(S_{it} = s_t \mid S_{it-1} = s_{it-1}\right) \prod_{t=1}^{T} P(Y_{it} = y_{it} \mid S_{it} = s_t) \right]. \end{aligned} \tag{8}$$

The likelihood to maximize is the theoretical likelihood but is hard to compute because it has NS^T elements (MacDonald and Zucchini 1997, p. 78). The computational traceability is only given for very small values of T. A slight rearrangement of equation (8) simplifies the computation and leads to

$$L_{iT} = P\left(Y_{i1} = y_{i1}, \ldots, Y_{iT} = y_{iT}\right) = \boldsymbol{\pi_i} \boldsymbol{\tilde{m}_{i1}} \boldsymbol{Q_{i,1\to 2}} \boldsymbol{\tilde{m}_{i2}}\, \boldsymbol{L} \boldsymbol{Q_{i,T\text{-}1\to T}} \boldsymbol{\tilde{m}_{iT}} \boldsymbol{1'}, \tag{9}$$

where $\boldsymbol{\pi_i}$ is the solution to $\boldsymbol{\pi_i} = \boldsymbol{\pi_i} \boldsymbol{\bar{Q}_i}$, s.t. $\sum_{s=1}^{NS} \pi_{is} = 1$, $\boldsymbol{\tilde{m}_{it}} = \begin{bmatrix} \tilde{m}_{it|s=1} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \tilde{m}_{it|s=NS} \end{bmatrix}$

with $\tilde{m}_{it|s} = \tilde{m}_{it|s=1}^{y_{it}} \left(1 - \tilde{m}_{it|s}\right)^{(1-y_{it})}$, y_{it} is the outcome of the dependent variable and $\boldsymbol{1'}$ is a $NSx1$ vector of 1s. The expression $\tilde{m}_{it|s}$ is the probability function of the Bernoulli distribution. This distribution is appropriate in our setting because the dependent variable is conversion or not which represent a binary variable.

Another problem that occurs in evaluating the likelihood is that the computation might suffer from numerical underflow. That means some elements of the function might be too small to be distinguishable from zero throughout the estimation process. Because of the multiplicative structure of the likelihood function this is a severe problem and has to be handled. A possible solution is to compute the scaled likelihood (MacDonald and Zucchini 1997, p. 79). The approach is to divide the joint state likelihood after each time period by a scale factor, accumulate the logarithms of these factors and add them to the logarithm of the likelihood. This is a convenient and stable way to compute the log-likelihood function from equation (9). Possible scale factors are to divide by the average vector element or the sum of the vector elements. In this study, we decide for the latter option because then the sum of the logarithmized scale factors is zero and we just have to compute the log-likelihood.

To obtain the log-likelihood function across individuals, one has to simply calculate the log-likelihood for each individual i and accumulate these log-likelihoods.

2.3.5 Endogeneity

Online advertising studies suffer from endogeneity because clickstream data are not random. Especially in display advertising, the targeting option is mostly a concern for endogeneity because an ad is served based on past browsing behavior. To mimic the targeting option, the researcher needs to know the exact targeting algorithms. The targeting algorithm is often not available because ad agencies do the targeting for advertisers. These agencies often do not reveal their exact approach on how they serve the ads.

To mitigate the problem, the researcher can use an instrumental variable approach to control for endogeneity (Heckman 1997). But the identification of valid instruments is challenging and almost impossible for online advertising (Abhishek et al. 2012; Rutz et al. 2012). Furthermore, invalid instruments make the endogeneity problem even worse (Rossi 2014).

We have several reasons that endogeneity is not a major concern in this study. First, a general concern is that the serving process of display ads is correlated with the interest of the consumer (Abhishek et al. 2012). Since we explicitly want to model the consumer's interest this is not a problem here. Second, we had a deeper look in our data and we did not find differences in targeting strategies across consumers. Therefore, if previous information is known about consumers, this information will be treated the same.

But there still might be a concern that previous ad encounters or previous website visits by the consumer influence the decision of the advertiser. One can argue that advertisers choose some ad messages or targeting options more often than others if the consumer has an observable history with the advertised brand. This history can consist of previous ad encounters or previous visits on the website. Therefore, these situations need to be controlled for. Therefore, we include two additional covariates in the state-specific equation: number of advertising exposures before actual exposure and number of website visits before exposure.

Usually, returns to advertising follow a diminishing pattern (Vakratsas and Ambler 1999). This means that the first exposure has the biggest impact and each subsequent exposure has less impact. To control for such diminishing patterns, we take the logarithm of both variables.[2]

Therefore, equations (3) – (5) change to

$$q_{its1} = \frac{\exp(\tau_{is1} - \boldsymbol{a}_{it}'\boldsymbol{\rho}_s - \boldsymbol{b}_{it}'\boldsymbol{v}_s - \boldsymbol{c}_{it}'\boldsymbol{\gamma}_s - \boldsymbol{z}_{it}'\boldsymbol{\xi}_s)}{1+\exp(\tau_{is1} - \boldsymbol{a}_{it}'\boldsymbol{\rho}_s - \boldsymbol{b}_{it}'\boldsymbol{v}_s - \boldsymbol{c}_{it}'\boldsymbol{\gamma}_s - \boldsymbol{z}_{it}'\boldsymbol{\xi}_s)} \tag{10}$$

$$\begin{aligned} q_{its2} = & \frac{\exp(\tau_{is2} - \boldsymbol{a}_{it}'\boldsymbol{\rho}_s - \boldsymbol{b}_{it}'\boldsymbol{v}_s - \boldsymbol{c}_{it}'\boldsymbol{\gamma}_s - \boldsymbol{z}_{it}'\boldsymbol{\xi}_s)}{1+\exp(\tau_{is2} - \boldsymbol{a}_{it}'\boldsymbol{\rho}_s - \boldsymbol{b}_{it}'\boldsymbol{v}_s - \boldsymbol{c}_{it}'\boldsymbol{\gamma}_s - \boldsymbol{z}_{it}'\boldsymbol{\xi}_s)} \\ & - \frac{\exp(\tau_{is1} - \boldsymbol{a}_{it}'\boldsymbol{\rho}_s - \boldsymbol{b}_{it}'\boldsymbol{v}_s - \boldsymbol{c}_{it}'\boldsymbol{\gamma}_s - \boldsymbol{z}_{it}'\boldsymbol{\xi}_s)}{1+\exp(\tau_{is1} - \boldsymbol{a}_{it}'\boldsymbol{\rho}_s - \boldsymbol{b}_{it}'\boldsymbol{v}_s - \boldsymbol{c}_{it}'\boldsymbol{\gamma}_s - \boldsymbol{z}_{it}'\boldsymbol{\xi}_s)}, \end{aligned} \tag{11}$$

$$\vdots$$

$$q_{itsNS} = 1 - \frac{\exp(\tau_{isNS-1} - \boldsymbol{a}_{it}'\boldsymbol{\rho}_s - \boldsymbol{b}_{it}'\boldsymbol{v}_s - \boldsymbol{c}_{it}'\boldsymbol{\gamma}_s - \boldsymbol{z}_{it}'\boldsymbol{\xi}_s)}{1+\exp(\tau_{isNS-1} - \boldsymbol{a}_{it}'\boldsymbol{\rho}_s - \boldsymbol{b}_{it}'\boldsymbol{v}_s - \boldsymbol{c}_{it}'\boldsymbol{\gamma}_s - \boldsymbol{z}_{it}'\boldsymbol{\xi}_s)}, \tag{12}$$

where $\boldsymbol{z}_{it}$ represent the additional variables and $\boldsymbol{\xi}_s$ captures the effect of these variable for each state.

2 One can also argue for an inverted U-shape modeling of these variables. This would imply that after a certain amount of exposures, each additional exposure has a negative influence. We did not choose this form of modeling since the effect whether a consumer is negatively touched by the ad is captured by the advertising variables message, format and targeting. Moreover we have multiple states that represent the consumer's interest. Therefore, each additional exposure can have different effects (positive or negative) depending on the actual state the consumer is in.

2.3.6 Recovering state membership

For advertisers to allocate the combination of ad characteristics and targeting options, they must know the state membership, where the consumer is in. Generally, there are two approaches: filtering and smoothing (Hamilton 1989). Whereas filtering uses the information up to the point of interest, smoothing uses all available data points, also the ones that lay ahead the point of interest. Since the advertisers can only use data up to the point they want to serve the ad, we use the filtering approach. It is based on the likelihood function and the probability to be in state s is:

$$P\left(S_{it} = s \mid Y_{i1} = y_{i1}, \ldots, Y_{iT} = y_{iT}\right) = \boldsymbol{\pi}_i \tilde{\boldsymbol{m}}_{i1} \boldsymbol{Q}_{i,1\to 2} \tilde{\boldsymbol{m}}_{i2}\, \boldsymbol{L} \boldsymbol{Q}_{i,t\text{-}1\to t,s} \tilde{\boldsymbol{m}}_{it|s} \,/\, L_{it}, \quad (13)$$

Where $\boldsymbol{Q}_{i,t\text{-}1\to t,s}$ is the sth column of the transition matrix and L_{it} the value of the likelihood function up to time t. The probabilities can be calculated for each state s and the consumer is allocated to the state with the highest probability.

2.3.7 Model estimation

The estimation is in a Bayesian framework where we model random effects in the transition matrix to consider consumer heterogeneity. As described above, this is important to distinguish individual heterogeneity from time dynamics. That is allowing the threshold parameters $\tau_{iss'}$ to vary across individuals. Therefore we have two sets of parameters: random effect parameters $(\boldsymbol{\theta}_i)$ and fixed effect parameters $(\boldsymbol{\psi})$. Since the threshold parameters vary, they are part of the random effect parameters and we can define $\boldsymbol{\theta}_i = \{\tau_{iss'}\}$, $s' \in \{1, \ldots NS - 1\}$. The fixed effect parameters are $\boldsymbol{\psi} = \{\boldsymbol{\rho}_s, \boldsymbol{\xi}_s, \beta_{0s}, \boldsymbol{\beta}_s\}$.

Priors

We use Gibbs Samplers and Metropolis Hastings algorithms to draw the model parameters. For both sets of parameters we need a starting value, the priors. We assume the random effect parameter $\boldsymbol{\theta}_i$ to be normally distributed with mean $\boldsymbol{\mu}_\theta$ and variance $\boldsymbol{\Sigma}_\theta$. Therefore the probability of $\boldsymbol{\theta}_i$ is

$$\boldsymbol{\theta}_i \sim N\left(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta\right) \Rightarrow P\left(\boldsymbol{\theta}_i\right) \propto \exp\left(0.5\left(\boldsymbol{\theta}_i - \boldsymbol{\mu}_\theta\right)' \boldsymbol{\Sigma}_\theta^{-1} \left(\boldsymbol{\theta}_i - \boldsymbol{\mu}_\theta\right)\right). \quad (14)$$

We also have to set starting values for the mean and precision parameters which determine $\boldsymbol{\theta}_i$. We also assume $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ to be normally distributed with mean $\boldsymbol{\mu}_{\boldsymbol{0}}$ and variance $\boldsymbol{V}_{\boldsymbol{0}}$. Therefore, the probability of $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ is

$$\boldsymbol{\mu}_{\boldsymbol{\theta}} \sim N\left(\boldsymbol{\mu}_0, \boldsymbol{V}_0\right) \Rightarrow P\left(\boldsymbol{\mu}_{\boldsymbol{\theta}}\right) \propto \exp\left(0.5\left(\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\mu}_0\right)' \boldsymbol{V}_{\boldsymbol{0}}^{-1}\left(\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\mu}_0\right)\right). \qquad (15)$$

We further assume the inverse of the covariance matrix, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}$, to be Wishart distributed, with df_0 degrees of freedom and scale matrix $\boldsymbol{S}_{\boldsymbol{0}}$. This leads to

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \sim W(df_0, \boldsymbol{S}_{\boldsymbol{0}}). \qquad (16)$$

We set the prior hyper-parameters $\{\boldsymbol{\mu}_{\boldsymbol{0}}, \boldsymbol{V}_{\boldsymbol{0}}, df_0, \boldsymbol{S}_{\boldsymbol{0}}\}$ in a way that they do not give any previous information. That means that the transition matrix in the beginning only consists of entries which resemble $1/NS$.

The priors for the fixed effects parameters are defined as follows: We also assume $\boldsymbol{\psi}$ to be normally distributed with mean $\boldsymbol{\mu}_{\psi}$ and variance $\boldsymbol{\Sigma}_{\psi}$. In a same way as for the random effect parameters the probability of $\boldsymbol{\psi}$ is

$$\boldsymbol{\psi} \sim N\left(\boldsymbol{\mu}_{\psi}, \boldsymbol{\Sigma}_{\psi}\right) \Rightarrow P\left(\boldsymbol{\psi}\right) \propto \exp\left(0.5\left(\boldsymbol{\psi} - \boldsymbol{\mu}_{\psi}\right)' \boldsymbol{\Sigma}_{\psi}^{-1}\left(\boldsymbol{\psi} - \boldsymbol{\mu}_{\psi}\right)\right). \qquad (17)$$

The prior hyper-parameters are again set in a way that they do not give any information. For the fixed effect parameters it means for the choice probabilities, that they always start at 0.5 (50% chance for converting and nor converting). The remaining parameters which estimate the effect of the ad characteristics and targeting options in the decision matrix, resp. the parameters that determine the short term influence on the choice probability are set to zero.

Posteriors

The full conditional distribution to sample from the random effects is

$$P\left(\boldsymbol{\theta}_i \mid \boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}, \boldsymbol{\psi}, data_i\right) \propto \exp\left(0.5\left(\boldsymbol{\theta}_i - \boldsymbol{\mu}_{\boldsymbol{\theta}}\right)' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\left(\boldsymbol{\theta}_i - \boldsymbol{\mu}_{\boldsymbol{\theta}}\right)\right) L\left(\boldsymbol{Y}_i \mid \boldsymbol{\theta}_i, \boldsymbol{\psi}\right), (18)$$

where $L\left(\boldsymbol{Y}_i \mid \boldsymbol{\theta}_i, \boldsymbol{\psi}\right)$ is the likelihood function from equation (9). The population mean is drawn from

$$\boldsymbol{\mu_\theta} \sim N\left(\boldsymbol{\mu_n}, \boldsymbol{V_n}\right), \tag{19}$$

Where $\boldsymbol{V_n^{-1}} = \left(\boldsymbol{V_0^{-1}} + N\boldsymbol{\Sigma_\theta^{-1}}\right)$ and $\boldsymbol{\mu_n} = \boldsymbol{V_n}\left(\boldsymbol{\mu_0 V_0^{-1}} + N\boldsymbol{\bar{\theta}\Sigma_\theta^{-1}}\right)$. N indicates the number of individuals.

The precision matrix is drawn from

$$\boldsymbol{\Sigma_\theta^{-1}} \sim W\left(df_1, \boldsymbol{S_1}\right) \tag{20}$$

where $df_1 = df_0 + N$ and $\boldsymbol{S_1^{-1}} = \sum_{i=1}^{N}\left(\boldsymbol{\theta_i} - \boldsymbol{\mu_\theta}\right)\left(\boldsymbol{\theta_i} - \boldsymbol{\mu_\theta}\right)' + \boldsymbol{S_0}^{-1}$.

The full conditional distribution to sample the fixed effects from is

$$\begin{aligned} &P\left(\boldsymbol{\psi} \mid \boldsymbol{\mu_\psi}, \boldsymbol{\Sigma_\psi}, \{\boldsymbol{\theta_i}\}, data\right) \propto \\ &\exp\left(0.5\left(\boldsymbol{\psi} - \boldsymbol{\mu_\psi}\right)' \boldsymbol{\Sigma_\psi^{-1}}\left(\boldsymbol{\psi} - \boldsymbol{\mu_\psi}\right)\right) L\left(\boldsymbol{Y} \mid \{\boldsymbol{\theta_i}\}, \boldsymbol{\psi}\right), \end{aligned} \tag{21}$$

where $L\left(\boldsymbol{Y} \mid \{\boldsymbol{\theta_i}\}, \boldsymbol{\psi}\right)$ is the sum over the likelihood function from equation (9) over all individuals i.

With the priors and posteriors, we can estimate the model using the Gibbs sampler and Metropolis Hastings algorithm. We use the Gibbs sampler for posteriors with a closed from (i.e., a known standard distribution) and the Metropolis Hastings algorithm for posteriors with no closed form. Both algorithms stem from the Bayesian statistics literature and are algorithms of the so called Markov Chain Monte Carlo (MCMC) methods (Rossi and Allenby 2003). In the following section, both, the Gibbs sampler and the Metropolis Hastings algorithm will be explained shortly. For a detailed description of the Gibbs sampler and the Metropolis-Hastings algorithms see for example Rossi et al. (2005, pp. 86-94).

First, we explain the general idea of the Gibbs sampler: Suppose we want to sample from a joint distribution $f\left(x_1, x_2\right)$. The closed form of that distribution is unknown, but we know the conditional distributions $f\left(x_1 \mid x_2\right)$ and $f\left(x_2 \mid x_1\right)$. Since only a part of the distribution is known, it is not possible to sample directly from the joint distribution. The Gibbs sampler proceeds as follows: Choose a starting value for x_2 and sample from the conditional distribution of x_1. The sampled parameter x_1 then is used to sample from the conditional distribution of x_2. Hence, we sample one parameter at a time given the other parameter. Since the starting value is chosen by the researcher and therefore not a part of the target distribution we want to sample from, the first draws of x_1 and x_2 have to be discarded. The discarded sample is called the burn-in period and is used to let the algorithm to get near the target distribution. The burn-in period can be the first thousands or ten thousands draws. It depends on the complexity of the target distribution. After discarding the burn-in sample, the algorithm approximately draws from the target distribution $f(x_1,x_2)$ (Greenberg 2008, pp. 91 - 92; Rossi et al. 2005, pp. 63 - 64).

When the conditional distribution is not a standard distribution that is easy to sample from, the Gibbs sampler cannot be applied. Then, a more general algorithm must be used. Thus, we explain the idea of the Metropolis-Hastings algorithm in the following section: Suppose we want to sample from the target distribution $f(x)$. The first step is to find a transition kernel $q(X,Y)$, which has $f(\cdot)$ as stationary distribution. The idea of the transition kernel is to draw a new proposal value and evaluate whether or not the new value belongs to the stationary distribution or not. Thus, the algorithm proceeds as follows: Choose a starting value X and draw a new value Y from the transition kernel. The probability of acceptance is

$$\alpha(X,Y) = \min\left\{\frac{f(Y)q(Y,X)}{f(X)q(X,Y)},1\right\}. \tag{22}$$

With probability α the proposed value Y will be accepted. Otherwise the algorithm continues with the previous value X. These steps repeat until the chain draws from the stationary distribution. Again, a burn-in period is used to let the chain learn from the data to reach high probability regions.

A prominent and often used kernel is the random-walk kernel. The idea is that the proposed value depends on the previous value plus an error term. The error term is often normally distributed with mean zero and covariance matrix $\sigma^2\Sigma$, where σ^2 is a scaling parameter of the covariance matrix that influences the step size. A prominent feature of the random walk transition kernel is the symmetry $q(X,Y) = q(Y,X)$, which simplifies the acceptance probability to

$$\alpha(X,Y) = \min\left\{\frac{f(Y)}{f(X)},1\right\}. \tag{23}$$

There are several ways to choose σ^2 and Σ. We will present an algorithm that automatically chooses the two parameters to reach the best performance of the Metropolis Hastings algorithm. Before we go into detail, we will present the whole model estimation procedure in the next section.

The estimation algorithm draws the parameters from the conditional distributions. The full model is:

$$\boldsymbol{\theta}_i \mid \boldsymbol{Y}_i, \boldsymbol{x}_i, \boldsymbol{c}_i, \boldsymbol{\psi}, \boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta \tag{24}$$

$$\boldsymbol{\mu}_\theta \mid \{\boldsymbol{\theta}_i\}, \boldsymbol{\Sigma}_\theta \tag{25}$$

$$\boldsymbol{\Sigma}_\theta \mid \{\boldsymbol{\theta}_i\}, \boldsymbol{\mu}_\theta \tag{26}$$

$$\boldsymbol{\psi} \mid \boldsymbol{Y}, \boldsymbol{x}, \boldsymbol{c}, \{\boldsymbol{\theta}_i\} \tag{27}$$

To draw $\boldsymbol{\theta}_i$ from equation (24), we use the Metropolis-Hastings algorithm, since the posterior has no closed form. We use a random walk Metropolis (RWM) algorithm with a Gaussian proposal distribution. The mean of the proposal distribution is the previous accepted value of $\boldsymbol{\theta}_i$, and the variance is $\sigma^2 \boldsymbol{\Lambda}$. Therefore the j-th draw of $\boldsymbol{\theta}_i$ is $\boldsymbol{\theta}_i^j = \boldsymbol{\theta}_i^{j-1} + N\left(0, \sigma_i^2 \boldsymbol{\Lambda}_i\right)$, where σ_i^2 is the scaling parameter of the RWM and $\boldsymbol{\Lambda}_i$ the covariance matrix of the proposal kernel. Ideally the choice of the scaling parameter and the covariance matrix for a RWM should yield an acceptance rate of 0.234 (Roberts et al. 1997). To omit a trial and error approach to reach the desired acceptance rate, there exist adaptive versions of Metropolis Hastings algorithms, that automatically tune the parameters to reach any desired acceptance rate and reduce autocorrelation of the draws (e.g., Atchade and Rosenthal 2005; Atchadé 2006; Haario et al. 2001). We include the adaptive algorithm of Atchadé (2006) and the algorithm is as follows:

1. Start the algorithm at some point $\boldsymbol{\theta}_{i_o}$ with adaptive tuning parameters $\boldsymbol{\mu}_0, \boldsymbol{\Gamma}_0, \sigma_0$.
2. At time j set $\boldsymbol{\Lambda}_j = \boldsymbol{\Gamma}_j + \varepsilon_1 \boldsymbol{I}_{dim\theta_i}$, where $\varepsilon_1 = 10^{-6}$ and $\boldsymbol{I}_{dim\theta_i}$ the identity matrix of rank $\dim(\boldsymbol{\theta}_i)$. The value of ε_1 is arbitrary, it just has to be a very small number. We stick to the proposal of the author.

3. Generate a new proposal $\boldsymbol{\theta}_{i_j} = \boldsymbol{\theta}_{i_{j-1}} + N\left(0, \sigma_j^2 \boldsymbol{\Lambda}_j\right)$ and generate $u \sim U\left(0,1\right)$, where $U\left(0,1\right)$ is the uniform distribution with minimum 0 and maximum 1.
4. Calculate the acceptance ratio:

$$\alpha_{\theta_i} = \min\left\{ \frac{\exp\left(0.5\left(\boldsymbol{\theta}_{i_j} - \boldsymbol{\mu}_\theta\right)' \boldsymbol{\Sigma}_\theta^{-1} \left(\boldsymbol{\theta}_{i_j} - \boldsymbol{\mu}_\theta\right)\right) L\left(\boldsymbol{Y}_i \mid \boldsymbol{\theta}_{i_j}, \boldsymbol{\psi}\right)}{\exp\left(0.5\left(\boldsymbol{\theta}_{i_{j-1}} - \boldsymbol{\mu}_\theta\right)' \boldsymbol{\Sigma}_\theta^{-1} \left(\boldsymbol{\theta}_{i_{j-1}} - \boldsymbol{\mu}_\theta\right)\right) L\left(\boldsymbol{Y}_i \mid \boldsymbol{\theta}_{i_{j-1}}, \boldsymbol{\psi}\right)}, 1 \right\}.$$

5. If $u < \alpha_{\theta_i}$ accept the proposed value, otherwise stick to the old value.
6. Update the adaptive tuning parameters:
   - 6.1. Set $\gamma_{j-1} = \dfrac{10}{j-1}$,
   - 6.2. $\boldsymbol{\Gamma}_j = \boldsymbol{\Gamma}_{j-1} + \gamma_{j-1}\left(\left(\boldsymbol{\theta}_{i_j} - \boldsymbol{\mu}_{j-1}\right)\left(\boldsymbol{\theta}_{i_j} - \boldsymbol{\mu}_{j-1}\right)' - \boldsymbol{\Gamma}_{j-1}\right)$,
   - 6.3. $\boldsymbol{\mu}_j = \boldsymbol{\mu}_{j-1} + \gamma_{j-1}\left(\boldsymbol{\theta}_{i_j} - \boldsymbol{\mu}_{j-1}\right)$,
   - 6.4. $\sigma_j = \sigma_{j-1} + \gamma_{j-1}\left(\alpha - \tau_{opt}\right)$, where $\tau_{opt} = 0.234$.
7. Go back to 2. and repeat.

The adaptive tuning parameters converge against some estimation parameters. As $j \to \infty$, $\boldsymbol{\mu}_j$ is the empirical mean of the sample and $\boldsymbol{\Gamma}_j$ will converge to the covariance matrix $\boldsymbol{\Sigma}_\theta$.

After generating $\{\boldsymbol{\theta}_i\}$, we can draw from the posterior distribution of $\boldsymbol{\mu}_\theta$. Since $\boldsymbol{\mu}_\theta$ follows a standard normal distribution, we just have to calculate $\boldsymbol{\mu}_n$ and $\boldsymbol{V}_n$ and perform a Gibbs update from $N\left(\boldsymbol{\mu}_n, \boldsymbol{V}_n\right)$.

In the next step we draw form the posterior distribution of $\boldsymbol{\Sigma}_\theta$. Its inverse, $\boldsymbol{\Sigma}_\theta^{-1}$, follows a standard Wishart distribution and we just have to compute df_1 and $\boldsymbol{S}_1$ and perform a Gibbs update from $\boldsymbol{\Sigma}_\theta^{-1} = W\left(df_1, \boldsymbol{S}_1\right)$.

In the last step we draw a sample from the posterior distribution of $\boldsymbol{\psi}$ which is given in equation (17). Similar to the draw of $\boldsymbol{\theta}_i$ the posterior is not available in closed form and we have to perform the same Metropolis Hastings algorithm. Again we use the adaptive version of Atchadé (2006) to automatically tune the RWM and reduce autocorrelation. The acceptance ratio for $\boldsymbol{\psi}$ is given by

$$\alpha_\psi = \min\left\{ \frac{\exp\left(0.5\left(\boldsymbol{\psi}_j - \boldsymbol{\mu}_\psi\right)' \boldsymbol{\Sigma}_\psi^{-1} \left(\boldsymbol{\psi}_j - \boldsymbol{\mu}_\psi\right)\right) L\left(\boldsymbol{Y}_i \mid \{\boldsymbol{\theta}_i\}, \boldsymbol{\psi}_j\right)}{\exp\left(0.5\left(\boldsymbol{\psi}_{j-1} - \boldsymbol{\mu}_\psi\right)' \boldsymbol{\Sigma}_\psi^{-1} \left(\boldsymbol{\psi}_{j-1} - \boldsymbol{\mu}_\psi\right)\right) L\left(\boldsymbol{Y}_i \mid \{\boldsymbol{\theta}_i\}, \boldsymbol{\psi}_{j-1}\right)}, 1 \right\}.$$

The Hidden Markov Model can be sensitive to the starting values of the estimation algorithm. To gain a better fit of the model we apply maximum likelihood estimation before estimating the hierarchical Bayes model. Since the likelihood function is complex and deriving the gradient and the Hessian is not obvious, we choose an estimation algorithm that is not based on gradients, but on function values. There are two possible algorithms that fit the need: the Nelder-Mead (NM) algorithm and simulated annealing (Khachaturyan et al. 1981; Nelder and Mead 1965). The NM algorithm suffers from the problem to be stuck in a local optimum for complex and high-dimensional likelihood functions. Therefore simulated annealing is the better option, since it searches for a global optimum. To estimate the starting values, we use the maxLik package in R, which allows using the simulated annealing algorithm. We also conduct a simulation study to test the model and the algorithm. The detailed procedure of the simulation can be found in the appendix.

2.4 Data

The data was provided by the Wharton Customer Analytics Initiative. The dataset is sponsored by Annalect, the data and analytics division of the Omnicom Group. Annalect is analyzing advertising campaigns for several companies. In our case the data are transaction data from an international travel and tourism company and span from December 2014 to January 2015. They use cookies to track consumers across different websites. Cookies are the industry standard and are often used in empirical advertising studies (e.g., Abhishek et al. 2012; Braun and Moe 2013).

The data consist of display impressions, display clicks, search clicks, email clicks, other clicks (where the data provider also does not known what "other" specially means) and website visits. A website visit can either result from a click on the different ad types or when the user entered the URL manually. Notice that a click on a certain ad type always results in a website visit. Since other ad types' impressions are usually not reported (as this is the case here), we create the covariates that control for other ad types as follows: We create covariates for search, email and other ad types that are set to zero. The covariate switches to one, as soon as a consumer encountered another ad type.

Furthermore, we have information about the targeting options. We have information about content amplification, content integration, behavioral targeting and retargeting, as well as other targeting options.[3] The advertiser treats these targeting options as mutually exclusive. We let other targeting options serve as the baseline category, where these four targeting options are evaluated against. We only know what kind of targeting was used, but not how the consumer was selected. Therefore, we assume that every consumer is targeted with the same (unknown) variables.

Moreover, every creative has a short description by Annalect. We code these descriptions to classify the message of the ad in brand-related, product-related and sales-related ads.[4] Brand-related ads are the ones that inform the consumer that travelling in general is something the consumer should consider. Product-related ads inform about a specific offer the company has and sales-related ads inform about the hard facts and include price information, discounts or a clear call to action. Some descriptives are ambiguous and are therefore coded as "Other" and serve as a baseline category. Last but not least, we also have information about the format. We differentiate the ads between non-obtrusive (standard) and obtrusive formats. Obtrusive formats include for example interstitials and roadblock units. Basically, every ad unit that is not standard is classified as obtrusive.

Our sample consists of 4,534 cookies with 187,499 observations in total. They were chosen along the following criteria. To increase the probability of choosing consumers who do not delete cookies, we exclude every cookie with less than 40 encounters with the ad.[5] These encounters contain all ad formats, including display, search and email. For these cookies, we extracted the display impressions and coded other ad encounters. Based on the description of the underlying dataset, an observation in our model comprises a display advertising impression, a display advertising click that results into a website visit or a

[3] Other targeting options include for example geo targeting and lookalike targeting. However, the sample sizes of these options are too small to analyze. We decide to include these options as "other" targeting strategies to not drop information about the ad exposures to consumers.

[4] Since we have a non-disclosure agreement with Annalect, we were not able to code these variables by multiple coders. It was not possible to change the description in a way that the original advertised brand was not recognizable.

[5] Consumers who delete cookies are a problem for advertisers. Since cookies contain information about the past browsing behavior, this information is gone after deleting the cookie. So far, there are no algorithms that are able to recover information from deleted cookies. Therefore, we choose are rather high amount of ad encounters to ensure that these are consumers did not delete their cookies along the observation period.

manually entered website visit. We combine clicks and visits since they result in the same scenario.

Table 2-2 gives an overview about the descriptive statistics of the sub-sample and the relative frequency of clicks and visits. The mean percentage of website visits resulting of clicks and manual website visits is 0.4%. This also represents the industry average. Model-free evidence already shows that targeting options (e.g., content amplification with 0.8%) and ad formats (obtrusive ads have a ratio of 4.6%) result in higher website visits. For ad messages, we only see so far that product related ads have a slightly lower ration (0.3%).

Table 2-2: Descriptive statistics of the sub-sample

Variables	Relative frequency for clicks and visits
Targeting approach (Baseline category: Other targeting options)	
Content integration	0.002
Content amplification	0.008
Behavioral Targeting	0.005
Retargeting	0.003
Ad format (Baseline category: non-obtrusive ad formats)	
Obtrusive ad	0.046
Ad message (Baseline category: Other ad messages)	
Brand-related	0.004
Product-related	0.003
Sales-related	0.004
Total	0.004

2.5 Results

2.5.1 Number of states

We estimate the model in a hierarchical Bayes framework. To let the estimation algorithm converge we use a burn-in period of 90,000 draws. Further, to avoid autocorrelation, we save every 10th iteration and use 1,000 draws to determine the posterior distribution. Thus, in total we let the algorithm run for 100,000 draws. To determine the number of states, we estimate several models with a different number of states and choose the model that has the best model fit based on the likelihood, log marginal density, log Bayes factor and the Deviance Information criterion (DIC) (Montoya et al. 2010; Netzer et al. 2008).

Table 2-3 shows the model statistics for different numbers of hidden states. All statistics indicate that a model with two hidden states fit the data the best. According to MacDonald and Zucchini (1997. p. 67) a two state model is often sufficient in practice. First, the amount of parameter estimates is not too high. With an increasing number of states, the amount of parameters to estimate increases tremendously. Second, this also has advantages in handling the model. Managers want rules that are easy to apply. With only two states, we can ensure that. We will get a parameter estimate for every ad characteristic and every targeting option for each state.

Table 2-3: Determining the number of states

Number of states	-2LL	-2Log marginal density	Log Bayes factor	DIC
1	8195.73	8224.54	-	8209.08
2	**6536.15**	**7266.64**	**478.95**	**6775.53**
3	6879.37	7272.39	-2.88	6988.59

2.5.2 Parameter estimates

Table 2-4 shows the posterior means and standard deviations of the estimated parameters. Since the interpretation of the parameter estimates is not straightforward, we calculate the transition matrices using equations (10) – (12) and the state dependent choice using equation (6). The resulting transition matrices are shown in Table 2-5. Figure 2-2 shows the histograms of the estimated parameters of the ad characteristics, targeting options and the endogeneity corrections. The covariates are not reported because they are modeled as binary variables and indicate whether a consumer had contact to other ad types or not. It is not differentiated how often they had contact with the ads. This information is available in the endogeneity correction variable that counts the consumer's contact with the advertised brand.

Table 2-4: Estimated parameters of the proposed model

	Parameter	State 1 ('cold' state)	State 2 ('hot' state)
	State intercept	**-7.04 (0.10)**	**1.70 (0.02)**
Targeting approach	Content integration	**-1.44 (0.43)**	-0.55 (0.67)
	Content amplification	0.75 (0.41)	0.29 (0.52)
	Behavioral targeting	**-1.78 (0.39)**	**1.59 (0.71)**
	Retargeting	**-2.83 (0.45)**	-1.01 (0.52)
Ad format	Obtrusive ad	**1.53 (0.49)**	**2.04 (0.65)**
Ad message	Brand-related ad	0.06 (0.62)	-0.32 (0.85)
	Product-related ad	**-1.61 (0.61)**	-0.46 (0.71)
	Sales-related ad	**0.56 (0.19)**	**-0.83 (0.38)**
Covariates	Email click	**-1.14 (0.55)**	0.42 (0.65)
	Search click	-0.39 (0.34)	-0.82 (0.75)
	Other click	0.18 (0.39)	0.81 (0.74)
Endogeneity correction	Previous ad encounters	**-1.39 (0.12)**	**1.14 (0.21)**
	Visits	**0.55 (0.21)**	**0.76 (0.38)**
	Threshold to cross states	3.87 (3.58)	1.75 (2.03)

(standard error in parenthesis, significant parameters in bold[6])

6 We consider a parameter as significant when the posterior mean is twice as great as the posterior standard devaince (Pieters and Wedel 2004).

Figure 2-2: Histograms of the transition variables

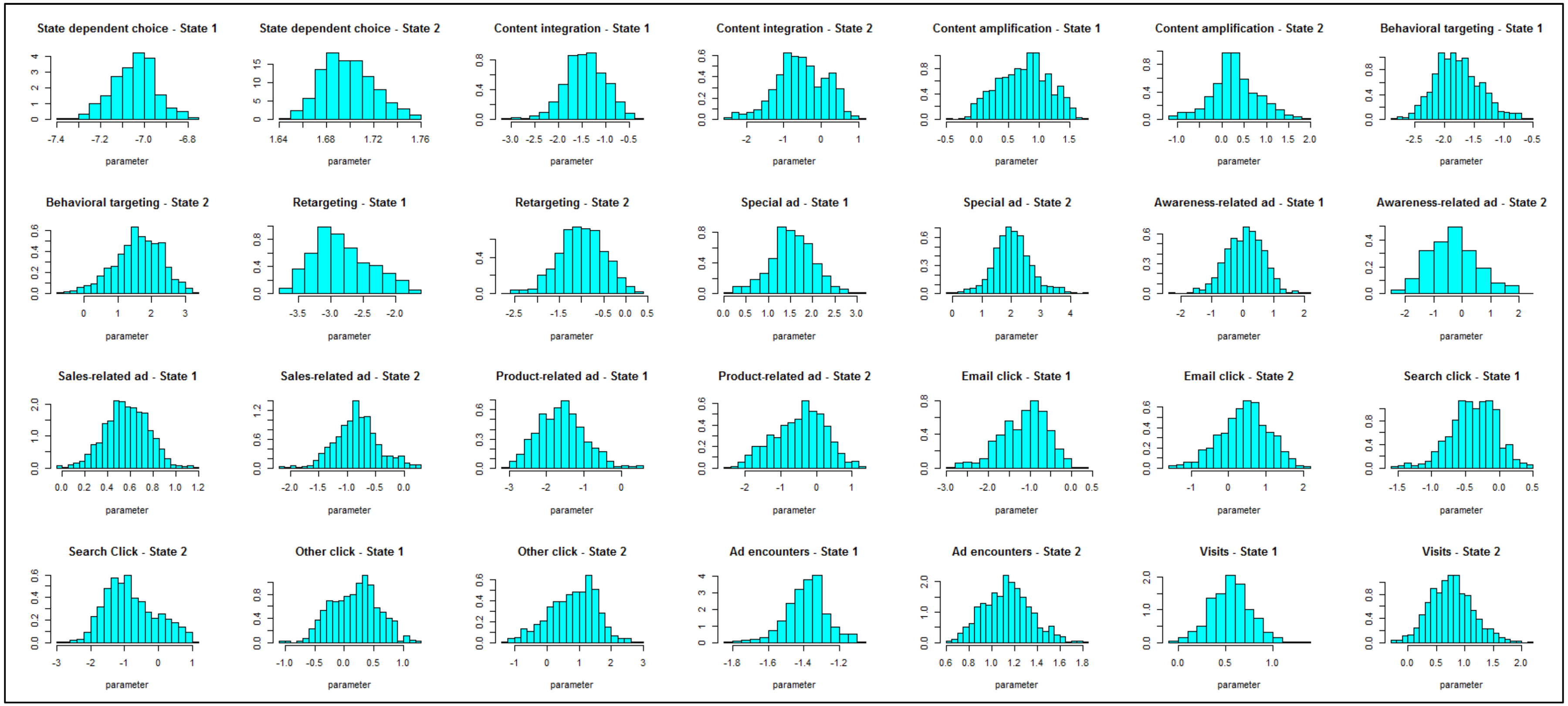


Table 2-5: Mean transition matrices for first impression

	Baseline		Content integration		Content amplification	
	t		t		t	
t-1	Cold	Hot	Cold	Hot	Cold	Hot
Cold	98.0%	2.0%	99.5%	0.5%	95.7%	4.3%
Hot	85.2%	14.8%	90.9%	9.1%	81.2%	18.8%

	Behavioral targeting		Retargeting		Obtrusive ad	
	t		t		t	
t-1	Cold	Hot	Cold	Hot	Cold	Hot
Cold	99.6%	0.4%	99.9%	0.1%	91.2%	8.8%
Hot	53.9%	46.1%	94.0%	6.0%	42.9%	57.1%

	Brand-related ad		Product-related ad		Sales-related ad	
	t		t		t	
t-1	Cold	Hot	Cold	Hot	Cold	Hot
Cold	97.8%	2.2%	99.6%	0.4%	96.5%	3.5%
Hot	88.8%	11.2%	90.1%	9.9%	92.9%	7.1%

	Previous ad encounters		Previous visit	
	t		t	
t-1	Cold	Hot	Cold	Hot
Cold	99.5%	0.5%	96.5%	3.5%
Hot	64.9%	35.1%	72.9%	27.1%

The results have to be interpreted with respect to the certain state and the probability of a consumer to click on the ad or to visit the website. This probability is different between state 1 and state 2. Inserting the state dependent choice parameters into equation (6), the probability to convert for state 1 is 0.09% and 17.08% for state 2. The probability of state 1 is way below the average ratio of a positive behavioral outcome, which is 0.4% in our data. The probability of state 2 is way above average. Therefore, we label state 1 as the “cold” interest state and state 2 as the “hot” interest state.

As expected the cold state is fairly “sticky”. This means, that the probability of staying in the cold state when being in the cold state is relatively high. Therefore, it is difficult to switch the interest state using display advertising. The hot state is not sticky, but this also makes sense. When a consumer is highly interested, he or she will normally get more information right afterwards. Therefore, there is a chance that at the next ad impression the consumer is less interested and transitions back to the cold state. Either, the consumer has all the information that is desired and the consumer is satisfied and does not need more information, or the gathered information is unattractive and therefore the interest in the product diminishes.

The results of the ad characteristics (ad message and ad format), targeting options and endogeneity variables can be compared to the baseline transition matrix which is calculated using the threshold parameters only. All the other matrices are calculated for one variable, holding all the other variables constant. The results suggest that brand-related ads increase the probability of transitioning from the cold state to the hot state. When being in the hot state, the probability of staying there decreases. Product related ads decrease both, the probability of switching from the cold state to the hot state and staying in the hot state when being there. When the advertiser shows sales-related ads when the consumer is in the cold state the probability of transitioning to the hot state increases. Surprisingly, when showing sales-related ads in the hot state, the probability of staying in the hot state decreases.

For targeting options, we find that content integrated ads decrease the probability of switching to the hot state when being in the cold state and staying in the hot state when being in the hot state. But ads that are targeted using content amplification increase both of the probabilities. Behavioral targeting decreases the probability of transitioning to the hot state when being in the cold state, but increases the probability of staying in the hot state when being in the hot state. Retargeting, however, decreases both probabilities.

Finally, we also included the logarithm of previous ad encounters and the logarithm of previous visits in the model. As expected, when a consumer is in the cold state and the advertiser serves more ads to the consumer, he or she might show reactance and the probability of migrating to the hot state decreases. The effect flips, when the consumer is in the hot state. Then, more encounters actually increase the probability of staying in the hot state. Previous visits on the advertiser's webpage increase both, the probability of transitioning from the cold state to the hot state and the probability of staying in the hot state.

The results imply a trade-off for advertisers. The probability is highest that a consumer is not interested in the product. This is shown by the sticky cold state. Now, the more ads the advertiser serves to the consumer, the lesser the probability of transitioning to the hot state. But when looking of the probabilities of the ad characteristics, some increase the probability of transitioning to the hot state. Thus, advertisers must carefully decide which ads and how many ads they show to the consumers. Using the model, it is possible to calculate the probability of transitioning to the hot state. When in the hot state, it is important to choose the right targeting strategy to reach the consumer while he or she is still interested. This leads to an increase in conversion (i.e., clicks and visits) behavior.

2.5.3 Comparison to other models

It may also be concern, that the latent interest states are not needed to explain advertising effectiveness. To tackle this concern, we compare the model to other alternatives. The first alternative is a standard logit model, followed by a random effects model. To prove, that our model explains the data better, we again choose different model evaluation criteria. Since the logit models are not Bayesian, we need other model fit statistics. Classic frequentist statistics include the likelihood, the Akaike Information Criterion and the Bayes Information Criterion (BIC, also referred to as the Schwarz Information Criterion). Table 2-6 indicates, that our model explains the data better than other models. This especially holds true, when having a look at the AIC and BIC, which penalize every additional variable that enters the model.

Table 2-6: Proposed model versus benchmark models

Model	-2LL	AIC	BIC
Logit model	8180.70	8208.70	8350.68
Random intercept logit model	7054.27	7084.27	7236.39
Proposed model	**6536.15**	**6602.15**	**6936.82**

2.5.4 Scenario analyses

To demonstrate the advantage of our model, we conduct several scenario analyses. We first want to show that the lift increases when serving display ad characteristics and targeting options that match the individual consumer's interest. Second, we simulate the behavioral outcome of consumer under several restrictions. We limit the number of obtrusive ads that can be served. These ads are often the result of a negotiation with the advertiser and the publisher who runs the website. Therefore, it is often the case that the advertiser does not have an unlimited number of obtrusive ads to serve. Furthermore, we also want to demonstrate that the model can be used to limit the amount of ads served to uninterested consumers. When a consumer is not interested in the product after several attempts, it is a waste of money to expose him to more display ads.

To do so, we use the parameter estimates from Table 2-4. We assume that every consumer starts in the cold state. For every impression, we use several combinations of ad characteristics and targeting options and calculate the probabilities of either staying in the cold state or shifting to the hot state. We limit the combination of ad characteristics and targeting options to the following eleven combinations: a pure obtrusive ad, content integration with awareness-, product- and sales-related ads, content amplification with awareness-, product- and sales-related ads, behavioral targeting with awareness-, product and sales-related ads, retargeting with sales-related ads and obtrusive ads.

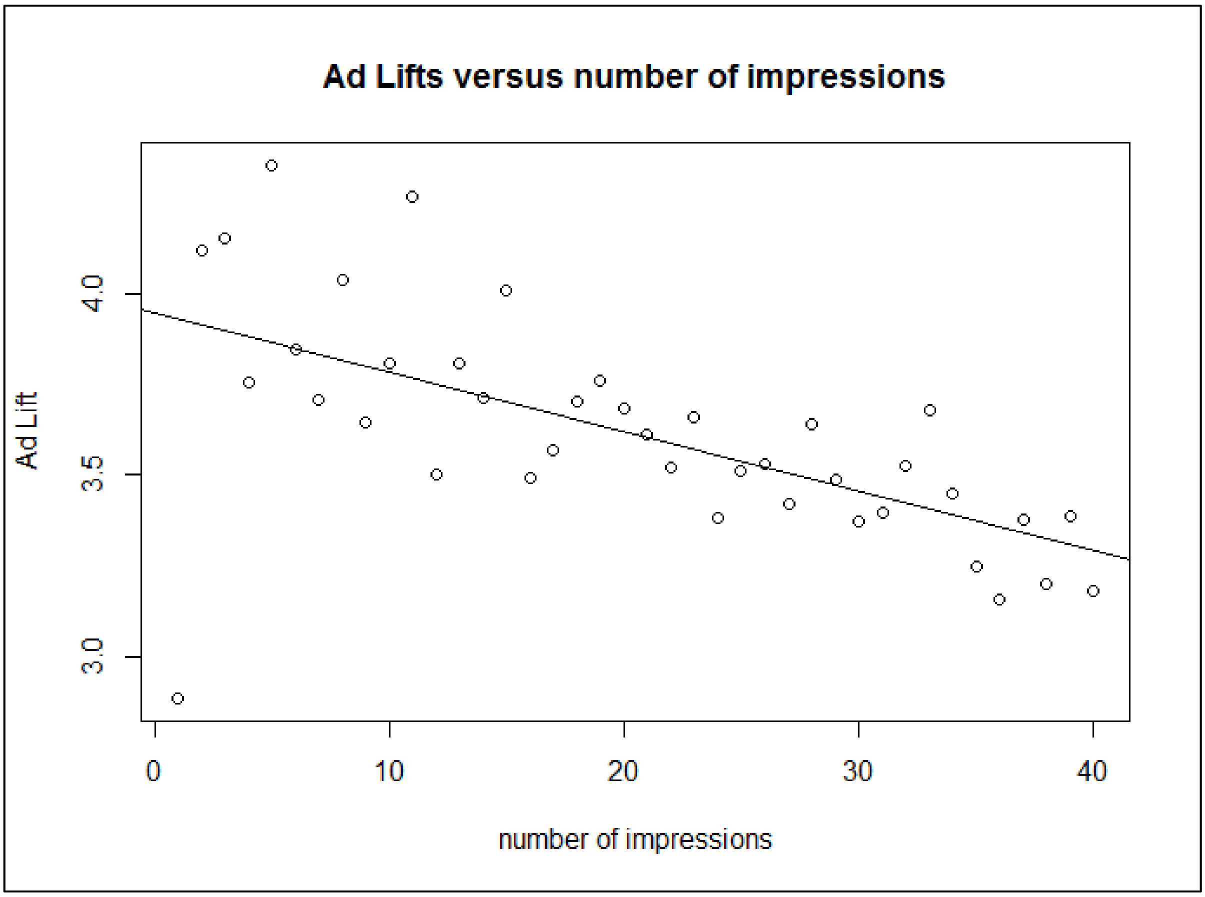
Based on the possible combinations, we calculate all transition probabilities. We display the ad that has the highest probability to shift the consumer to the hot state. Afterwards we simulate the behavioral outcome of the consumer based on the state specific choice probability. An impression counter tracks the amount of served ads which will be considered at the next impression.

This procedure continues until we reach the maximum number of iterations. We use 10,000 simulated individuals and set the impressions history from one to forty iterations. We track all the choices and calculate the choices over the impressions. This ratio is compared to the observed ratio in the dataset. Afterwards, we calculate the lift in choices using the following equation:

$$lift = \frac{choice\_ratio\_simulated}{choice\_ratio\_observed} - 1 \qquad (28)$$

This first simulation study shows that the simulated choice rate is actually higher, than the observed choices in the dataset. Figure 2-3 depicts the lift dependent on the number of served impressions. We observe that the lift has a decreasing trend when the number of observation increases. With an increasing number of observations the probability of transitioning to the hot state decreases. Therefore, the decreasing trend is what we expect.

Figure 2-3: Simulated lift and number of served impressions



Now, we limit the exposure of obtrusive ads to a certain amount. Table 2-7 shows, that even when we limit the amount of obtrusive ads, the choice rate is still higher than the choices in the observed dataset. We observe that the effect is very similar for fewer impressions. But as impressions increase, the lift ratio also increases with an increasing number of obtrusive ads.

Table 2-7: Ad lift with obtrusive ad threshold

Obtrusive ad threshold/impressions	5	10	15	20	25	30	35	40
1	3.73	3.20	2.70	2.48	2.40	2.28	2.10	1.81
2	3.87	3.52	3.01	2.85	2.57	2.51	2.45	2.23
3	3.96	3.80	3.15	3.29	3.13	2.67	2.65	2.58
4	3.78	3.67	3.71	3.28	2.94	2.95	2.52	2.66
5	3.81	4.04	3.55	3.36	2.92	2.97	2.71	2.98

To decide when to stop serving ads to consumers, we define a threshold probability. When the probability of being transferred to the hot state is lower than the previously defined threshold (and therefore the chance of having an interested consumer is low), the simulation stops for the individual. We do that to calculate an individual frequency cap for every individual.

Table 2-8 shows the lifts for different thresholds. As the threshold probability increases, the lift also increases. This happens, because if the probability of transitioning to the hot state is below the threshold, the serving of ad stops. Thus, the question arises how many ads are served to an individual. Table 2-9 shows how many impressions were served to a consumer using different thresholds for a maximum of ten impressions. We see that when the probability threshold is too low, every consumer sees every ad (every individual receives 10 ad impressions). However, when the threshold gets higher, more and more consumers stop seeing display ads (e.g., when the advertiser decides the threshold to be 0.005, 8,603 individuals receive seven ads, 540 individuals receive eight ads etc.). It goes up to a point, where the probability is so high that no consumer sees the ad (last row of Table 2-9).

Advertisers can use this simulation to calculate how many ads a certain consumer has to see or when it is better to stop serving ads. Lifts go up tremendously, but this due to the fewer impressions that are served in total to the consumers. The advertiser can use this to design schedules for different campaigns. The threshold can be lower for pure awareness campaigns, where they use other metrics than clicks and websites visits. However, if the campaign is purely focused on clicks and visits, the probability should be higher to generate higher choice ratios due to fewer impressions.

Table 2-8: Lifts for different probability thresholds and impressions

Probability threshold/impressions	5	10	15	20	25	30	35	40
0.001	4.06	3.83	3.41	3.28	3.30	3.97	4.44	5.02
0.005	4.05	5.12	7.71	8.08	10.17	10.80	11.69	11.75
0.01	3.87	6.78	9.67	11.30	11.84	12.02	14.50	15.55
0.05	10.57	15.51	18.82	20.90	21.47	21.67	23.80	24.23

Table 2-9: Number of served impressions for different probability thresholds

Probability threshold/impressions served	0	1	2	3	4	5	6	7	8	9	10
0.001											10,000
0.005								8,603	540	50	807
0.01						9,019	84	58	56	36	747
0.05		9123	216	123	70	51	36	36	23	8	314
0.1	10,000										

2.6 Summary

2.6.1 General discussion

The aim of this chapter was to assess the effect of display ad characteristics and targeting options on consumers' behavioral response considering the consumers' unobservable interest in the product. We developed a model that includes the unobservable interest in the product in an advertising scenario. We used a real-world data set to estimate the model and conducted scenario analyses to see whether we can improve clicks and visits of consumers. We find two interest states in the model: a cold state, where the consumer is less interested and a hot state, where the consumer is more interested in the product. We do not find a state, where a consumer is highly likely to react to online display advertising. But since click-rates are low and website visits attributed to display ads are also low, this is not surprising.

Moreover, we find that sales-related messages are most effective when the consumer is in the cold state. Interestingly brand-related ads only slightly increase the switching probability. Moreover, the ad message is not significant when the consumer is in the hot state. Then, the targeting strategy is more important. Strategies like content amplification and behavioral targeting increase the probability of a positive reaction of the consumer. Furthermore, a good way to reach the consumer is the usage of obtrusive formats. These ads have a positive influence in both, the cold and the hot state. Content integrated ads have a negative influence on consumer's interest in both, the cold and the hot state. Furthermore, we find that retargeted ads have a negative effect in both, the cold and the hot state. This is not surprising because retargeting can be annoying and is seldom in the interest of the consumer. It may work when the consumer is close to purchase a suggested by Lambrecht and Tucker (2013). But in our case we were more interested in the reaction behavior before an actual purchase, namely a click or a website visit.

To sum it up, we give managers a new method, with which they can calculate the unobservable interest state of the consumer using only clickstream data. Based on the identified interest, they can fine tune how to target a consumer and what kind of ad is served. Moreover, we included contact to other advertising exposures, as well as previous websites and the effect of multiple ad encounters. It is a dynamic model, where only the interest threshold of the consumer has to be calculated.

2.6.2 Managerial implications

The HMM which estimates the latent interest in the product has several important insights for managers. We observe that the latent interest changes the way consumers react to several ad characteristics and the way they were targeted. Advertisers should design the display ads to match the consumer's interest state. They should also track how often a consumer was in contact with the brand. Every ad encounter that did not cause a positive behavioral outcome decreases the probability of a click or website visit. The advertising industry should take a step back of showing as many ads as possible to consumers and should start showing *relevant* ads. We find in our scenario analyses that showing less, but relevant ads increase the positive behavioral outcome (i.e., a click or a website visit). The threshold when to stop showing ads has to be defined in order to match the communication goal of the advertiser.

Especially the trending targeting strategy of retargeting should be limited. We find that it is neither effective in the cold state, nor in the hot state. This is in line with previous findings that retargeting is only effective when consumers are close to purchase (Lambrecht and Tucker 2013).

In terms of ad message, the display ad should have sales-related information. Surprisingly, we find that in the low interest state sales-related ads increase the interest in the product. A possible explanation is that consumers directly want to see an offer that gives them information about the costs. Since display ads have only limited space to tell a message, price related information may be the strongest argument to have a deeper look at the product.

2.6.3 Limitations and further research

Like every empirical study, this chapter faces several limitations. First of all, advertisers have more information about the cookies than we have in our dataset. Therefore, they can run the model with more variables to fine-tune their serving algorithms. These variables can include size of brand name (Pieters and Wedel 2004), pictorials (Pieters and Wedel 2004), information content (Urban et al. 2014), animations and format (Hong et al. 2004; Kuisma et al. 2010), size and position (Kuisma et al. 2010), color schemes (van der Lans et al. 2014; Wedel and Pieters 2015), and visual complexity (Pieters et al. 2010).

Furthermore, this chapter used data from the past and no actual field experiment, where we can control the serving of the display impressions. However, we conducted some simulation studies to show that choice rates will be higher using our model. Future research should use this model in a field experiment to support the results of our simulations.

Another limitation is the use of cookies. For now this is a limitation for every online advertising study using clickstream data since a cookie does not necessarily represent a single consumer. Families can share a computer and therefore share one cookie. Also, some users tend to delete their cookies on a regular basis. We tried to tackle this point by using a sufficient individual path length, but still this remains to be an issue. Also modeling wise, cookies can cause several problems. In field experiments, we start at a certain date to analyze the data. Therefore, a left-censoring problem might apply. We tried to tackle this issue by assuming that every consumer starts with a low interest, but further research could try to control for the censoring problem.

Future research should also use this dynamic model to an integrated marketing communication approach. In this chapter, we only used display advertising and controlled for other advertising effects as far as we had the information in the data. But it would be interesting to combine search campaigns, social media campaigns and other digital (or non-digital) exposures and add them to the model.

Despite those limitations, this study provides an important contribution to theory and practice because it shows that display advertising can be more effective when it matches the consumer's latent interest state. This is positive for both, the consumer and the advertiser because consumers see less irrelevant ads and companies can save costs by not serving those ads.

3 User device and ad response: The moderating role of ad position

3.1 Introduction

Consumers use different devices to visit websites (Ratcliff 2014). Whereas in the past, consumers mostly used desktop computers and laptops to go online, nowadays the usage of smartphones and tablets increases. In the following, we call desktop computers and laptops traditional devices to go online and smartphones and tablets mobile devices.

Approximately 90% of US adults own a smartphone and spend more than 4 hours a day using it (Chang 2015; McDonough 2016). Thus, a new opportunity for marketers arises to reach the target audience. Marketers exploit the new technological features to use location based services (e.g., Andrews et al. 2015; Fong et al. 2015; Luo et al. 2014). But they also still use classical advertising approaches like display ads (Bart et al. 2014), and these approaches still account for the highest shares in mobile advertising spending (eMarketer 2015).

Yet, opinions about the effectiveness of display ads served on mobile devices are mixed (Heine 2013; di Girogi 2015). The reason for the discrepancy is that marketers currently do not know how to use the mobile channel for advertising. Some companies just try do to "something" and see what happens without giving it a deeper thought (Patel et al. 2013). It is common practice that marketers use similar approaches for mobile and traditional devices when it comes to display ads (Del Rey 2012). However there is evidence, that consumers react differently to ads served on mobile devices compared to traditional devices (Ghose, Goldfarb, et al. 2013; Shankar and Balasubramanian 2009) Previous studies indicate that search costs are higher for mobile compared to the traditional devices because of the smaller screen size (Ghose, Goldfarb, et al. 2013). As a consequence, the position of an ad might be more critical for ads served on mobile than on traditional devices.

Despite being highly relevant for marketers, there is no study that compares consumers' reactions to ads served on mobile and traditional devices. Literature on mobile ad effectiveness rather focusses on the effectiveness of location-based advertising messages (Andrews et al. 2016; Fong et al. 2015; Luo et al.

2014). Thus, we have little knowledge whether marketers should consider the type of device a consumer is using when serving display ads.

It is aim of this study to examine whether the effectiveness of display ads is affected by the device consumers are using. We use clickstream data of 14,611 cookie IDs to investigate how consumers react to ads served on a German nutrition-related website. About 52% of the website visitors used traditional devices, and about 48% used mobile devices. We used two different ads and two different positions on the website to study whether consumers react differently to ads served on traditional compared to mobile devices. The results show that in general consumers click less on display ads when using a mobile device. However, a display ad served further down the website increases the probability to be clicked for the mobile channel compared to the traditional channel. There is no significant difference between ads at the top of the page between these two channels. These insights contribute to our understanding about the effectiveness of ads served on mobile devices.

The remainder of the study is organized as follows. We first discuss differences between traditional and mobile devices and the potential effect of these differences for ad effectiveness. Afterwards, we present the conceptual framework and derive hypotheses for the study. Then, we describe the empirical study and present the results. This chapter closes with managerial implications and limitations of the study.

In this chapter, we find that mobile display advertising in general is less effective in terms of clicks. However, the ad position moderates the effectiveness of the device. The effectiveness also differs for returning visitors. This means there are differences between the traditional and the mobile channel.

3.2 Differences between the mobile and traditional online channel

The following section describes the differences between traditional and mobile devices. More specifically, the devices differ with respect to the usage situation (Hart 2014), technological features (Dhar and Varshney 2011) and consumer behavior (Lambrea 2016).

3.2.1 Difference in usage situation

Consumers have different motives to use a service or technology. The Uses and Gratifications (U&G) theory aims to explain why consumers use particular media to satisfy specific needs (Katz 1959). It has been applied to mass media like TV, radio or newspaper (Berelson 1949; Canrtril and Allport 1935; Rubin 1984). Previous research finds that consumers use mass media mostly for experience of the usage itself or for the information that they gather during the process. These gratifications are classified as process gratification and content gratification (Sutanto et al. 2013). Another classification for media usage contains instrumental and ritualistic use of media (Rubin 1984). Where instrumental use is goal-directed and purposeful, ritualistic use is habitual and diversionary (Hiniker et al. 2016). Stafford et al. (2004) further identify social gratification as a unique gratification related to Internet use. Results related to mobile Internet usage suggest that gains in efficiency and accessibility are the primary gratifications that arise from mobile usage (Stafford and Gillenson 2004).

Above and beyond the differences in motivations and gratifications, consumers use their mobile device in different situations compared to traditional devices (Hart 2014). While they use traditional devices most often sitting at a desk or at work, mobile devices are often used either at home sitting on the couch or on the run. Mobile devices are also used to bridge time gaps like waiting for a bus. This has an important implication for advertisers. Whereas a consumer is more focused on the device while using traditional devices, he or she might be easily distracted while using mobile devices (Hart 2014). This difference in usage situations also affects consumers' decision-making process when it comes to buying decisions. Consumers use mobile devices for information search but rather do not buy a product using the mobile device – especially when the buying decision requires more consideration (Wang et al. 2015).

Finally, the differences in usage situations affect the time when consumers use the devices. Mobile device usage peaks during morning and evening and desktop devices are mostly used over the day (Chaffey 2016).

3.2.2 Difference in technological features

The devices differ with respect to technological features. As the name "mobile" suggests, the user carries around the device wherever he or she goes. The location of the device can easily be tracked by GPS or other location technologies (Dhar and Varshney 2011).

Furthermore, the physical appearance of the devices is fundamentally different. Whereas traditional devices have a rather big screen, mobile devices have rather small screens (Hart 2014). A classical desktop computer has an external monitor ranging from 20” to 30”, whereas a laptop is normally smaller, ranging from netbooks (9”-11”) to normal laptops (13”-17”). Mobile devices are smaller, ranging from 4”-7” for smartphones and 7”-10” for tablets. This also leads to a different pixel resolution. Typically, mobile screens have fewer pixels compared to traditional devices.

Another difference is how consumers interact with the devices (Grewal et al. 2016): traditional devices can be used via keyboard or mouse, whereas most mobile devices only offer a touchscreen. Those touch-based interactions can increase ownership effects, meaning that those devices are perceived as more personal (Brasel and Gips 2014).

3.2.3 Difference in consumers’ reactions to advertising

Consumers’ may react differently to ads served on mobile and traditional devices (Ghose, Han, et al. 2013). First, a reason might be that mobile screens are smaller, and the *ad format* might play a crucial role. There are different guidelines for traditional display ads and mobile display ads (IAB 2015). Second, *ad position* might also affect consumers’ reactions to ads served on mobile and traditional devices. Since mobile websites have a different design compared to traditional websites, advertisers face different decisions where to place an ad on the website (Hart 2014). Third, in a cluttered Internet environment marketers have to stand out to reach the consumer. Consumers can evaluate within 100 milliseconds whether something is an ad or not (Pieters and Wedel 2012). Thus, another factor that can influence the behavioral outcome is whether the ad is *obtrusive* or not. Making advertising more salient can influence the consumer’s perception of the ad.

Ad format

There are several studies for traditional online advertising that try to overcome the “banner blindness” phenomenon. Display ads that feature pictorials receive higher attention than pure text ads (Goodrich 2011). According to the format, skyscraper formats (vertical ad formats) receive a higher visual attention than banner formats (horizontal ad format) (Kuisma et al. 2010). When coming to more obtrusive formats, purchase intention increases (Goldfarb and Tucker 2011). But the effects depend on what kind of target audience processes the display ads. The targeting option seems to moderate the effect of ad formats. When serving obtrusive ads, the context of the website should not match the context of the display ad (Goldfarb and Tucker 2011). Also standard formats differ in their response based on the target group (Bruce et al. 2016). Whereas, horizontal ads are best for retargeted consumers, rectangular ads work best when the advertiser targets the audience via age.

It is unclear how these formats work in a mobile environment. Banner formats and skyscrapers are not listed in the mobile guidelines for smartphones (IAB 2015). However, a skyscraper-like ad can be used on tablets. Here it is important to differentiate between those devices. Whereas a smartphone usually has a portrait screen orientation (height is larger than breadth), a tablet (and a computer) inherits a portrait screen orientation (breadth is larger than height). Until now, there is no academic evidence which ad formats work best for the mobile devices.

Ad position

Prior research indicates that ad position is important in general (Drèze and Hussherr 2003) but actually academic research on ad position effects is sparse. Research shows that ads served on the right of a website are viewed more often than ads that are placed above the text (Simola et al. 2011; Goodrich 2011). A possible explanation can be that consumers skip the ad that is placed above the text they are interested in. An ad that is on the right can still be in the peripheral vision of a consumer because it is still close to the entire text. Studies show that consumers even process advertising unconsciously when they do not look at them (Yoo 2008). Ads served on mobile devices are often placed within the text and might, thus, be processed more often. However, there is no study that investigates the effect of different ad positions considering the devices on which the ad is served.

Obtrusiveness

Conditional on the ad format animated ads differ in attention. Animated skyscraper ads increase the attention toward a display ad, whereas vertical banners decrease the attention (Kuisma et al. 2010). The website where the display ad is served also influences the effectiveness. Obtrusive ads on a content-related website are less effective than obtrusive ads on websites which are unrelated to the content of the ad (Goldfarb and Tucker 2011). Overall, previous research seems to suggest that obtrusive ads are often not effective (Kuisma et al. 2010).

3.3 Conceptual framework and hypotheses

The Uses and Gratifications theory together with the Elaboration Likelihood Model (ELM) can be used to derive hypotheses regarding differences in consumers' reactions to display ads served on mobile and traditional devices. The ELM is a persuasion theory and explains attitude formation and attitude changing (Petty and Cacioppo 1986). Within the ELM there are two different routes of processing that affect the attitude formation and change: *central route processing* and *peripheral route processing*. When consumers follow central route processing, they are motivated, able and have the opportunity to process the ad. The depth of processing is high, resulting in a high use of cognitive resources. Contrarily, when consumers follow peripheral route processing, they have little motivation, ability and opportunity to process the ad. The depth of processing is rather low and does not use much cognitive resources.

In general, advertising is presented in a noisy, cluttered environment (Pieters et al. 2002, 2007). This also holds for display advertising. The challenge is to attract consumer's attention to be able to elicit a positive reaction towards the ad (i.e., click). Such a favorable reaction seems more likely when consumers follow central route processing. Contextual targeted ads, that is the ad content fits the content of the website, seems to favor central route processing as suggested by Bart et al. (2014).

Central route processing further requires the ability and the opportunity to process the ad (Maclnnis and Jaworski 1989). However, given the smaller size of ads served on mobile devices and the situation when mobile devices are used – often crowded environments or attention is distracted from ad processing to secondary task – consumers may be less able and have less opportunity to follow central route processing when using a mobile device (Bart et al. 2014; Grewal et al. 2016; Shankar et al. 2010). Thus, a positive reaction towards the ad is less likely. This leads to the following hypothesis:

H1: Using a mobile device has a negative effect on the probability to click on contextual targeted ads.

Ad position might moderate the effect of type of device on consumers' reactions to display ads. Ads at the top of a website impede reading the actual content of the website. Given the smaller screen size of mobile devices, an ad at the top of the website impedes the reading of the actual content of the website even more than on mobile devices compared to traditional devices. As a consequence, consumers may scroll down immediately to grasp the content they are actually interested in. Such a behavior reduces the probability that the ad is processed and that consumers react positively to the ad (MacInnis and Jaworski 1989). Thus, we propose:

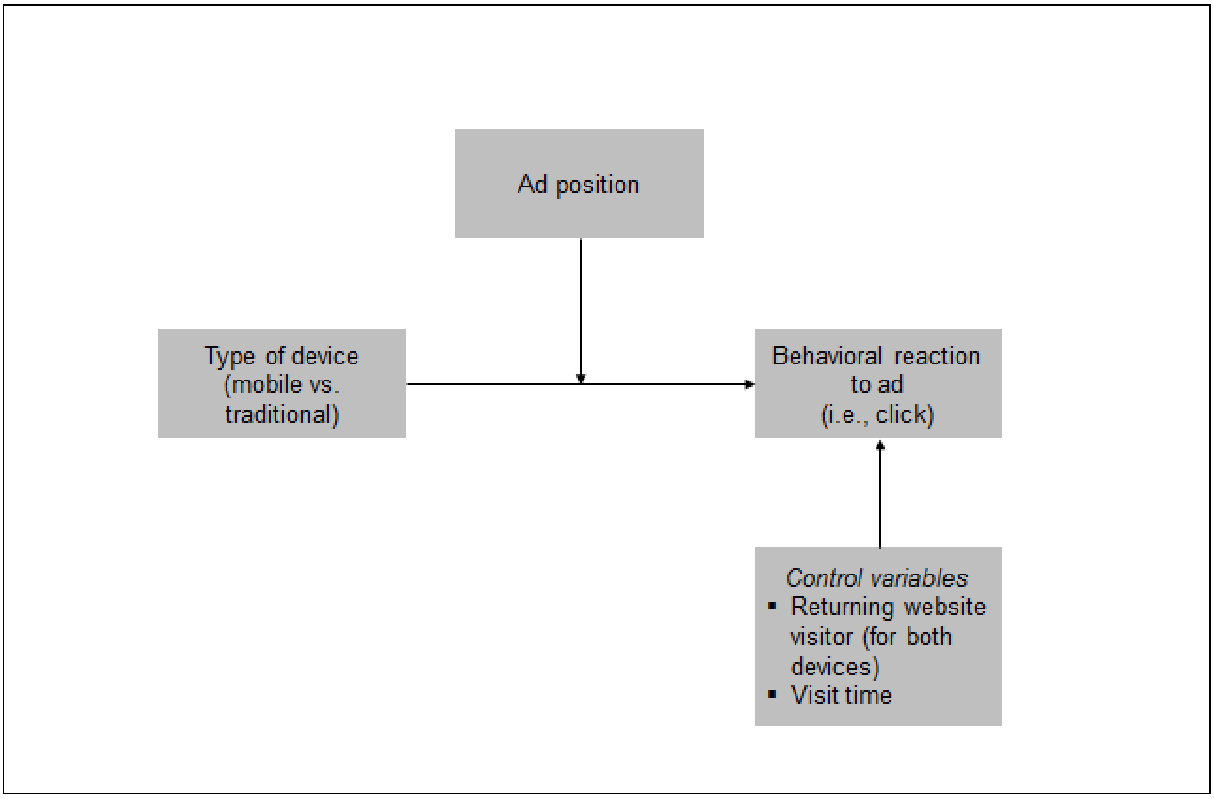
H2a: The negative effect of using a mobile device on the probability to click on contextual targeted ads is strengthened when the ad is positioned at the top of the website.

However, positioning a contextual targeted ad within the text of a website might result in differential effects. Consumers who are interested in the website content, focus on the actual content and pay attention on the reading task. When the attention is shifted to the task of extracting information from the website, consumers often perceive telepresence where they actually forget where they are (Novak et al. 2000; Steuer 1992). This decreases distraction probability and increases the ability for mobile users to process information. Mobile screens are smaller than traditional screens and have a lower pixel resolution (Ghose, Goldfarb, et al. 2013). When the ad has an equal amount of pixels, the share of the screen that is used by the display ad is higher for mobile screens. Therefore, the opportunity for processing the display ad is higher for mobile users. Therefore, for contextual targeted ads that are embedded in the text, processing and appositive reaction to the ad might be more likely when using a mobile device. Therefore, we suggest:

H2b: The negative effect of using a mobile device on the probability to click on contextual targeted ads is weakened when the ad is embedded in the text of the website.

Since consumers' reaction to an ad might depend on their familiarity with the website and their interest in the content of a website, we control for returning visits and visit time. Consumers who are returning to a website are probably highly interested in the content of the website. This interest might positively influence the processing of contextual targeted ads (Goldfarb and Tucker 2011; Lambrecht and Tucker 2013). Returning visitors of the website are already familiar with the website structure and may experience less cognitive burdens, increasing both the opportunity and ability to process the ad information – independent of the device used. Thus, returning visitors compared to first time visitors may have a higher likelihood to process the ad and thus also a higher likelihood to react positively to the ad. Moreover, the visit time (i.e., time spend on the website) is an indicator for a consumer's interest in the content of the website, and might affect consumers' reaction to an ad positively. The conceptual framework is depicted in Figure 3-1.

Figure 3-1: Conceptual framework



3.4 Study design

We design a quasi-experiment to investigate whether the type of device used to visit a website affects consumers' reactions (i.e., click) to an ad. More specifically, we consider an informational website and focus on contextual targeted ads to ensure a general interest in the ad content. The experiment was conducted on the German website www.meinstoffwechsel.com. During the time of the experiment, the website consisted of 25 webpages, from which 21 pages had editorial content and banner ads. The remaining 4 webpages did not include any banner ad. The website itself deals with health topics and has therefore a very specific target audience. This means that consumers who visit the website have a certain interest in health related topic and are more interested in the content than an average user. This differentiates the website from news portals which have all kinds of consumers on their webpage.

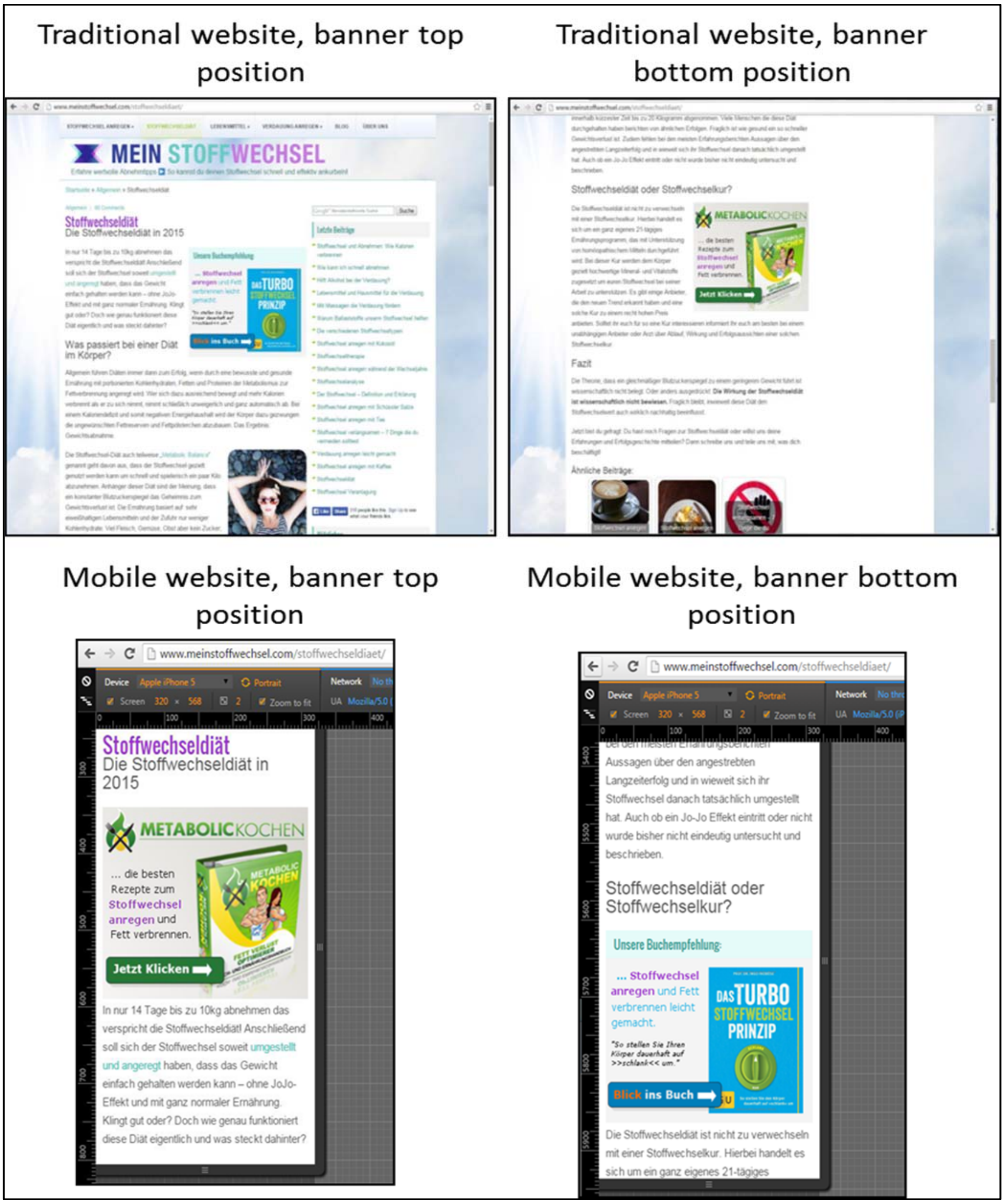
We placed two banner ads on the 21 webpages on the website. The first ad was placed at the top of the website and the second ad was placed within the editorial content, further down at the page.

The data was collected between August 8th 2015 and September 6th 2015. Within that period, the two banner ads were the only ads that were shown on that website. The ad position was randomized.

The banner ads itself were designed to match the website's content to increase clicks (Goldfarb and Tucker 2011). In particular, both ads advertised books that focus on human metabolism. One ad advertised a book that was published by a regular publisher and the other ad advertised an ebook by two fitness coaches. To avoid too much annoyance and obtrusiveness, both ads were static (Goldfarb and Tucker 2011; Kuisma et al. 2010). We used standard ad formats for both ads, that is, 300 x 250 pixels. According to Google (2015) this format works well for desktop and mobile websites.

Figure 3-2 shows how the website looked like in the traditional online channel and the mobile channel.

Figure 3-2: Screenshot of the experimental environment



We collected clickstream information to extract consumers' reactions. Since we used a field experiment, we have no further information about the individual consumer visiting the website. But in most cases this will be the situation for advertisers.

The ad exposure is exogenously determined because the position of the ad (top or bottom) and the ad creative (published book vs. ebook) are completely randomized. However, the choice of the device (mobile vs. traditional) is beyond our control. Thus, the participants self-select the device they are using based on their motivations. We, thus, have to control for self-selection effects.

3.4.1 Data collection

The data was collected by tracking cookies that are stored on the users' device during the first visit. Cookies are small text files that contain a certain sequence of numbers to uniquely identify the consumer. This is a standard procedure in web-tracking and a reliable way to track users without requiring a registration on the webpage. The consumer does not notice that a cookie is placed onto the device. If a consumer does not want to have cookie placed on the device, they can change the browser options to not allow cookies from external websites. Furthermore, a small hint has to be placed on the website to inform the consumer about the cookies.

It is often the case, that web crawlers, bots and spiders gather information from a website. That is often used to gather large scale information or extract certain parts of a website. In this experiment, it also happened that the website was visited by such technology. Therefore, we exclude the most common bots, crawlers and spiders.

When consumers visited the website, the first display ad was served at the top position. This counts as the first impression. An impression at the bottom position only counts when the consumer actually scrolled down and had the possibility to see ad. This is common practice, e.g., Google only counts an impression when at least 50% of the ad are visible for more than a second (Google 2015b). The final dataset comprises information about the total number of impressions per ad position and the total number of sessions. We only have the information which ad was served at what position when the ad was actually clicked. Thus, we cannot calculate effects on the individual level for each impression but for every cookie.

3.4.2 Data description

In total, 14,641 unique consumers visited the website during the data gathering period. We excluded 30 cookies from the analyses because they did not have an impression count for the top position. This can be due to measurement errors. From the remaining 14,611 cookies, 7,595 visited the website using a traditional device and 7,016 visited the website using a mobile device. The share of cookies that had a positive ad response (i.e., click) is 4.5% for mobile users and 9.5% for desktop users. The average impressions for the top (bottom) position are 1.55 (1.13) for mobile users and 1.70 (1.26) for desktop users. Most of the consumers visited the website only once during the data gathering period. The average number of sessions is 1.70 for mobile users and 1.17 for desktop users. The average visit time is 404 seconds across all sessions for mobile users and 299 seconds across all sessions for desktop users. Table 3-1 shows the descriptive statistics for both devices, including the aggregated information.

Table 3-1: Descriptive statistics

	Mobile				Desktop			
	Mini-mum	Maxi-mum	Mean (relative frequency)	Variance	Minimum	Maxi-mum	Mean (relative frequency)	Variance
Clicks per cookie	0	1	0.04	-	0	1	0.09	-
Impressions top position	1	16	1.55	1.41	1	29	1.70	2.26
Impressions bottom position	0	16	1.13	1.16	0	25	1.26	1.33
Number of sessions	1	56	1.70	12.36	1	44	1.17	2.31
Visit time (in seconds)	6	7,231	404.00	479,883.00	6	9,732	299.00	314,106.00
Returning visitor	0	1	0.14	-	0	1	0.06	-

3.4.3 Research method

The purpose of this study is to examine whether there are differences between the ad response behavior of consumers across devices. Consumers visit the website and see one out of two ads at the top of the website. When they scroll down, they see another ad at the bottom position. The desired outcome is a click on either ad. We define the binary outcome using the standard latent variable formulation:

$$y = \begin{cases} 1, & \text{if } y^* > 0 \\ 0, & \text{otherwise} \end{cases}, \tag{29}$$

where y^* is an unobservable variable and y whether the ad was clicked or not. This variable is influenced by the device, the impression counts for the top and bottom position, whether it is a returning visitor or not, the interaction effect of device and the aforementioned variables, and the visit time to control for the duration a consumer spend on the website. The equation for the unobservable variable y^* is:

$$\begin{aligned} y^* = \alpha \\ &+\beta_1 mobile \\ &+\beta_2 impression\_top \\ &+\beta_3 impression\_bottom \\ &+\beta_4 returning\_visitor \\ &+\beta_5 mobile * impression\_top \\ &+\beta_6 mobile * impression\_bottom \\ &+\beta_7 returning\_visitor * mobile \\ &+\beta_8 visit\_time \\ &+\varepsilon \end{aligned} \tag{30}$$

where α is the intercept, β_i are the regression coefficients and ε is the error term. We use a probit model and, thus, ε follows a normal distribution.[7] We

[7] We also estimated the model using extreme value distributed errors and the results remain similar.

consider an interaction effect of the device and returning visitor for the sake of being complete. However, we expect that this effect is not significant.

Endogeneity in non-linear models

The choice of the device is endogenous. In a normal linear regression, one can account for endogeneity using instrumental variables among other approaches (Wooldridge 2010). Using instruments, it is common to predict the endogenous variable and replace the variable using the predicted values of a first auxiliary regression (Wooldridge 2010, p. 97). Yet, this approach is not suitable for non-linear models (Danaher et al. 2015; Wooldridge 2010, p. 597). An approach to control for endogeneity in non-linear models is using probit residuals (Danaher et al. 2015). To do so, we have to identify variables that explain the endogenous variable using a standard probit model and calculate the residuals. Define $\widehat{mobile} = \Phi(\gamma + Z'\delta + \zeta)$, where $\widehat{mobile}$ is the estimated probability to use the mobile channel, Φ is the cumulative distribution function of the normal distribution, γ is the intercept, Z a vector of covariates, δ the effect of the covariates and ζ is the error term. Calculate the residuals $res = mobile - \widehat{mobile}$ and insert this new variable into equation (30) Note that the new variable is inserted in addition to the endogenous variable. Therefore, equation (30) changes to:

$$
\begin{aligned}
y^* = \alpha & \\
&+\beta_1 mobile \\
&+\beta_2 impression\_top \\
&+\beta_3 impression\_bottom \\
&+\beta_4 returning\_visitor \\
&+\beta_5 mobile * impression\_top \\
&+\beta_6 mobile * impression\_bottom \\
&+\beta_7 returning\_visitor * mobile \\
&+\beta_8 visit\_time \\
&+\beta_9 res \\
&+\varepsilon
\end{aligned} \tag{31}
$$

Equation (31) now inherits the “pure” effect of the mobile channel and additionally an endogeneity effect that corrects the model.

Relying on clickstream data, we have no psychographics or any other information about the consumer who visits the website. But industry reports found that mobile web traffic follows a certain pattern throughout the day (Leonard 2013). That is, there is a recurring pattern at which time of the day mobile traffic is increasing or decreasing. Therefore, we include the time of the day in hours and whether the site was visited on the weekend or not into the covariate vector Z. Time of the day consists of 23 dummy variables with midnight as the base category and weekend is a binary variable that indicates whether it is Saturday or Sunday.

Interaction effects in non-linear models

The interpretation of interaction effects needs some attention in non-linear models (Ai and Norton 2003). Statistically, an interaction is the cross-derivative of the expected value of the dependent variable for continuous interactions or discrete differences for binary interactions. But the structure of a non-linear model does not allow for such an interpretation. In the probit model the link function is the standard normal cumulative distribution. In the following, we will explain how one can interpret the results using one interaction term (for impressions of the upper position). The conditional mean of the ad response y is:[8]

$$\begin{aligned} & E\left(y \mid impression\_top, mobile, \boldsymbol{X}\right) \\ & = \Phi\begin{pmatrix} \alpha + \beta_1 impression\_top + \beta_2 mobile + \\ \beta_{12} mobile * impression\_top + \boldsymbol{X}\beta \end{pmatrix} \\ & = \Phi\left(u\right) \end{aligned} \tag{32}$$

where Φ is the standard normal cumulative distribution and $\boldsymbol{X}$ the remaining variables of equation (31).

8 For the sake of simplicity, the parameter labeling of the regression coefficients are slightly different in the following.

Since the device is a dummy variable and the number of impressions is continuous, the interaction effect is defined as follows (Norton et al. 2004):

$$\begin{aligned}
\frac{\Delta \dfrac{\partial \Phi(u)}{\partial impression_top}}{\Delta mobile} &= \frac{\left[(\beta_1 + \beta_{12} mobile)\Phi'(u)\right]}{\Delta mobile} \\
&= (\beta_1 + \beta_{12}) \\
&\quad * \Phi'\left[(\beta_1 + \beta_{12}) impression_top + \beta_2 + \boldsymbol{X}\beta\right] \\
&\quad - \beta_1 \Phi'(\beta_1 impression_top + \boldsymbol{X}\beta).
\end{aligned} \tag{33}$$

Equation (33) shows that the interaction effect is different from just the parameter of the interaction effect. There are four implications for interaction effects in this case (Ai and Norton 2003; Norton et al. 2004). First, even when the interaction coefficient β_{12} is zero, the whole interaction effect can be nonzero. Second, the significance cannot be assessed by simply calculating the t-statistic for the interaction coefficient. Third, the interaction effect depends on the value of the remaining independent variables. Fourth, the interaction effect can be positive and negative for different values of the covariates.

An alternative way is to compute the marginal effects. The marginal effects describe how the conditional probability changes when you change the value of an independent variable, holding all other independent variables constant. In non-linear models the marginal effects can be calculated as follows:

$$\zeta = \frac{\partial \Phi(u)}{\partial x_i} = \beta_i \Phi(u). \tag{34}$$

This expression also holds for the interaction term. Equation (34) indicates that the marginal effects always depend on the values of the other independent variables.

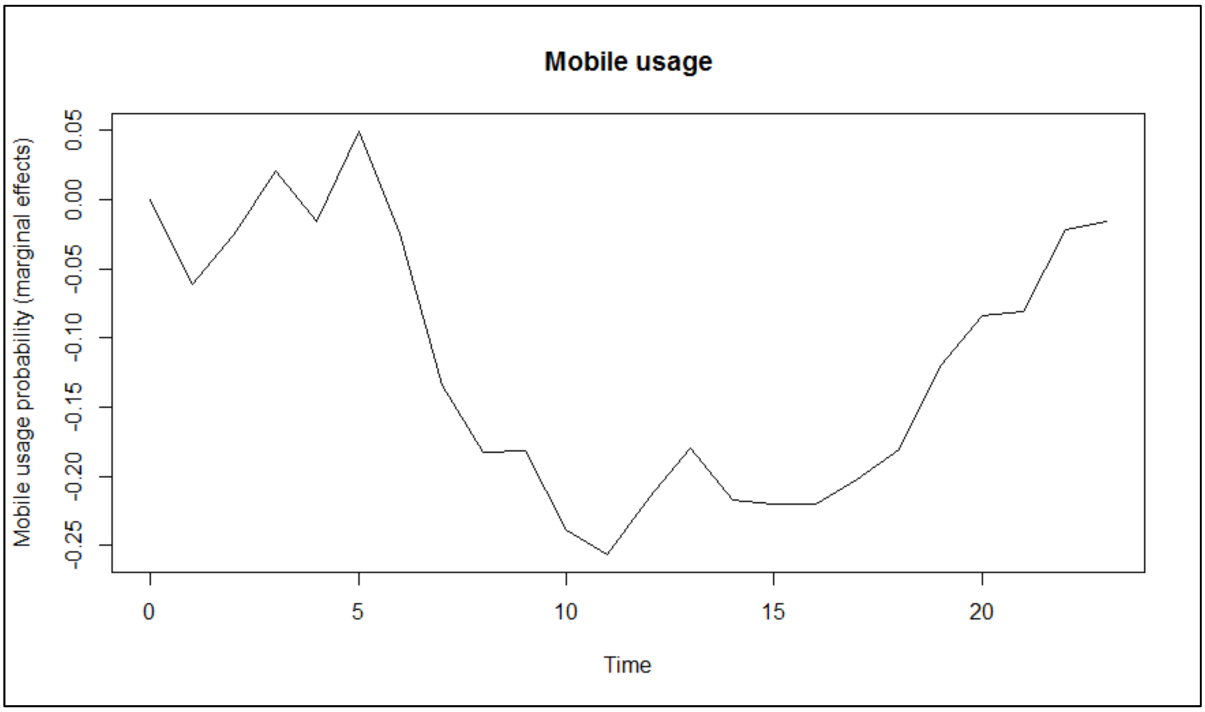
3.5 Results

We first address the endogeneity issue. Table 3-2 shows the result of the first probit regression. To interpret the parameters, we report the average marginal effects instead of the parameter estimates. Consumers use the mobile device more on weekends. For daytime, we also find a pattern of usage. Having midnight as the baseline category, we observe that in the morning hours the probability to use the device decreases. During working hours, the probability decreases even more. It slightly increases at a time where people finish work until the late evening hours.

Table 3-2: Marginal effects of mobile device usage

Variable name	Marginal effect	Standard error	p value
Weekend	0.029	0.009	0.001
1am – 2am	-0.062	0.047	0.186
2am – 3am	-0.025	0.062	0.682
3am – 4am	0.021	0.074	0.780
4am – 5am	-0.016	0.077	0.834
5am – 6am	0.050	0.063	0.434
6am – 7am	-0.025	0.046	0.586
7am – 8am	-0.134	0.035	0.000
8am – 9am	-0.183	0.030	0.000
9am – 10am	-0.182	0.029	0.000
10am – 11am	-0.239	0.026	0.000
11am – 12am	-0.257	0.025	0.000
12am – 1pm	-0.215	0.027	0.000
1pm – 2pm	-0.180	0.028	0.000
2pm – 3pm	-0.217	0.027	0.000
3pm – 4pm	-0.220	0.027	0.000
4pm – 5pm	-0.220	0.027	0.000
5pm – 6pm	-0.202	0.028	0.000
6pm – 7pm	-0.180	0.029	0.000
7pm – 8pm	-0.121	0.030	0.000
8pm – 9pm	-0.084	0.030	0.005
9pm – 10pm	-0.081	0.030	0.007
10pm – 11pm	-0.021	0.032	0.501
11pm – 12am	-0.016	0.034	0.637

Figure 3-3: Mobile device usage over daytime



After controlling for mobile device usage, we estimate the proposed model. We have two different display ads in the experiment. First, we test whether we can pool the data. Applying a Chi-Square test on the click-rate of both ads shows that the click-rates are not significantly different $\left(\chi^2 = 1.86,\ p = 0.17\right)$. Thus, we can pool our data across the two different ad creatives.

In the following analyses, we use 80% of the sample as our training sample and the remaining 20% as our validation sample (Hosmer and Lemeshow 2000). Table 3-3 shows the parameter estimates and marginal effects of the training sample. The marginal effects are calculated for every data points and are averaged over all effects.

Table 3-3: Parameter estimates and marginal effects

	Parameter (β)			Marginal Effects (ζ)		
Variable	Value	S. E.	p value	Value	S. E.	p value
Intercept	-1.650	0.112	<0.001	-	-	-
Mobile	-0.805	0.231	<0.001	-0.091	0.029	0.002
Top position	0.395	0.021	<0.001	0.045	0.002	<0.001
Bottom position	-0.249	0.026	<0.001	-0.028	0.003	<0.001
Returning visitor	-0.175	0.106	0.099	-0.018	0.010	0.068
Visit time	-0.023	0.020	0.238	-0.003	0.002	0.238
Mobile x top position	0.031	0.036	0.390	0.004	0.004	0.390
Mobile x bottom position	0.101	0.042	0.016	0.011	0.005	0.016
Mobile x returning visitor	-0.072	0.144	0.615	-0.008	0.015	0.599
Endogeneity correction	0.273	0.229	0.234	0.031	0.026	0.234

In support of H1, we observe that the probability to click on an ad is 9.1% lower when the consumer uses a mobile device compared to a traditional device $(\zeta = -0.091,\ p = 0.002)$.

In general, the probability to click on an ad increases by 4.5% for every impression that is served at the top position $(\zeta = 0.045,\ p < 0.001)$. On the contrary, the probability to click on the ad decreases by 2.8% for every impression that is served at the bottom position $(\zeta = -0.028,\ p < 0.001)$. H2a stated that the negative effect of using a mobile device on the probability to click on contextual targeted ads is strengthened when the ad is positioned at the top of the website. However, the interaction is close to zero and not significant $(\zeta = 0.004,\ p = 0.390)$. Yet, we observe that the probability to click on an ad increases by 1.1% for every impression at the bottom position when using a mobile device $(\zeta = 0.011,\ p = 0.016)$, supporting H2b.

The probability of clicking on a display ad is lower for returning visitors compared to first time visitors. Our results state that returning visitors have a 1.8%

lower probability to click on a display ad; however this result is only significant on the 10% level $(\zeta = -0.018,\ p = 0.068)$. Visit time does not affect the probability to click on the ad significantly $(\zeta = -0.003,\ p = 0.238)$.

3.6 Robustness check

We consider an alternative model specification to test the robustness of the results. One disadvantage of the proposed model above is that it assumes a linear relationship between the amount of ads served at the top and bottom position. To get a deeper understanding of the model, we change the specification of the top and bottom impressions. To capture non-linear effects of the impressions, we categorize them as dummy variables: there is a dummy variable for every outcome of the impression count up to five impressions. Six and more impressions are condensed to one variable. Notice, that for the top position, every individual has seen an ad, thus, one impression serves as the base category. For the bottom position not every consumer has scrolled so far to see an ad, thus, zero impressions serve as the base category here. Thus, the equation for the unobservable variable y^* changes to:

$$\begin{aligned} y^* = {} & \alpha + \beta_1 mobile + \sum_{i=2}^{6} \beta_{2,i} impression\_top_i \\ & + \sum_{j=1}^{6} \beta_{3,j} impression\_bottom_j + \beta_4 returning\_visitor \\ & + \sum_{i=2}^{6} \beta_{5,i} mobile * impression\_top_i \\ & + \sum_{j=1}^{6} \beta_{6,j} mobile * impression\_bottom_j \\ & + \beta_7 returning\_visitor * mobile + \beta_8 visit\_time + \beta_9 res + \varepsilon \end{aligned} \tag{35}$$

For the interaction effects, two dummy variables interact. The calculation of the interaction effect now follows the following formula (Norton et al. 2004):

$$\begin{aligned}\frac{\Delta^2\Phi(u)}{\Delta impression\_\mathrm{top}_i\,\Delta mobile} = {} & \Phi(\beta_1+\beta_2+\beta_{12}+\boldsymbol{X}\beta) \\ & -\Phi(\beta_1+\boldsymbol{X}\beta) \\ & -\Phi(\beta_2+\boldsymbol{X}\beta) \\ & +\Phi(\boldsymbol{X}\beta)\end{aligned} \tag{36}$$

Estimating (35) leads to the results presented in Table 3-4.

The substantial results do not change. Yet, we observe is that there is indeed a threshold for the amount of display ads that should be served at the bottom position in the mobile channel.

The first three impressions increase the probability to click on an ad $(\zeta_1 = 0.083,\ p < 0.001;\ \zeta_2 = 0.041,\ p = 0.027;\ \zeta_1 = 0.062,\ p = 0.018)$. But from the fourth impressions onwards, a display ad at the bottom position served on a mobile device does not increase the probability to click on an ad significantly. We do not find any significant effect for the interaction of mobile device and impressions at top positions. This is consistent with the finding in the linear model.

Table 3-4: Parameter estimates and marginal effects (dummy model)

	Parameter (β)			Marginal Effects (ζ)		
Variable	Value	S. E.	p value	Value	S. E.	p value
Intercept	-1.040	0.137	<0.001	-	-	-
Mobile	-1.214	0.275	<0.001	-0.129	0.034	<0.001
Top position: 2 impressions	0.855	0.078	<0.001	0.107	0.012	<0.001
Top position: 3 impressions	1.368	0.088	<0.001	0.231	0.022	<0.001
Top position: 4 impressions	1.370	0.129	<0.001	0.243	0.034	<0.001
Top position: 5 impressions	1.316	0.157	<0.001	0.231	0.041	<0.001
Top position: ≥6 impressions	1.522	0.163	<0.001	0.285	0.046	<0.001
Bottom position: 1 impression	-1.111	0.074	<0.001	-0.125	0.011	<0.001
Bottom position: 2 impressions	-0.464	0.083	<0.001	-0.039	0.006	<0.001
Bottom position: 3 impressions	-0.759	0.105	<0.001	-0.055	0.005	<0.001
Bottom position: 4 impressions	-0.673	0.155	<0.001	-0.048	0.007	<0.001
Bottom position: 5 impressions	-0.927	0.219	<0.001	-0.057	0.007	<0.001
Bottom position: ≥6 impressions	-0.705	0.208	0.001	-0.049	0.009	<0.001
Returning visitor	-0.301	0.112	0.007	-0.026	0.009	0.002
Visit time	-0.041	0.019	0.031	-0.004	0.002	0.031
Mobile x Top position: 2 impressions	0.058	0.128	0.650	0.006	0.013	0.657
Mobile x Top position: 3 impressions	0.121	0.147	0.410	0.013	0.016	0.434
Mobile x Top position: 4 impressions	0.462	0.208	0.026	0.057	0.031	0.067
Mobile x Top position: 5 impressions	0.500	0.258	0.053	0.063	0.040	0.116
Mobile x Top position: ≥6 impressions	0.350	0.285	0.219	0.041	0.039	0.292
Mobile x Bottom position: 1 impression	0.667	0.122	<0.001	0.083	0.018	<0.001
Mobile x Bottom position: 2 impressions	0.354	0.139	0.011	0.041	0.018	0.027
Mobile x Bottom position: 3 impressions	0.499	0.172	0.004	0.062	0.026	0.018
Mobile x Bottom position: 4 impressions	-0.055	0.270	0.838	-0.005	0.025	0.833
Mobile x Bottom position: 5 impressions	0.798	0.325	0.014	0.115	0.063	0.069
Mobile x Bottom position: ≥6 impressions	0.357	0.349	0.306	0.042	0.048	0.382
Mobile x returning visitor	-0.019	0.151	0.901	-0.002	0.015	0.900
Endogeneity correction	0.345	0.254	0.173	0.034	0.025	0.173

3.7 Model evaluation

To evaluate the overall model performance, we compute further model statistics. We start evaluating the model fit with the help AIC and BIC. The AIC for the proposed model is 5003.2, and the BIC equals 5076.87. We further use the holdout sample to examine whether the proposed model has a high predictive validity. Since we have unbalanced class sizes (we have much more non clicks than clicks), we have to determine an appropriate cutoff value. A cutoff value discriminates clickers from non-clickers. When the predicted probability to click crosses the threshold of the cutoff value, the user will be classified as a clicker and when the threshold is not crossed, the consumer will be classified as a non-clicker. We calculate the cutoff value in a way, that both, the detection of zeros and ones are best. We do so by maximizing the Youden Index which is defined as sensitivity + specificity – 1 (Youden 1950). The sensitivity is the probability of being classified as a clicker when the consumer really is a clicker. The specificity is the probability of being classified as a non-clicker, when the consumer actually is a non-clicker. Maximizing this statistics leads to an optimal cutoff value of 0.025. For unbalanced class sizes, it is normal that the cutoff value is far from 50%. Using this cutoff value for discrimination, leads to the following classification table.

Table 3-5: Classification matrix for holdout sample

	Predicted: non-clicker	Predicted: clicker	Total
Observed: non-clicker	1931	785	2716
Observed: clicker	4	203	207
Total	1935	988	2923

Most of the clickers are identified correctly. However, there are some non-clickers who are falsely identified as clickers. The sensitivity is 98.1% and the specificity equals 71.1%. Table 3-6 shows different Pseudo R^2 measures for the proposed model.

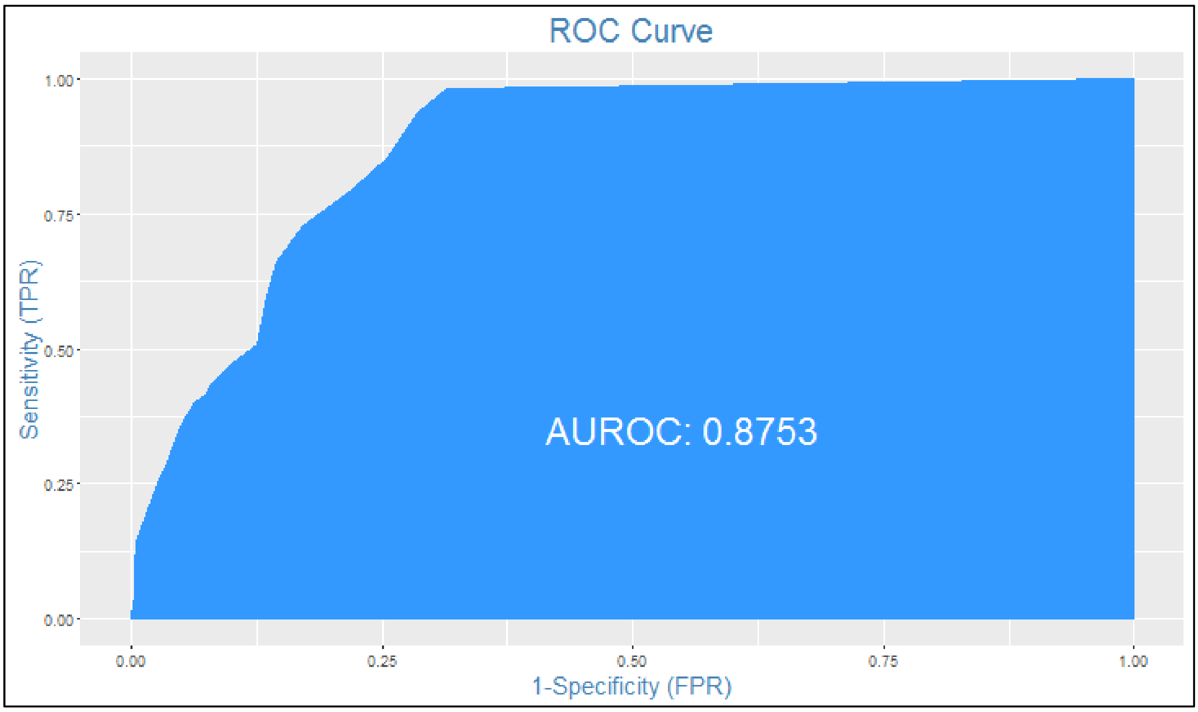
Table 3-6: Pseudo R^2 measures for the proposed model

Measure	Value
McFadden	0.28
Cox and Snell	0.13
Nagelkerke	0.33

These measures can be distorted when the class sizes differ (Christensen et al. 2014). Therefore, we continue with the Receiver Operator Characteristics (ROC) curve, which is not influenced by an unbalanced class size. The ROC curve is a graphical representation of the prediction performance with varying values of sensitivity and specificity. Thus, it is independent of a cutoff value. The goodness of fit can be evaluated with the area under the ROC curve. As sensitivity and specificity are always between zero and one, the value gives a good indication of the performance. A model without a predictive power has a slope of one and the area under the curve is 0.5. According to Hosmer and Lemeshow (2000) the area under the ROC curve should be greater than 0.7. Figure 3-4 shows the ROC curve for the holdout sample. The area under the ROC curve is 0.88 and therefore indicates a good discrimination between clickers and non-clickers.

Considering all performance metrics, the model performance appears to be good.

Figure 3-4: ROC curve for the holdout sample



3.8 Discussion

This study enhances our knowledge about the effectiveness of display ads explicitly considering the device on which the ad is served. We designed a field experiment on a German website to test whether consumers' reactions to contextual targeted ads depends on the device used. The results show that consumers are less likely to react to ads when they are using a mobile device. This is consistent with previous findings (Bart et al. 2014). However, we also find that ad position has an influence on consumers' reactions (i.e., clicks). Whereas the top position has a higher probability to be clicked, we do not find a significant difference between mobile and traditional devices. In general, the bottom position has a lower probability to be clicked. But if a consumer visits a website with a mobile device, the probability increases but only up to three impressions. Thus, marketers should rather aim for the top position on a website. However, when a consumer visits the webpage with a mobile device, they can also serve ads effectively at the bottom position but should track of the served impressions. Such an approach can be efficient for marketers since ads at the lower half of the webpage are usually cheaper (Evans 2009). We also find that returning visitors have a lower probability to click on the ad.

As all studies, this study faces some limitations. We only have aggregated information on the cookie level and no individual level data. Future research could replicate this study taking panel effects into account. Furthermore, we only have information about the clicked display ad and not about the impressions that were not clicked.

This study shows that there are differences in consumers' reactions to display ads depending on the device they are using. Future research could extend this study by testing display ads that feature unique mobile channel features. For example the mobile coupon literature shows that position effects are important (Molitor et al. 2015). Advertisers can design display ads with mobile coupon features and target them via geo-location targeting. The mobile user does not have to register at a service provider making acquisitions costs smaller. Another future challenge for marketers is to identify users across devices. Consumers often use multiple devices to access a website. It is a tough call for marketers to follow a unique user through all devices. But being able to do so, would allow for more fine-grained insights into consumers' reactions to display ads.

4 Value of user generated content and social shopping tools

The previous studies investigated how marketers can attract consumers to their website with the help of display advertising. Once they succeed to attract a consumer to their website, marketers face the challenge to stimulate purchases. To help consumers to make a purchase decision, most online retailers offer additional information next to product information. Such additional information may be product reviews and Q&A boards but also social shopping tools like collaborative shopping and sharing opportunities. For managers, it is critical to know whether these additional information and tools are related to consumers' purchasing behavior. Such knowledge provides insights into the effectiveness of the additional information and tools.

4.1 Introduction

For many retailers, the online channel has become an important distribution channel (e.g., Ailawadi and Farris 2017; Verhoef et al. 2015). Moreover, many manufacturers use their own (online) retail stores and sell their products directly to the consumers (e.g., NIKE, Hugo Boss) bypassing retail intermediaries. When retailers or manufacturers use the Internet as a distribution channel,[9] they face two major shortcomings of that particular distribution channel. First, there is a *lack of social experience* (e.g. Evans et al. 1996; Zhu et al. 2010). Social interaction of shopping is seen as one of the prime motivations for consumers to actually visit stores (Darden and Dorsch 1990) and has been found to be a crucial influencer of consumers' shopping behavior (Arnold and Reynolds 2003). While shopping, consumers have a desire to communicate with others, to share ideas and ask for feedback with their shopping companions, and to enjoy the time together with friends and family (Rohm and Swaminathan 2004; Tauber 1972). Second, there is a *lack of information about the product*. That is, customers cannot experience the product by touching or trying it. Hence, a multi-sensory experience is missing, potentially affecting consumers' shopping behavior when shopping online (Barlow et al. 2004).

[9] For simplicity, we subsequently speak of retailers, which include intermediaries as well as manufacturers that have their own retail store.

Prior studies claim that retailers have to develop ideas on how to deliver superior value to their customers through innovations that go beyond satisfying basic needs like buying a product (Reinartz et al. 2011). Such innovations should also overcome the above mentioned shortcomings of the online channel. Retailers have, thus, started to integrate user generated content (e.g., reviews, Q&A boards) into their online shops to overcome the lack of information about products. For example, customer reviews provide information about a product above and beyond the information offered by the retailer. Additionally, reviews evaluate overall product performance and specific aspects of a product. Q&A boards also address gaps in the information provided on the website. As such, reviews and Q&A boards facilitate the purchase decision for the customer by reducing uncertainty in the pre-purchase phase. This reduction in uncertainty might stimulate purchases and, furthermore, might affect product returns. If consumers make better informed decisions, the likelihood that they return a product might decrease (Minnema et al. 2016). Thus, user generated content on a retailer's website might affect customer revenue eventually. Previous studies show that product reviews affect the sales and return rate of a product (e.g., Babić Rosario et al. 2016; Floyd et al. 2014; Minnema et al. 2016). However, these studies implicitly assume that consumers consider product reviews in their purchase decisions if they are available on a retailer's website. Yet, providing reviews or Q&A boards on a retailer's website does not ensure that customers actually consider them. Given this limitation of previous studies, we lack knowledge whether the actual consideration of user generated content affects customers purchase and return behavior and ultimately their revenue. Moreover, we do not know whether revenue differs for customers who consume (passive use) or contribute (active use) user generated content. Such knowledge, however, provides interesting managerial insights. First, whether managers should stimulate the use of user generated content in the purchase decision. Second, we gain insights for targeting decisions.

In addition, retailers offer social shopping tools in their online shops to improve the social experience while shopping online. Examples for social shopping tools are collaborative shopping (Zhu et al. 2010) or the option to share products of interest with one's own social network (e.g., facebook share button). Research on social shopping tools is scarce, and focuses on the design of such tools not taking the effect on customer revenue into account (Kim et al. 2013; Zhu et al. 2010). Therefore, it is unknown whether social shopping tools can actually tackle the challenge of a lacking social experience and ultimately influence purchase and return decisions.

It is the aim of this study to examine the effect of customers' use of user generated content and social shopping tools on customer revenue considering return behavior. In doing so, we consider customers' gross revenue and product returns as well as and customers' net revenue. When studying the effects of user generated content and social shopping tools, it is moreover important to account for potential self-selection effects. We use propensity score matching to control for such effects.

Our contributions are threefold: (1) we shed light into individual-level effects of using user generated content and social shopping tools on customer revenue, (2) we consider customers' return behavior and thus the effect of user generated content and social shopping tools on the quality of customers' buying decisions, and (3) we differentiate between active and passive use of user generated content and social shopping tools when studying the effects.

We find that customers who use user generated content produce higher gross revenues for the retailer - after controlling for self-selection effects. Surprisingly, these customers also have higher return rates. However, they still generate significantly higher net revenues for the retailer. Social shopping tools have little effect on customers' purchasing and return behavior. Knowing how new technologies affect online customers' revenue and return behavior is of utmost importance and provides insights for retailers' and manufactures' (multi-)channel strategies.

This chapter is organized as follows: In the next section we develop the hypotheses. Then, we describe the data and research method. Afterwards, we discuss the results of the empirical study. Finally, the chapter closes with a summary, limitations, and areas for future research.

4.2 Conceptual background

4.2.1 Literature review

There is a vast amount of literature on user generated content and especially on online product reviews as one type of user generated content. Multiple studies have investigated the effect of online product reviews on sales. The results of these studies have been summarized by recently published meta-analyses

(Babić Rosario et al. 2016; Floyd et al. 2014). These meta-analyses show that, on average, user generated content (i.e., online product reviews) is positively correlated with sales.

However, previous research has paid little attention on the effect of product reviews on customers' decision to return a product. Only recently, a study takes the return behavior into account (Minnema et al. 2016). Minnema et al. (2016) show that product reviews have an effect above and beyond customers' purchase decision and also affect return decisions: if a product has an overly high review valence, the probability increases that the consumers send it back (Minnema et al. 2016). A possible reason is that the consumers have overly high product expectations that cannot be met with the actual product performance leading to dissatisfaction resulting in a product return.

However, all studies implicitly assume that consumers consider product reviews in their decision when they are available on a webpage. That implies that we have actually little knowledge on the effects of user generated content at the individual level. We do not know whether customers who actually consider user generated content in their purchase decisions are the better customers for a retailer, that is, that they buy more and return less because they make better informed decisions. If this is the case, the customers who actually contribute content to a retailer's webpage generate value for the retailer by providing valuable information to other consumers. Moreover, we lack knowledge about those customers who contribute content. Are those customers more profitable for the firm? We address this gap in previous research with this study.

Furthermore, we have little knowledge about the effects of social shopping tools (e.g., collaborative shopping, social share buttons) because literature on such tools is scarce. Previous studies focus on the design of social shopping tools to improve the social shopping experience. To facilitate the overall shopping enjoyment, voice chat seems to be a good design tool to increase the co-presence, i.e., the feeling of actually shopping together (Kim et al. 2013). This ultimately influences the overall shopping enjoyment and the intention to use such tools to shop together. Moreover, previous research shows that a seamless navigation is critical to ensure a pleasurable social shopping experience (Zhu et al. 2010). Yet, the studies do not investigate the effect of such tools on purchases and product returns at the consumer-level. We address this gap in literature with this study.

4.2.2 Hypotheses

This research aims to investigate the effects of user generated content and social shopping tools on customer gross revenue and return behavior, and ultimately customer net revenue (i.e., customer revenue – value of product returns). There are two ways for the customers to use user generated content. They can read posts and opinions of other customers, which we define as passive use. Moreover, they can also post their opinions and make them accessible to everyone, which is an active use of user generated content. This distinction also holds for social shopping tools. A customer can invite somebody to a digital shopping trip (i.e. collaborative shopping) or share content within a social network. Because of the active character, we define this behavior as active use. Contrary, they can get invited by friends or receive shared content, which is an example of passive use of social shopping tools.

User generated content provides additional information to customers that complements retailer information (e.g., information about price, color, size) and focuses on actual product experiences (Hennig-Thurau and Walsh 2004). Hence, *customers reading user generated content* may be better informed about products and their performance and may feel more comfortable in making a purchase decision. Thus, the probability of buying may increase. Moreover, user generated content represents a part of the broader personalized virtual community support, and finally improves the perception of social presence of the website for those consumers who read it (Kumar and Benbasat 2006; Mathwick 2002). Social presence is the extent to which a psychological connection is formed between a website and its customers, and research shows that perceived social presence increases consumers' intention to shop on that website (Kim et al. 2013). Furthermore, customers reading user generated content may perceive the website as more useful what might also affect their intention to shop at the website positively (Bickart and Schindler 2001; Kumar and Benbasat 2006). We thus propose:

> H1a: Passive use of user generated content (i.e., consuming) affects customers' gross revenue positively.

Consumers contribute user generated content (i.e., active use) because of social benefits like interacting with other consumers and the concern for others but also to vent negative feelings (Hennig-Thurau et al. 2004). Consequently, customers' contributions will differ in their valence. Contributing negative user generated content indicates negative expectation disconfirmation and thus customer dissatisfaction with the product. Nevertheless, contributing negative content requires some effort from the customer and may serve as an indication that the customer is still interested in having a relationship with the retailer (Knox and van Oest 2014).

Contributing positive content creates and improves the feeling of being part of a community that may, in turn, influence a customer's intensity of the relationship with the firm positively (McAlexander et al. 2002). Moreover members of an online community contribute word-of-mouth because of a sense of engagement (Ray et al. 2014). Engaged customers tend to be more valuable for the firm, than less engaged customers because of more transactions (Kumar et al. 2010). We thus propose that contributing content to a website will strengthen the customer relationship with the firm what, in turn, results in higher revenues (Rishika et al. 2013).

H1b: Active use of user generated content (i.e., contributing) affects customers' gross revenue positively.

A customer's decision to return a product is influenced by pre-purchase expectations. After buying the product, the customer experiences the product and forms his/her post-purchase evaluation of the product. This post-purchase evaluation may either confirm or disconfirm a customer's pre-purchase expectations (e.g., Oliver 2009). In the latter case, a customer may decide to return a product because of expectation disconfirmation.

Reading user generated content may influence customers' product expectations by adjusting pre-purchase expectations. Assuming that user generated content reflects actual product performance, consumers should form expectations that are consistent with their post-purchase product experiences. Thus, the likelihood of a gap between pre-purchase expectations and post-purchase evaluations becomes smaller. Customers reading user generated content are thus more likely to be satisfied with the product, and less likely to return it (Oliver 2009). Therefore we propose:

H2: Passive use of user generated content (i.e., consuming) affects customers' return rates negatively, i.e., the return rates are lower.

A customer's decision to contribute content is affected by his/her post-purchase product evaluation (Moe and Schweidel 2012). Customers are more likely to write reviews when their post-purchase evaluation is either high or low (Anderson 1998; Dellarocas and Narayan 2006; Moe and Schweidel 2012). Satisfaction or dissatisfaction with the product may affect customer loyalty to the retailer. While this effect is likely to be positive when customers are satisfied with the product (Yuen and Chan 2010), the effect is probably negative when customers' expectations about product performance are not met. Hence, the effect of contributing content on customers' product returns might be ambiguous. We therefore refrain from formulating a hypothesis.

Online retailers design *social shopping tools* to enhance the shopping experience. They serve customers' hedonic shopping needs. While shopping, customers are able to interact with their friends. Despite being physically separated, the customers recognize the presence of their friends. The degree of the awareness of other people in the online shopping context is called co-presence, the main driver of social presence (Biocca et al. 2001). Based on the feeling of not being alone and the hedonic character of social shopping tools, the shopping experience is enhanced (Kim et al. 2013). This makes the whole shopping more fun by increasing the intrinsic motivation probably leading to a state of flow (Hoffman and Novak 1996; Kim et al. 2013). Because of the positive shopping experience, the intention to shop at that website may increase. For those consumers who passively use social shopping tools, these tools may serve as advertising. Being invited by others to join them while shopping on that website and getting to know what products close others like is a very interactive process. Since interactivity is a main concept of flow, the occurrence of flow is very likely (Hoffman and Novak 1996; Novak et al. 2000). Moreover, this may also inspire consumers to shop at that site (Luo 2005). However, the difference of being invited or inviting friends to use social shopping tools is rather small. The process of shopping together remains the same and interactivity emerges also for passive users. Thus, we expect no difference in the actual behavior of active and passive users, and thus propose:

H3: Customers using social shopping tools (i.e., active and passive use) generate higher gross revenues compared to those not using these tools.

The use of social shopping tools does usually not provide additional information about product performance. Consumers interact with their shopping companion who may not have more information about the product than already provided on the website. Therefore, we do not expect an effect of using social shopping tools on the quality of consumers' purchase decisions by reducing a potential expectation-disconfirmation.

H4: The use of social shopping tools does not affect customers' return behavior.

4.3 Data

We use data from a large Dutch online retailer that offers user generated content as well as social shopping tools on its website, namely reviews, Q&A boards, the option to share content via social media, and collaborative shopping. We have information about the use (i.e., no, passive, active use) of user generated content and social shopping tools as well as demographical, attitudinal and behavioral data for a sample of 2,498 customers. This information was collected through an online survey, and customers of the retailer were randomly asked to participate in the study. In this survey, further information about the customers was collected such as their internet skills, shopping attitude, primary use of internet, recommendation intention and demographic variables.

Additionally, we have information about gross and net revenues as well as of product returns for the same sample for a period of about 2 years - covering the period before the survey took place. A customer's gross revenue represents his or her spending before returning any product. Net revenue is the revenue from a customer after product returns are taken into account. Based on the information about gross and net revenue, we calculate the share of product returns, which is the monetary value of product returns (gross – net revenue) divided by gross revenue. We use the share of returns instead of the value of returns because the share of returns is not affected by purchase volume. In our analysis, we use the logarithm of gross and net revenue to approximate a normal distribution. We therefore only consider customers with net revenue larger than zero, leaving us with a sample of 2,452 customers.

4.4 Research method – Propensity score matching

4.4.1 Average treatment effect and assumptions

The fundamental problem in determining the effect of active users of user generated content and social shopping tools on revenue compared to passive and non-users is that we can only observe the revenues of the customers in one state. Thus, we are not able to determine the gross and net revenue of an active user as if the same customer would be a passive or non-user. Therefore, it is not possible to compare the revenues of the customers in different states (e.g., active or passive usage) directly because it is very likely that these customers differ in their characteristics.

Nevertheless, we are interested in the difference in revenue of an active (passive) user compared to the revenue of the same user when using the tools passively (avoiding the tools). More formally, we want to calculate the average treatment effect on the treated (ATT):

$$ATT_k = E_i\left(y^1_{i,k} \mid d_i = 1\right) - E_i\left(y^0_{\mathrm{i},k} \mid d_i = 1\right),\ \forall k \in K \tag{37}$$

where d_i indicates whether the customer i is an active user or not. $E_i\left(y^1_{\mathrm{i},k} \mid d_i = 1\right)$ is the expected value of all active customers i for the (observed) outcome variable k (here customer revenue), and $E_i\left(y^0_{\mathrm{i},k} \mid d_i = 1\right)$ is the expected value of all active customers i for the (unobserved) outcome variable k as if they were not active.

The expected outcome for untreated customers is only a valid estimate for the counterfactual outcome if no self-selection effects exist. This is, for example, the case in experimental settings when the researcher can randomly assign people to a certain treatment. However, in this study, the customer decides whether to use user generated content or social shopping tools, thus we have to consider a selection bias caused by self-selection effects (Heckman and Navarro-Lozano 2004; Mithas and Krishnan 2009). This bias corresponds to the average self-selection effect (SE).

More formally:

$$\left[E_i\left(y_{i,k}^0 \mid d_i = 1\right) - E_j\left(y_{\mathrm{j},k}^0 \mid d_j = 0\right)\right] = ATT_k + SE_k,\ \forall k \in K \tag{38}$$

The right-hand side of (38) shows that if the researcher does not control for possible self-selection effects, the ATT can be overestimated or underestimated, depending on the influence of the self-selection effects.

4.4.2 Controlling for selection bias

The problem of controlling for self-selection effects in non-experimental research designs is a very important issue in the economics and econometrics literature (Gensler et al. 2013; Heckman et al. 1997; Heckman and Navarro-Lozano 2004; Mithas and Krishnan 2009; Rosenbaum and Rubin 1983, 1985). The state-of-the art method to overcome this problem is propensity score matching (Mithas and Krishnan 2009; Rosenbaum and Rubin 1983). The goal is to find a 'statistical twin' of a treated customer that has not received treatment. A statistical twin is a customer with similar observed characteristics to ensure the comparability of the outcome variables and therefore provide a valid estimate of the counterfactual outcome. But the more characteristics the researcher observes, the lower is the probability to find a similar customer. Therefore researchers condensed the characteristics to one single number, the so-called propensity score. The propensity score is calculated by a logit or probit model and represents the probability of receiving treatment (Rosenbaum and Rubin 1983). The customers are matched based on the propensity score. It is assumed that the closer the propensity scores of two customers are, the more similar they are in their characteristics. Therefore, the ATT for an outcome variable k when using propensity score matching results in:

$$ATT_k = E_i\left(y_{i,k}^1 \mid d_i = 1, \hat{p}(z_i)\right) - E_j\left(y_{\mathrm{j},k}^0 \mid d_j = 0, \hat{p}(z_i)\right),\ \forall k \in K \tag{39}$$

where $p(\cdot)$ is the estimated propensity score, and $E_j\left(y_{\mathrm{j},k}^0 \mid d_j = 0, \hat{p}(z_i)\right)$ is the estimate for the counterfactual outcome.

An assumption linked to propensity score matching is that it is sufficient to estimate the propensity score by observable covariates. Thus, unobservable characteristics driving the active content generation and social shopping tool usage of customer are ignorable. Only if this 'strong ignorability' assumption holds, the difference between active users and passive, respectively non-users for each value of the propensity score is an unbiased estimate of the treatment effect (Rosenbaum and Rubin 1983).

The idea of balancing the dataset to compare the outcome of the two groups can be evaluated by the percentage reduction in bias (PRB) using the following formula (Rosenbaum and Rubin 1985):

$$PRB_m = \left(1 - \frac{\overline{x}_{d_i=1,m}^{after} - \overline{x}_{d_i=0,m}^{after}}{\overline{x}_{d_i=1,m}^{before} - \overline{x}_{d_i=0,m}^{before}} \right) * 100, \ \forall m \in M \tag{40}$$

The larger the reduction in bias, the better is the comparability of the groups (Rosenbaum and Rubin 1984).

To conduct propensity score matching, the outcome of interest, the treatment variables and the covariates to determine the propensity score need to be defined (Mithas and Krishnan 2009).

Outcome

In the forthcoming analyses the outcomes of interest are the gross and net revenue of the customers. To ensure valid test statistics and results we take the logarithm of the revenue to approximate a normal distribution.

Treatment

The treatment is the active, passive or non-use of user generated content or social shopping tools. For our purpose, a non-user does not use any of the technologies provided on the website. A passive user only passively uses a tool without using another tool of the same kind actively and without using a tool of the other kind in any way. For example a passive user of user generated content is not allowed to use social shopping tools passively or actively. This ensures that we estimate the isolated effect of the technologies and avoid a biased influence of the other tool. An active user therefore is defined as somebody who uses at least one tool of the categories actively without using a tool of the other category. An active user is allowed to use a technology of the same category passively. For example, a customer who writes reviews can also read reviews. Table 4-1 shows the distribution among user generated content and social shopping tools. It also highlights that customers use social shopping tools less than user generated content.

Table 4-1: Customers' use of user generated content and social shopping tools

		Social shopping tools		
		no use	passive use	active use
User generated content	no use	23.49%	0.90%	1.55%
	passive use	44.37%	1.39%	1.92%
	active use	23.04%	1.06%	2.28%

Covariates

As mentioned above, the propensity score is estimated via observable covariates. Shopping in the internet is different from shopping in a brick-and-mortar store. In general, when people use (online) technologies, they need skills to do so (Fulk 1993). Every customer is able to shop in brick-and-mortar stores, but it requires some knowledge to use an online shop (Hoffman and Novak 1996; Novak et al. 2000). Additionally, they need to have certain knowledge to use user generated content and social shopping tools on the website of an online retailer. Therefore, we consider the internet skills based on the scale by Novak et al. (2000). We expect that customers, who are more familiar with the internet, use user generated content and social shopping tools more actively and provide higher revenue.

People have different attitudes toward online shopping. Some customers only shop because they have a certain need. Others shop because they are seeking for joy and fun (Hirschman and Holbrook 1982). Childers et al. (2001) studied the hedonic and utilitarian motivations for online shopping and found that both are important predictors of online attitudes. Since social shopping tools are established to make the online shopping more fun and desirable, we propose that hedonic consumers are more likely to use social shopping tools. User generated content is provided to give information to the customers and reduce uncertainty in the buying process (Li et al. 2011). They target the utilitarian needs of a customer and we expect that utilitarian users are more likely to use user generated content. Therefore, we take the attitude toward online shopping into account when estimating the propensity score (Voss et al. 2003).

Built on utilitarian and hedonic attitudes toward online shopping, we also consider a customer's main reasons to use the internet since customers have different reasons to use the internet and visit an online store (Moe 2003). We asked for relaxation, sharing, contact, research, commerce and job reasons. The first three categories represent hedonic reasons and the latter three represent utilitarian reasons to use the internet.

The internet is a huge network which allows customers to articulate themselves (Hennig-Thurau et al. 2004). A common measure for loyalty in the marketing practice is the net promoter score, where the customer is asked how likely he or she would recommend the online shop to a friend (Reichheld 2003). The more likely the customer is to recommend the online shop to a friend, the more likely is he or she to recommend the online shop on the internet by actively use any of the provided tools. Because of a good relationship with the retailer, they are also assumed to generate higher revenues. Therefore, we take the recommendation intention of consumers into account using the Net Promoter question.

Additionally, we consider demographical data, namely age and gender, since they are important in characterizing the customer and avoid bias because of unobserved heterogeneity. We also control for the number of categories in which the customer made a purchase. Table 4-2 gives an overview of the covariates used for determining the propensity score.

Table 4-2: Covariates considered estimating the propensity score

Variable	Definition
internet skills	internet skills of the customer measured with six items based on Novak et al. (2000); Cronbach's alpha = 0.815 ▪ I am extremely skilled at using the web. ▪ I consider myself knowledgeable about good search techniques on the Web. ▪ I know somewhat less than most users about the web (reversed item) ▪ I know how to find what I am looking for on the Web. ▪ How would you rate your skill at using the Web, compared to other things you do on the computer? ▪ How would you rate your skill at using the Web, compared to the sport or game you are best at?
utilitarian attitude	utilitarian attitude toward online shopping measured with five items based on Voss et al. (2003); Cronbach's alpha = 0.848 Online shopping is… … effective/not effective … helpful/not helpful … functional/not functional … necessary/not necessary … practical/not practical
hedonic attitude	hedonic attitude toward online shopping measured with five items based on Voss et al. (2003); Cronbach's alpha = 0.752 Online shopping is… … not fun/fun … dull/exciting … not delightful/delightful … not thrilling/thrilling … unenjoyable/enjoyable

Table 4-2: Covariates considered estimating the propensity score (continued)

Variable	Definition
use of internet: relax	reason if the customer uses the internet for relaxation (1=yes, 0=no)
use of internet: research	reason if the customer uses the internet for research (1=yes, 0=no)
use of internet: shopping	reason if the customer uses the internet for shopping (1=yes, 0=no)
use of internet: share	reason if the customer uses the internet for sharing content (1=yes, 0=no)
use of internet: connect	reason if the customer uses the internet for contact (1=yes, 0=no)
use of internet: job	reason if the customer uses the internet for job purposes (1=yes, 0=no)
sum of categories	number of categories in which the customer bought during the observation period
share fashion	percentage share of purchases within the fashion category
recommendation intention	recommendation intention of a customer measured on a 11-point scale
gender	gender of the customer (1=female, 0=male)
age	age of the customer in years

To ensure that the multiple items actually measure the desired construct, we first conduct a factor analysis. The details can be found in Appendix B.

To check the internal reliability, we use Cronbach's alpha (Cronbach 1951). The test statistics should be greater than 0.7, which is the case here for all extracted factors (see Table 4-2). Since all values exceed 0.7, the internal consistency of the constructs is ensured. We proceed with calculating the mean of the different items and use this statistic in the following analyses.

Table 4-3 shows the reduction in bias for every variable for user generated content and social shopping tools. The reduction in bias is substantial for the user generated content comparisons. However, for social shopping tools, five variables have a negative reduction in bias. These effects are caused by the relative small samples when analyzing social shopping tools. To be consistent in the way the propensity score matching is employed, we, however, keep all variables in the propensity score model.

To sum it up, we have the gross and net sales as well as the product returns as outcome variables. The active, passive or no use is the treatment variable and the above mentioned covariates determine the treatment.

Table 4-3: Reduction in bias

	User generated content			Social shopping tools	
Covariates	passive vs. no use	active vs. passive use	active vs. no use	passive vs. no use	active vs. no use
internet skills	95.8	88.5	99.2	94.7	47.9
use of internet:					
relax	58.7	45.8	70.2	-41.4	83.3
research	89.8	98.3	96.9	85.3	95.1
commerce	99.6	93.0	96.3	98.3	62.9
share	95.4	96.4	89.2	92.9	87.1
contact	99.3	92.3	91.9	25.5	4.6
utilitarian attitude (online shopping)	94.0	89.0	97.0	-97.7	83.9
hedonic attitude (online shopping)	91.9	91.0	95.8	1.3	98.0
recommendation intention	96.0	95.1	89.3	91.3	70.9
gender (1: female, 0: male)	9.3	95.8	75.2	-513.3	-80.6
age	97.4	57.9	98.8	-409.6	48.7
sum of categories	80.5	92.4	87.7	66	-101.6

4.4.3 Estimating the treatment effect

There are several algorithms to match treated and untreated customers. Popular ones are (1) the one-nearest neighbor algorithm, which considers just one matching partner; (2) the n-nearest neighbor algorithm that includes n matching partners, (3) the radius matching algorithm, which considers every matching partner within a defined caliper and (4) the Gaussian kernel algorithm, which considers all untreated customers as matching partners by assigning a weight based on the distance to the treated customer. There is no single best solution which algorithm to take. Matching with more than one partner leads to an increased bias in the estimated ATT, but also to a decreased variance (Caliendo and Kopeinig 2008). However, asymptotically all algorithms yield similar results (Zhao 2004).

The most common approach in the recent literature is kernel matching (Caliendo et al. 2012; Gensler et al. 2012; Mithas and Krishnan 2009). We use the kernel matching algorithm with a Gaussian kernel using a bandwidth as suggested by Silverman (1986).

To further ensure comparability of the two groups we calculate the ranges of the propensity scores for the two groups. Every case falling outside the overlapping distribution of active, passive and non-users, the region of common support, is dropped from the analysis because there is no potential matching partner (Heckman et al. 1997). As mentioned before the treatment variable consists of three states: (i) not using any technology, (ii) only passively using any technology, and (iii) actively using any technology. Since the propensity score is calculated with a binomial logit model, an appropriate approach to deal with multiple treatment cases is to estimate a series of binomial models (Lechner 2001).

4.5 Results

4.5.1 Characteristics of users of user generated content

The propensity score models allow for describing the users of user generated content and social shopping tools (Table 4-4). For user generated content, passive users compared to non-users have higher internet skills (β = 0.211, p = 0.001), use the internet for research (β = 0.329, p = 0.005) and shopping (β = 0.338, p = 0.021), which are utilitarian motivations to use the internet, buy in more categories ($\beta = 0.130$, $p = 0.009$), and recommend the online shop to others ($\beta = 0.096$, $p = 0.014$). Additionally they are younger ($\beta = -0.018$, $p < 0.001$). Active users of user generated content compared to passive users use the internet for research ($\beta = 0.241$, $p = 0.055$) and to connect with others ($\beta = 0.275$, $p = 0.019$), representing hedonic and utilitarian motivations. Additionally, they buy in more categories ($\beta = 0.139$, $p = 0.002$) and are more prone to recommend the online shop to others ($\beta = 0.094$, $p = 0.033$). Active users of user generated content tend to be women ($\beta = 0.411$, $p = 0.009$). When comparing active users to non-users of user generated content, it turns out that active users have higher internet skills ($\beta = 0.309$, $p < 0.001$) and use the internet for research ($\beta = 0.593$, $p < 0.001$), shopping ($\beta = 0.489$, $p = 0.010$) and to connect with others ($\beta = 0.368$, $p = 0.007$), representing utilitarian and hedonic motivations. Additionally, they tend to have a hedonic attitude toward online shopping ($\beta = 0.144$, $p = 0.061$), buy in more categories ($\beta = 0.257$, $p < 0.001$), and are more prone to recommend the focal retailer to others ($\beta = 0.180$, $p < 0.001$). Again, active users tend to be women ($\beta = 0.350$, $p = 0.061$).

For social shopping tools, passive users of social shopping tools use the internet less for research ($\beta = -0.933$, $p = 0.049$) and shopping ($\beta = -1.050$, $p = 0.026$) compared to non-users. Furthermore, they tend to buy less within the fashion category ($\beta = -0.011$, $p = 0.070$). Active users are categorized by the higher hedonic attitude toward online shopping ($\beta = 0.482$, $p = 0.018$).

Table 4-4: Estimation results of the logit model

	user generated content			social shopping tools	
	passive vs. no use	active vs. passive use	active vs. no use	passive vs. no use	active vs. no use
internet skills	0.211 (0.001)	0.052 (0.446)	0.309 (0.000)	0.015 (0.953)	-0.289 (0.135)
utilitarian attitude	0.011 (0.878)	0.055 (0.493)	0.029 (0.739)	-0.230 (0.447)	0.042 (0.851)
hedonic attitude	0.105 (0.117)	0.065 (0.341)	0.144 (0.061)	0.324 (0.228)	0.482 (0.018)
use of internet: (base: job) relax	0.099 (0.372)	-0.000 (0.999)	0.143 (0.285)	0.088 (0.854)	0.270 (0.468)
research	0.329 (0.005)	0.241 (0.055)	0.593 (0.000)	-0.933 (0.049)	-0.524 (0.142)
shopping	0.338 (0.021)	0.119 (0.491)	0.489 (0.010)	-1.050 (0.026)	-0.342 (0.400)
share	-0.029 (0.811)	0.007 (0.952)	-0.019 (0.894)	0.013 (0.979)	0.465 (0.209)
connect	0.073 (0.521)	0.275 (0.019)	0.368 (0.007)	-0.353 (0.452)	0.001 (0.998)
sum of categories	0.130 (0.009)	0.139 (0.002)	0.257 (0.000)	-0.390 (0.106)	-0.039 (0.816)
share fashion	0.001 (0.699)	0.002 (0.251)	0.003 (0.163)	-0.011 (0.070)	0.001 (0.884)
recommendation	0.096 (0.014)	0.094 (0.033)	0.180 (0.000)	0.247 (0.169)	0.177 (0.186)
gender	-0.148 (0.320)	0.411 (0.009)	0.350 (0.061)	0.466 (0.472)	-0.142 (0.776)
age	-0.018 (0.000)	0.008 (0.107)	-0.007 (0.226)	-0.006 (0.748)	0.019 (0.198)
N	1663	1653	1140	597	613
Pseudo R^2	0.042	0.025	0.098	0.108	0.068

p-values in parentheses.

4.5.2 Effects of user generated content on customer revenue and share of product returns

Table 4-5 shows the effects of using user generated content and social shopping tools on the logarithm of customers' gross and net revenue and the share of product returns. Passive users of user generated content produce significantly higher gross revenues for the retailer than customers who do not use user generated content ($\beta = 0.162$, $p = 0.006$). This result supports H1a. Moreover, we propose in H1b that contributing additionally to consuming user generated content (active use) results in higher gross revenues ($\beta = 0.240$, $p < 0.001$). Since we find that active users of user generated content provide higher revenues, H1b is supported. However, we do not find support for H2 that customers consuming user generated content (passive use) have lower product returns compared to those not using it ($\beta_{\text{passive}} = 0.028$, $p = 0.042$; $\beta_{\text{active}} = 0.049$, $p = 0.002$). Actually, passive and active users of user generated content have a higher share in returns than customers not using it. Yet, customers who use user generated content still produce significantly higher net revenues ($\beta_{\text{passive}} = 0.113$, $p = 0.048$; $\beta_{\text{active}} = 0.344$, $p < 0.001$), indicating, that, despite the higher return rates, the customers buy much more than customers who do not use user generated content.

With respect to social shopping tools, we proposed that customers who use social shopping tools either passively or actively generate higher gross revenues compared to those not using these tools. However, we do not find support for H3 in our data ($\beta_{\text{passive}} = 0.405$, $p = 0.214$; $\beta_{\text{active}} = 0.076$, $p = 0.765$). Yet, we find support for H4 that the use of social shopping tools does not influence customers' return behavior ($\beta_{\text{passive}} = 0.057$, $p = 0.432$; $\beta_{\text{active}} = 0.016$, $p = 0.731$).

Table 4-5: Effects of using user generated content and social shopping tools

		ln(gross revenue)	share of returns	ln(net revenue)
user generated content	passive vs. non-users	0.162 (0.006)	0.028 (0.042)	0.113 (0.048)
	active vs. non-users	0.421 (0.000)	0.049 (0.002)	0.344 (0.000)
	active vs. passive users	0.240 (0.000)	0.015 (0.226)	0.220 (0.000)
social shopping tools	passive vs. non-users	0.405 (0.214)	0.057 (0.432)	0.260 (0.418)
	active vs. non-users	0.076 (0.765)	0.016 (0.731)	0.061 (0.792)

p-values in parentheses.

4.5.3 Heterogeneity

To further investigate the effects of user generated content, we conduct additional validation and robustness checks. First, we assess if there is heterogeneity within the treatment effects. Following the approach of Rosenbaum and Rubin (1984), we from subclasses, the so-called strata, based on the estimated propensity score. Based on the lowest and highest propensity score within the range of common support, the initial solution consists of five groups with the same range (Dehejia and Wahba 2002). Based on the subclasses, we check if there are significant differences within the subclasses to ensure a comparability of the data. Following the argumentation of Becker and Ichino (2002), we use the 1% level to test the differences. The more variables are included in the estimation of the propensity score the less likely it is to balance every covariate in the sample. Assuming the balancing property of every variable is mutually independent and we would use the 5% level with 12 variables, the probability of rejecting the balancing property although it holds true is $\binom{12}{1}(0.05)^1(0.95)^{11} = 0.341$.

Table 4-6: Assessing heterogeneity using different strata

active versus non-users

Stratum	Number of non-users	Number of active users	Ln(gross revenue) of non-users	Ln(gross revenue) of active users	Difference
1	69	12	5.722	5.905	0.184
2	172	101	5.806	6.304	0.498***
3	186	143	6.107	6.437	0.330**
4	112	214	6.230	6.829	0.599***
5	32	94	7.041	7.468	0.427**

Stratum	Number of non-users	Number of active users	Ln(net revenue) of non-users	Ln(net revenue) of active users	Difference
1	69	12	5.199	5.617	0.418
2	172	101	5.366	5.854	0.487***
3	186	143	5.682	5.912	0.230*
4	112	214	5.791	6.263	0.472***
5	32	94	6.575	6.893	0.317*

active versus passive users

Stratum	Number of passive users	Number of active users	Ln(gross revenue) of passive users	Ln(gross revenue) of active users	Difference
1	139	36	5.656	6.146	0.491**
2	334	142	6.055	6.319	0.264**
3	394	205	6.380	6.521	0.141*
4	188	136	7.037	7.247	0.210**
5	30	44	7.221	7.768	0.548**

Stratum	Number of passive users	Number of active users	Ln(net revenue) of passive users	Ln(net revenue) of active users	Difference
1	139	36	5.321	5.863	0.542**
2	334	142	5.561	5.841	0.280***
3	394	205	5.876	5.970	0.094
4	188	136	6.493	6.654	0.161*
5	30	44	6.790	7.125	0.335*

passive versus non-users

Stratum	Number of non-users	Number of passive users	Ln(gross revenue) of non-users	Ln(gross revenue) of passive users	Difference
1	29	16	5.438	5.487	0.048
2	30	22	6.033	5.838	-0.195
3	52	44	5.817	6.136	0.319
4	187	269	5.947	6.046	0.099
5	208	492	6.187	6.399	0.211**
6	65	233	6.445	6.569	0.124

Stratum	Number of non-users	Number of passive users	Ln(net revenue) of non-users	Ln(net revenue) of passive users	Difference
1	29	16	4.871	5.102	0.231
2	30	22	5.436	5.370	-0.066
3	52	44	5.310	5.624	0.314
4	187	269	5.517	5.581	0.064
5	208	492	5.747	5.892	0.145*
6	65	233	6.445	6.569	0.124

This results in five strata for both active versus passive users and active versus non-users. The comparison of passive versus non-users results in six strata. Table 4-6 shows the results of the sub-classification. For both, active versus passive and active versus non-users the gross and net revenue of active users are significantly higher than their control groups with exception of strata 1. Here, the difference is still positive in all cases but not significant, which might be caused by the low number of observations in these strata. In contrast, the comparison between passive and non-users is only significant in one stratum, which means that the significance is mostly driven by this stratum. However, the other strata are still positive with the exception of the second strata. Still, there is indication that an effect exists.

4.5.4 Robustness check – Sensitivity analysis

Propensity score matching is based on the conditional independence assumption (Rosenbaum and Rubin 1983). This assumption implies that no relevant customer characteristics that influence the use of user generated content and social shopping tools are disregarded when estimating the propensity score. If there are unobserved characteristics that influence the use of user generated content and social shopping tools, a hidden bias might exist (omitted variable bias), which would suggest that the found differences in revenue and share of returns would not represent causal effects. Thereby, it is likely that there are additional characteristics that drive customers' use of user generated content and especially social shopping tools which we were not able to consider in this empirical study (Mithas and Krishnan 2009). The question is how severe the omitted variable bias is. To assess the potential bias, we conduct a sensitivity analysis as suggested by Rosenbaum (1987). We calculate the Rosenbaum bounds using Wilcoxon sign-rank tests for the estimated effects. Table 4-7 shows the critical odds ratios at which the conclusions would alter. There are two potential changes: significant results might become insignificant at a certain odd ratio, and insignificant results can also become significant. The critical odd ratio should be as high as possible. Other studies using the same sensitivity analysis find critical odds ratios between 1.1 and 2.5 (DiPrete and Gangl 2004; Mithas and Krishnan 2009; Mithas et al. 2005). In our study the critical odds ratios for the use of user generated content vary between 1.0 and 2.1. Therefore the results are in line with previous research. The effects of active versus no-use of user generated content are the most robust effects, while the effects of passive versus no-use of user generated content are most susceptible to hidden bias. This means that there might be other variables, which explain, why consumers read user generated content. However, these are only worst case scenarios.

The effects with respect to social shopping tools are very sensitive. Considering that we only compare relatively small samples and that we do not find any significant effects, this result indicates that the use of social shopping tools potentially has an effect on customer revenue and return behavior.

Table 4-7: Critical odds ratio assess the sensitivity of the effects of user generated content and social shopping tools

		Critical odds ratio		
		ln(gross revenue)	ln(net revenue)	share of returns
user generated content	passive vs. non-users	1.3^{-}	1.2^{-}	1.1^{-}
	active vs. non-users	2.1^{-}	1.9^{-}	1.3^{-}
	active vs. passive users	1.5^{-}	1.5^{-}	1.1^{+}
social shopping tools	passive vs. non-users	1.0^{+}	1.2^{+}	1.7^{+}
	active vs. non-users	1.4^{+}	1.5^{+}	1.6^{+}

- significant effect becoming insignificant;
+ insignificant effect becoming significant

4.6 Discussion

4.6.1 Summary and managerial implications

Many retailers (i.e., intermediaries and manufacturers) provide innovative technologies in their online shops to facilitate the shopping process for the customers and enhance their shopping experience to finally affect firm performance. We analyze two kinds of innovative technologies: user generated content and social shopping tools. Our analyses show that implementing such tools can pay off for a retailer. Passive users of user generated content provide significantly higher revenue than non-users. This might be because user generated content provides additional information to customers that complements retailers' information and focuses on actual product experiences. Moreover, active users produce higher revenues than passive and non-users of user generated content. As such, compared to passive users, shoppers that also contribute user generated content (i.e. active users) seem more connected to the retailer by feeling being part of a community, which intensifies the relationship and results in higher revenues. In contrast to our expectations, the return rates are also higher for passive and active users of user generated content compared to non-users, whereas the return rates do not differ between active and passive users. But ultimately, passive and active users of user generated content produce higher net revenues for the retailer. As such, our results show that reading or even providing user generated content is done by "better" customers in that they simply buy more, with the result that they also return more. This effect seem to outweigh the effect that active or passive users of user generated content might be more knowledgeable with respect to the actual product performance, which should lead to lower disconfirmation between pre-purchase expectations and post-purchase experiences, higher satisfaction, and finally lower return rates. The use of social shopping tools seems to not affect customers' revenues on average in our study.

Overall, our results suggest that innovative online shopping technologies help to motivate purchase decisions. In any case, managers should stimulate the active use of user generated content since it provides value for other customers (passive users) and value to the retailer. For example, our data provider was thinking about introducing a loyalty program mainly based on contributing user generated content opposed to purchase behavior. Since user generated content also motivates passive users of this content to shop more, retailers should make this content easily available and also stimulate shoppers to read this content when browsing on their website. Stimulating user generated content comes with the cost of higher product returns, which is not problematic since net revenue is still higher than net revenue of non-users. While our study's insights are of interest to pure online retailers as well as manufacturers that use their own online retail stores, it is also of interest to multichannel retailers as well as manufacturers that have online as well as offline shops. The reason is that firms' multichannel strategy should be integrated and cannot be considered in isolation. Multichannel firms that receive customers' feedback via user generated content in their online store can use this content also in their offline store and vice versa. For example, Burberry offers technologies such as tablet computers in their offline store to inform their offline shoppers about Burberry's products and the content by other users (Lindner 2016). As such, they stimulate the feeling of being part of a community, which leads to higher commitment with the firm and finally more purchases. Now, Burberry should also stimulate those offline shoppers to write reviews on their website. Multichannel firms should also integrate their return process (i.e., products bought online might be returned offline), As such, they reduce customers' return costs, have more shoppers in their offline shops, and have also the opportunity to personally interact with their customers and discuss potential problems with the products that have led to the return of the product. All of the above can potentially lead to an increased relationship with the customers and finally more purchases.

4.6.2 Limitations and future research

This study faces some general limitations which provide avenues for future research. First, the number of users of social shopping tools is rather small what might be a reason for the insignificant findings. Social shopping tools are innovative features and therefore, very few customers actually use them. When an increasing number of customers use social shopping tools, this study should be replicated to confirm or object this study's claim that the use of social shopping tools does not affect customer revenue on average. Second, the design of user generated content and social shopping tools cannot be studied based on the available data. Therefore we are not able to investigate the effects of different designs of these tools. Future research may investigate whether specific designs of user generated content and social shopping tools are more effective in stimulating purchases than others. Additionally, we cannot measure content available in different channels. Future studies should investigate effects of user generated content generated and used in different channels. Third, we find in our study that engaged customers are very profitable for a company. However, we do not know what affects customer engagement. It is beyond the scope of this study to develop strategies to do so, but future research should examine how firms' can engage customers. Fourth, user generated content reflects the customers' opinion. Future research should investigate how this information can be considered in supportive online purchase environments (Xiao and Benbasat 2007).

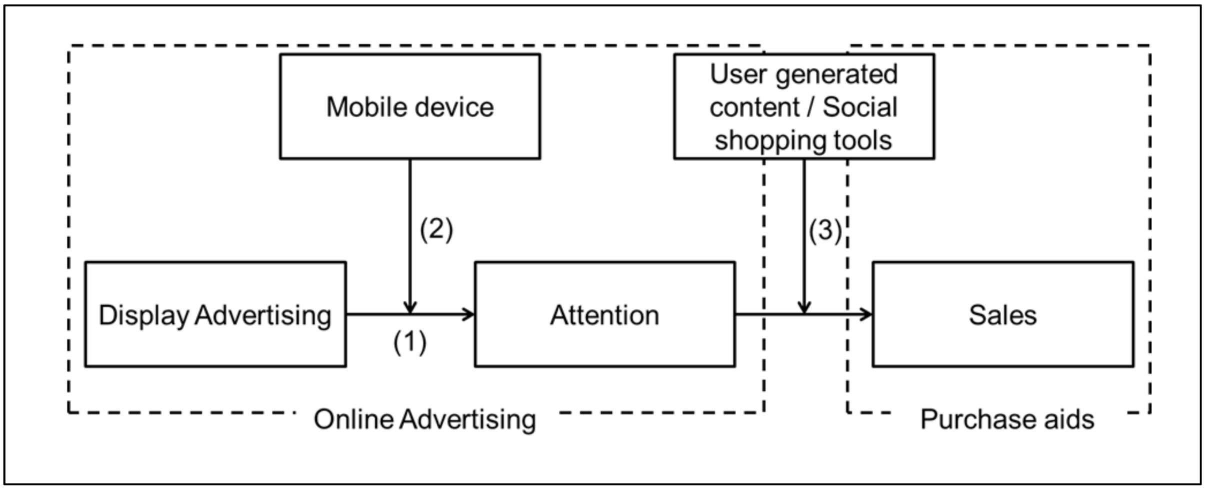
Despite these limitations, this study contributes to the existing literature by shedding light on the effects of customers' active and passive use of user generated content and social shopping tools on customer revenue and return behavior.

5 General discussion

Today, consumers shop a lot online (eMarketer 2016). The online channel faces several challenges for marketers. First, the online channel is cluttered. Therefore, it is difficult to stand out and generate attention for the product. A common way to generate attention is display advertising (IAB 2015). Display ads are served in the traditional online channel (using PCs and laptops) and the mobile channel (using smartphones and tablets). These channels differ in usage situation and consumer behavior and therefore advertisers have to be aware of the differences when serving display ads. When marketers succeeded and consumers visit the webpage, marketers face the problem that consumers rely on other consumers' opinions (Babić Rosario et al. 2016; Floyd et al. 2014).

This dissertation aimed to tackle these challenges using three different studies. The first study investigated whether display ads are more effective when they match the consumer's interest in the product. The second study answers the question if consumers click mobile ads differently compared to traditional display ads. Last, but least, the third study dealt with whether consumers spend more money when using purchase aids such as user generated content or social shopping tools. Figure 5-1 shows again the framework of this dissertation.

Figure 5-1: Outline of dissertation



We shortly summarize the three studies and end this chapter by giving an outlook to the future.

5.1 Study 1: Effect of ad characteristics and targeting options on display ad effectiveness taking the unobservable interest into account

Advertisers face the problem that they do not know how interested the consumer is in the advertised product. A good display ad should match the communication goal of the advertiser but also match the interest of the consumer. We propose that based on the unobservable interest, different ad characteristics should be served and different targeting options should be used. We define the ad characteristics in terms of ad message and ad format. To estimate the unobservable interest, we develop a Hidden Markov Model (Netzer et al. 2008). We are interested in the transitions between certain interest states and how ad characteristics and targeting options influence these transitions. That is, managers can identify ad characteristics and targeting options that positively influence the interest state and therefore increase the probability of a reaction of the consumer. On the other hand, they can also identify ad characteristics and targeting options that decrease the interest and therefore decrease the probability of a reaction.

After building up the model, we conduct a simulation study to test how well the model can identify the parameters. We prove that our model fits the data better than simpler approaches like the traditional logit model or the random intercept logit model. Afterwards we run several simulations based on the estimated parameters and show that we can increase the chance of a reaction of the consumer.

Our dataset consists of clickstream data from an American travel and tourism company. In total, we have information about approximately 80 million cookies with over 300 million impressions. We draw a sample out the dataset to make the estimation feasible.

We find that ad characteristics and targeting options indeed influence the interest in the product. Based on the interest state, the targeting option plays a huge role in consumers' behavior. The ad format also has different effects on reaction based on the latent interest state. Interestingly the message only has an effect when the consumer has a low interest in the product, but not when he or she has a high interest. We also control of the number of served impressions and previous clicks, as well as for other marketing exposures like search or email advertising. We find that more impressions have a negative impact when the consumer is not interested, but when the interest is present; an additional exposure increases the chance to react. As expected, consumers that reacted before have a higher probability to be interested in the product and react again.

Our model is the base for several simulations. Advertisers can simulate, whether it is reasonable to serve an additional ad to a consumer or not. The model enables advertisers to plan an individual frequency cap for each consumer. We help advertisers to show the right display ad to the right consumer at the right time in the right format.

Summing up, we built a new model to explicate the latent interest in an advertised product based on clickstream data. These kinds of data are the only data available for advertisers. Therefore, an understanding of effects within an ad campaign can be crucial for campaign success.

5.2 Study 2: User device and ad response: The moderating role of ad position

Mobile marketing is on the rise and consumers visit the internet more often with mobile devices. Research in the mobile channel is sparse. But researchers agree that the mobile channel is conceptually different from the traditional online channel (Ghose, Goldfarb, et al. 2013). But still, advertisers treat these channels equally. We want to show that the behavioral outcome of consumers is different regarding the channel they are using while accessing the internet. Furthermore, we want to study how the reaction to display advertising differs between the two channels. Therefore, we conduct a field experiment on a German nutrition website. We control for the position of the display ad. We place one ad at the top of the website and another one further down. What is beyond our control is the decision of the consumer with which device he or she enters the website. Thus, we first try to explain the mobile usage with the time of the day on which the consumer visits the website and whether it is weekend or not. Then we built up a traditional probit model to explain the reaction on a display ad (i.e. whether the consumer clicks on the display ad or not). We find that in general consumers click less on a display ad when using a mobile device. However, a display ad further down the website increases the probability to be clicked up to a certain amount of impression for the mobile channel compared to the traditional channel. There is no significant difference between ads at the top position between these two channels.

We also test the prediction validity of the probit model using a holdout sample which consists of 20% of the sample. All validation measures indicate a good prediction accuracy of the proposed model.

Therefore, we help advertisers to understand, that consumers process display advertising differently when using a mobile device compared to a stationary device. Since ads at the bottom are clicked more often, advertisers should buy these inventories when the consumer visits a website with a mobile device.

5.3 Study 3: Value of user generated content and social shopping tools

Many retailers and manufacturers use the Internet as a distribution channel above and beyond their offline retail stores. Using the online channel as a distribution channel faces retailers and manufacturers with two challenges: Consumers using the online channel cannot evaluate the products beforehand and, moreover, shopping lacks the social experience of offline stores. Retailers try to overcome these limitations by offering user generated content and social shopping tools in their online shops. This chapter aims to examine the effect of these two tools on customer revenue, and customers' return behavior. We differentiate between customers' active, passive and no use of these tools and also control for self-selection effects using propensity score matching. Controlling for self-selection effects is important because active users might have different attitudes compared to passive users or non-users.

We find that customers who use user generated content produce higher gross revenues for the retailer. Surprisingly, these customers also have higher return rates. However, they still generate significantly higher net revenues for the retailer. Moreover, the effects are smaller when controlling for self-selection effects. We also find that social shopping tools do not affect the customer's revenue on average. In general, user generated content indeed has a positive effect on purchase decision, but a fraction of this effect can be explained through covariates that determine the self-selection effect.

Therefore, retailers should stimulate the use of user generated content because active users seem to be more loyal to the company and buy more. Furthermore, they provide important information for passive consumers that they can use for their purchase decision. Even though social shopping tools are not significant, retailers should still be careful on not using these tools. One reason for the insignificance might be the small number of users.

5.4 Outlook to the future

In this dissertation we shed light on how marketers can convince online consumers to purchase. We start with a study that shows that serving display ads that match a consumer's interest can increase the positive behavioral outcome of consumers. Advertisers should take a step back on showing too many ads to consumers, but start thinking what consumers want to see. For example, consider the targeting strategy of retargeting. Our study shows that it decreases the probability of a positive behavioral outcome. This may be surprising, but with a deeper look it at, this results makes sense. Retargeting mostly works automatically, where agencies track a visit at an online shop via cookies. When they realize, that a consumer looked at a product, but did not buy it, they serve ads to that certain consumer displaying either the product itself or other products from the online shop. This can interfere with the consumer's privacy concerns. Just transfer this situation to an offline example. When a consumer visits a supermarket and considers several products, but does not buy them, no advertiser will follow this consumer and serve him or her ads from these products. But this is exactly how retargeting works. This form of advertising can work, but only when consumers are close to purchase and showed interest on third party platforms (Lambrecht and Tucker 2013).

This example transfers to other forms of advertising as well. This leads to an additional problem for marketers: The increasing installation of ad blockers. Since advertisers become more and more aggressive, consumers protect themselves by blocking intrusive ads. This can cause serious harm to both, advertisers and publishers, since advertisers cannot reach the desired target audience and publishers earn their money with serving ads on their websites. One approach is to convince the consumer to deactivate the ad blocker for certain websites. Using an appropriate message increases the probability that consumers are willing to allow ads on certain websites (Schumann et al. 2014).

To sum it up, the advertising industry is still a huge and relevant business, but the environment changes fast. Therefore, the industry has to think about how to serve ads to the consumer. In our opinion, the industry practice should take a step back from annoying the consumer with irrelevant ads and move a step toward the serving of ads that fits the need of the consumer. Our model shown in chapter 2 is a first move toward that direction. We believe that an advertiser that does not intrusively serves irrelevant ads to a consumer will benefit in a long run.

The multichannel complexity is also a challenge that needs to be addressed. So far, there is no differentiation on how display ads are served between the traditional online channel and the mobile channel (Del Rey 2012). Marketers need to be aware of these differences and design different strategies to reach the consumer.

Furthermore, the attribution problem between these channels is not solved yet. For most advertisers, it is almost impossible to differentiate mobile users from online users. So far, consumers are tracked using cookies. But since they are always stores locally, it is difficult to connect a consumer that uses a traditional computer with the equivalent mobile device. It is even more difficult to connect all the offline activities to that certain consumer. This is a future challenge for marketers since they influence each other (Ghose, Han, et al. 2013; Guitart and Hervet 2016).

Last, but not least, product managers should be as transparent as possible. Consumers rely on other consumers' opinions and this influence their decision to purchase. Online shops should give consumers the opportunity to connect. This can be done by giving a platform for user generated content or social shopping tools. Especially user generated content can give hints of negative properties of a product or service. This information can be used to create a better experience for consumers. Positive information can be used in advertising as a quality sign.

To summarize this dissertation, the three key learnings are:

1. Advertisers should serve display ads that are relevant to the consumer,
2. Advertisers should consider the differences between the mobile and the traditional online channel,
3. Retailers should give consumers the opportunity to interact with each other on the website.

Appendix A: Simulation study

We conduct several simulations to test whether the model can be identified and estimate all the parameters. First, we generate the parameters of the transition matrix. We specify the state specific intercept parameters $\tau_{iss'}$ and blur them with a normal distribution with mean zero and true covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{\boldsymbol{true}}$. This ensures that every individual in the simulation has a different parameter which captures heterogeneity in the model. Then we set the true values of the ad characteristics $\boldsymbol{\rho_s}$, the targeting options $\boldsymbol{\upsilon_s}$ and generate these variables with a Bernoulli distribution.

Based on the generated parameters, we have the individual transition matrix for each individual *i* and can simulate a path of different states. Based on the state, we can simulate a possible outcome for individual at time *t* in state *s* based on the conditional choice parameters.

In detail we simulate a model with two and three hidden states. The mean transition matrices and state dependent emission probabilities are shown in Table A-1.

Table A-1: Design for simulation study

#states	Covariance matrix	Initial distribution	Transition matrix	State dependent choice
2	$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{\boldsymbol{true}} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}$	$\boldsymbol{\pi} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$\boldsymbol{Q} = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$	$\boldsymbol{m} = \begin{bmatrix} 0.1 \\ 0.9 \end{bmatrix}$
3	$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{\boldsymbol{true}} = \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.2 \end{bmatrix}$	$\boldsymbol{\pi} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$	$\boldsymbol{Q} = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{bmatrix}$	$\boldsymbol{m} = \begin{bmatrix} 0.05 \\ 0.45 \\ 0.90 \end{bmatrix}$

The model is able to identify all the model parameters successfully. Table A-2 shows the results for the parameter values for the two and three states model. No parameter is significantly different at the 0.05% level, meaning that every parameter could be recovered. Convergence was assesses through visual inspection. The trace plots and histograms of the fixed effects parameters are shown in Figure A-1 and Figure A-2.

The adaptive algorithm of Atchadé (2006) takes care of the autocorrelation problem that inherits Random Walk Metropolis Hastings algorithms. Together with skipping 10 observations from the posterior distribution, there is no autocorrelation for the two states model and only little autocorrelation for the three states model. Figure A-3 and Figure A-4 show the autocorrelation functions for the fixed effects for the two models.

Using the simulated data also allows us to check for the model's ability to recover the latent state of the consumer. Using the filtering approach from equation (13) shows that the model is able to recover the latent state most of the time. Figure A-5 and Figure A-6 depict the simulated and estimated state of one sim-ulated data chain. In this example, 96.7% of the states were identified right in the two states model and 83.3% of the states were identified right in the three states model.

Table A-2: Simulation results

Parameter	2 state model (7 parameters)		Three state model (30 parameters)	
	True value	Parameter estimate (2.5%, 97.5%)	True value	Parameter estimate (2.5%, 97.5%)
τ_{11}	2.2	2.08 (1.16, 2.99)	2.2	2.40 (1.46, 3.33)
τ_{12}	-	-	-0.3	-0.13 (-1.03, 0.77)
τ_{21}	-2.2	-2.15 (-3.12, -1.18)	-2.9	-2.51 (-3.79, -1.23)
τ_{22}	-	-	1.8	1.72 (0.87, 2.58)
τ_{31}	-	-	-2.9	-2.41 (-3.18, -1.64)
τ_{32}	-	-	-0.4	-0.88 (-2.11, 0.32)
$\tilde{\beta}_{01}$	-2.2	-2.26 (-2.39, -2.14)	-2.9	-2.95 (-3.22, -2.71)
$\tilde{\beta}_{02}$	1.5	1.52 (1.48, 1.57)	1.0	1.08 (0.96, 1.18)
$\tilde{\beta}_{03}$	-	-	0.9	0.86 (0.71, 0.99)
σ_{11}	0.2	0.21 (0.10, 0.41)	0.2	0.22 (0.06, 0.55)
σ_{12}	0.0	-0.02 (-0.18, 0.10)	0.0	-0.02 (-0.23, 0.09)
σ_{13}	-	-	0.0	-0.15 (-0.74, 0.02)
σ_{14}	-	-	0.0	0.03 (-0.07, 0.20)
σ_{15}	-	-	0.0	-0.04 (-0.23, 0.06)
σ_{16}	-	-	0.0	-0.10 (-0.57, 0.07)
σ_{22}	0.2	0.24 (0.10, 0.45)	0.2	0.21 (0.08, 0.44)
σ_{23}	-	-	0.0	0.00 (-0.22, 0.41)
σ_{24}	-	-	0.0	0.00 (-0.11, 0.10)
σ_{25}	-	-	0.0	-0.01 (-0.16, 0.13)
σ_{26}	-	-	0.0	0.05 (-0.16, 0.29)
σ_{33}	-	-	0.2	0.41 (0.10, 1.40)
σ_{34}	-	-	0.0	-0.02 (-0.28, 0.11)
σ_{35}	-	-	0.0	0.04 (-0.12, 0.36)
σ_{36}	-	-	0.0	0.11 (-0.16, 0.97)
σ_{44}	-	-	0.2	0.19 (0.07, 0.37)
σ_{45}	-	-	0.0	-0.06 (-0.19, 0.03)
σ_{46}	-	-	0.0	-0.03 (-0.26, 0.11)
σ_{55}	-	-	0.2	0.15 (0.07, 0.29)
σ_{56}	-	-	0.0	0.05 (-0.07, 0.26)
σ_{66}	-	-	0.2	0.35 (0.09, 0.88)

Figure A-1: Trace plot and histogram of the fixed effect parameters (2 states)

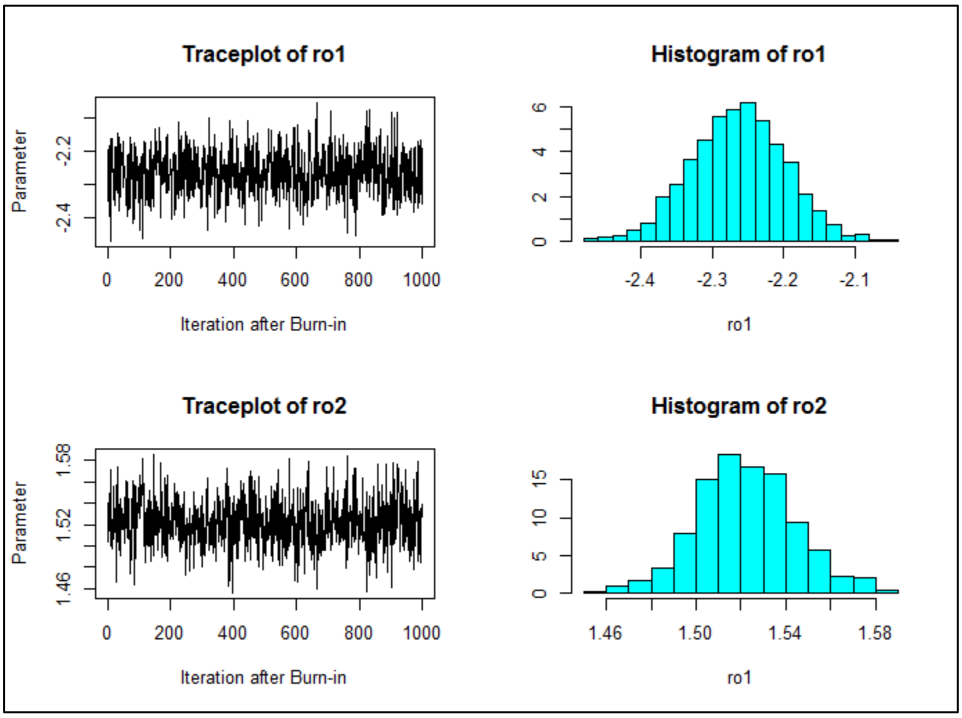


Figure A-2: Trace plot and histogram of the fixed effect parameters (3 states)

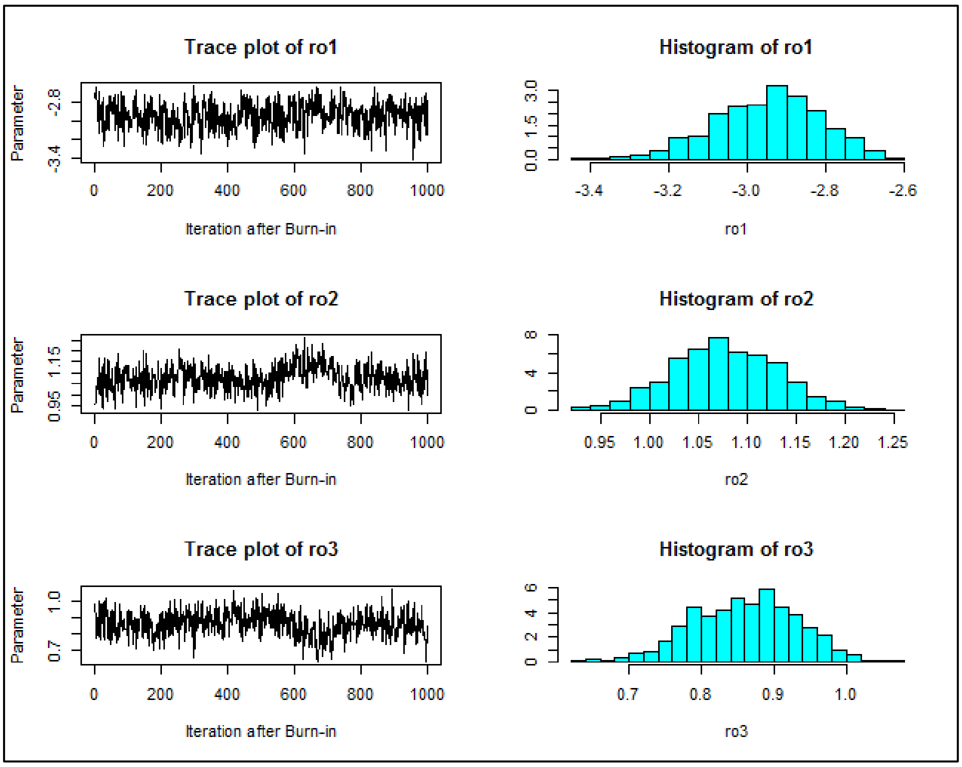


Figure A-3: Autocorrelation-function for the fixed parameters of the two states model

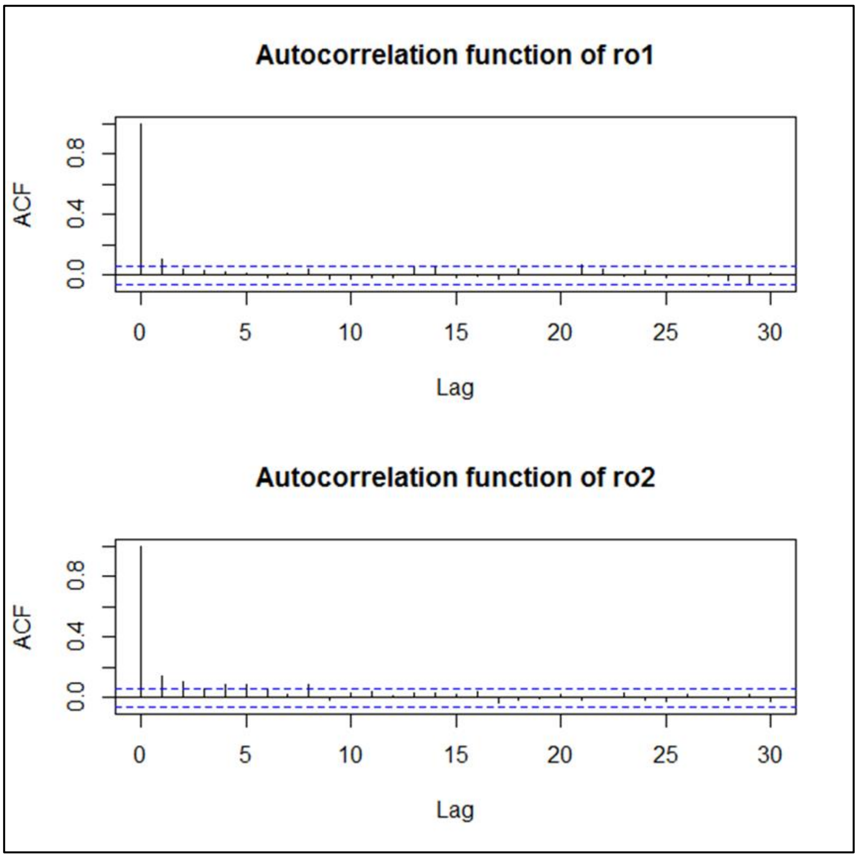


Figure A-4: Autocorrelation-function for the fixed parameters of the three states model

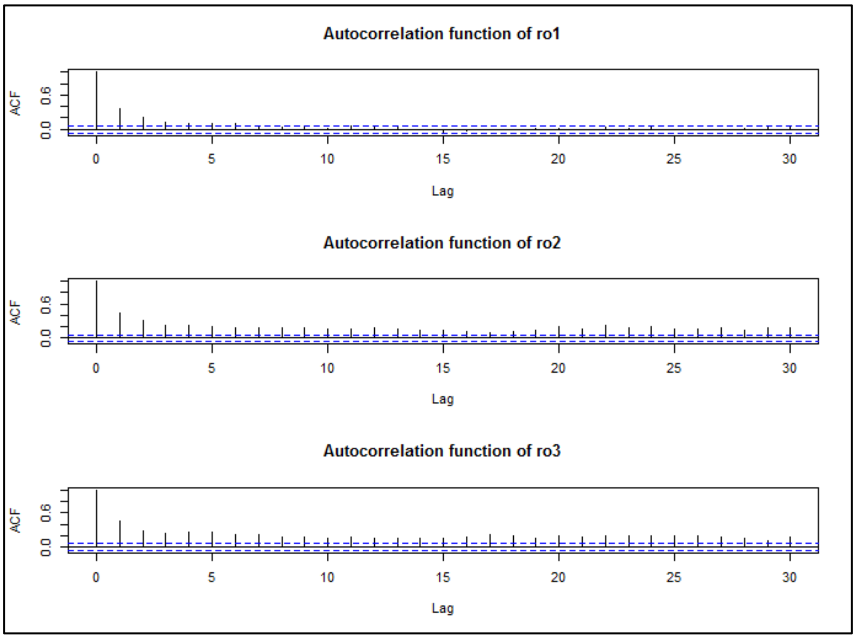


Figure A-5: Simulated vs. estimated state membership (2 states)

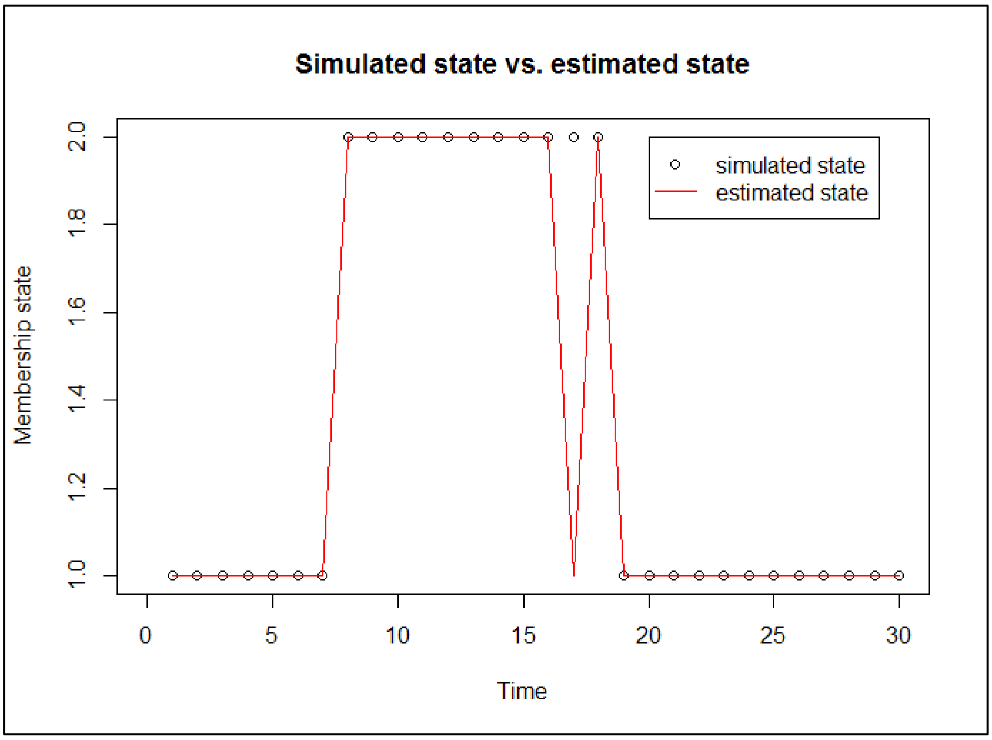
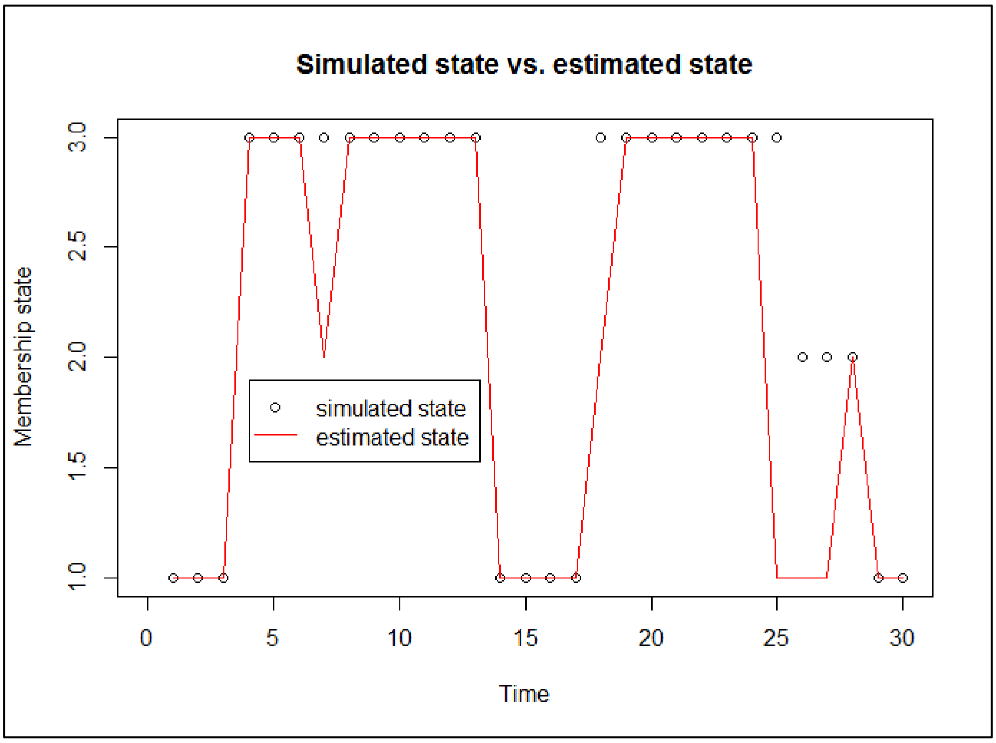


Figure A-6: Simulated vs. estimated state membership (3 states)



Appendix B: Factor analysis

To ensure the suitability of the data for a factor analysis, we test the data using the Kaiser-Meyer-Olkin criterion of sampling adequacy (Kaiser and Rice 1974) and Barlett's test of sphericity (Bartlett 1937). The measures of sampling adequacy are conducted for each variable to test whether a variable should be included in the factor analysis and for the whole dataset to test the suitability of the whole dataset. Both measures should exceed the threshold of 0.5.

To get a better understanding of the structure of the data, Table B-1 shows the correlation matrix of the items. A first look suggests that items that belong together indeed have a high correlation and items that do not belong together have a rather low correlation.

Table B-2 shows the anti-image correlation matrix with the measures of sampling adequacy on the diagonal. All values on the diagonal are above 0.5 indicating that all variables should be included into the analysis. The overall suitability is KMO = 0.850 indicating a meritorious fit (Kaiser and Rice 1974, p. 112).

Bartlett's test of sphericity tests whether the data matrix is the identity matrix. If this is the case, the variables are unrelated and the factor analysis should not be conducted. The test statistic is $\chi^2 = 16184.394$ (p-value < 0.01), indicating that the data are suited for a factor analysis.

After conducting a factor analysis, we have to determine the number of final factors. Since we want to validate three constructs, we aim for three factors. To objectively determine the number of factors, we use the latent root criterion (Guttman 1954; Kaiser and Dickmann 1959) and the visual scree plot (Cattell 1966). Table B-3 shows the eigenvalues and explained variances per number of factors. As the latent root criterion extracts a factor for every eigenvalue greater than one, three factors are extracted. This accounts for approximately 57% of the variance, which is a sufficient value. Figure B-1 shows the scree plot which plots the eigenvalues and the total number of factors. If there is a notable "elbow", the number of factors left from the elbow is extracted. Again, this criterion extracts three factors. To summarize, we aim for three factors and both criteria extract three factors.

The next step for validation of the constructs is to check whether the items load on the desired factors. Therefore, we calculate the factor loadings of each item on the factor. For a clear interpretation, we also calculate the rotated factor loading using the varimax rotation (Kaiser 1958). Table B-4 shows the unrotated and rotated factor solution. From the rotated factor solution, we can see that every item loads on the desired factor.

Table B-1: Correlation matrix of items

	Internet skills 1	Internet skills 2	Internet skills 3	Internet skills 4	Internet skills 5	Internet skills 6	Utilitarian attitude 1	Utilitarian attitude 2	Utilitarian attitude 3	Utilitarian attitude 4	Utilitarian attitude 5	Hedonic attitude 1	Hedonic attitude 2	Hedonic attitude 3	Hedonic attitude 4	Hedonic attitude 5
Internet skills 1	1.000	0.745	0.372	0.425	0.532	0.404	-0.208	-0.094	-0.194	-0.126	-0.132	-0.079	-0.080	-0.145	-0.052	-0.143
Internet skills 2	0.745	1.000	0.357	0.581	0.538	0.423	-0.278	-0.110	-0.236	-0.164	-0.202	-0.103	-0.112	-0.164	-0.058	-0.184
Internet skills 3	0.372	0.357	1.000	0.193	0.395	0.344	-0.194	-0.084	-0.194	-0.089	-0.108	-0.065	-0.017	-0.063	0.001	-0.092
Internet skills 4	0.425	0.581	0.193	1.000	0.334	0.265	-0.181	-0.081	-0.162	-0.107	-0.180	-0.103	-0.076	-0.130	-0.064	-0.149
Internet skills 5	0.532	0.538	0.395	0.334	1.000	0.634	-0.257	-0.106	-0.246	-0.142	-0.176	-0.147	-0.119	-0.177	-0.078	-0.193
Internet skills 6	0.404	0.423	0.344	0.265	0.634	1.000	-0.188	-0.088	-0.184	-0.138	-0.142	-0.141	-0.094	-0.135	-0.054	-0.160
Utilitarian attitude 1	-0.208	-0.278	-0.194	-0.181	-0.257	-0.188	1.000	0.369	0.682	0.276	0.573	0.285	0.185	0.260	0.104	0.336
Utilitarian attitude 2	-0.094	-0.110	-0.084	-0.081	-0.106	-0.088	0.369	1.000	0.411	0.209	0.310	0.279	0.236	0.270	0.276	0.316
Utilitarian attitude 3	-0.194	-0.236	-0.194	-0.162	-0.246	-0.184	0.682	0.411	1.000	0.335	0.604	0.308	0.198	0.286	0.169	0.353
Utilitarian attitude 4	-0.126	-0.164	-0.089	-0.107	-0.142	-0.138	0.276	0.209	0.335	1.000	0.311	0.121	0.146	0.153	0.163	0.197
Utilitarian attitude 5	-0.132	-0.202	-0.108	-0.180	-0.176	-0.142	0.573	0.310	0.604	0.311	1.000	0.331	0.169	0.290	0.097	0.381
Hedonic attitude 1	-0.079	-0.103	-0.065	-0.103	-0.147	-0.141	0.285	0.279	0.308	0.121	0.331	1.000	0.489	0.582	0.375	0.608
Hedonic attitude 2	-0.080	-0.112	-0.017	-0.076	-0.119	-0.094	0.185	0.236	0.198	0.146	0.169	0.489	1.000	0.587	0.552	0.503
Hedonic attitude 3	-0.145	-0.164	-0.063	-0.130	-0.177	-0.135	0.260	0.270	0.286	0.153	0.290	0.582	0.587	1.000	0.499	0.686
Hedonic attitude 4	-0.052	-0.058	0.001	-0.064	-0.078	-0.054	0.104	0.276	0.169	0.163	0.097	0.375	0.552	0.499	1.000	0.437
Hedonic attitude 5	-0.143	-0.184	-0.092	-0.149	-0.193	-0.160	0.336	0.316	0.353	0.197	0.381	0.608	0.503	0.686	0.437	1.000

Table B-2: Anti-image correlation matrix

	Internet skills 1	Internet skills 2	Internet skills 3	Internet skills 4	Internet skills 5	Internet skills 6	Utilitarian attitude 1	Utilitarian attitude 2	Utilitarian attitude 3	Utilitarian attitude 4	Utilitarian attitude 5	Hedonic attitude 1	Hedonic attitude 2	Hedonic attitude 3	Hedonic attitude 4	Hedonic attitude 5
Internet skills 1	0.799	-0.575	-0.115	0.015	-0.166	-0.018	-0.026	0.015	0.022	-0.001	-0.039	-0.018	-0.025	0.043	0.006	-0.013
Internet skills 2	-0.575	0.770	-0.059	-0.409	-0.122	-0.055	0.087	-0.015	-0.018	0.039	0.019	-0.046	0.046	-0.001	-0.030	0.027
Internet skills 3	-0.115	-0.059	0.925	0.025	-0.134	-0.109	0.037	0.010	0.064	-0.004	-0.040	0.008	-0.033	-0.016	-0.023	0.010
Internet skills 4	0.015	-0.409	0.025	0.837	-0.023	-0.013	-0.015	-0.004	-0.012	-0.008	0.067	0.019	-0.029	0.005	0.028	0.010
Internet skills 5	-0.166	-0.122	-0.134	-0.023	0.836	-0.486	0.045	-0.027	0.044	-0.006	-0.016	0.003	0.008	0.024	0.008	0.012
Internet skills 6	-0.018	-0.055	-0.109	-0.013	-0.486	0.814	-0.015	0.002	-0.007	0.048	0.003	0.051	0.001	-0.013	-0.017	0.013
Utilitarian attitude 1	-0.026	0.087	0.037	-0.015	0.045	-0.015	0.848	-0.113	-0.446	-0.015	-0.233	-0.017	-0.034	0.003	0.079	-0.041
Utilitarian attitude 2	0.015	-0.015	0.010	-0.004	-0.027	0.002	-0.113	0.925	-0.166	-0.046	-0.023	-0.047	-0.010	0.007	-0.146	-0.055
Utilitarian attitude 3	0.022	-0.018	0.064	-0.012	0.044	-0.007	-0.446	-0.166	0.835	-0.126	-0.298	-0.032	0.021	-0.012	-0.044	-0.010
Utilitarian attitude 4	-0.001	0.039	-0.004	-0.008	-0.006	0.048	-0.015	-0.046	-0.126	0.910	-0.132	0.067	-0.032	0.024	-0.089	-0.029
Utilitarian attitude 5	-0.039	0.019	-0.040	0.067	-0.016	0.003	-0.233	-0.023	-0.298	-0.132	0.873	-0.095	0.046	-0.030	0.102	-0.121
Hedonic attitude 1	-0.018	-0.046	0.008	0.019	0.003	0.051	-0.017	-0.047	-0.032	0.067	-0.095	0.904	-0.167	-0.191	-0.020	-0.277
Hedonic attitude 2	-0.025	0.046	-0.033	-0.029	0.008	0.001	-0.034	-0.010	0.021	-0.032	0.046	-0.167	0.863	-0.254	-0.333	-0.063
Hedonic attitude 3	0.043	-0.001	-0.016	0.005	0.024	-0.013	0.003	0.007	-0.012	0.024	-0.030	-0.191	-0.254	0.863	-0.165	-0.400
Hedonic attitude 4	0.006	-0.030	-0.023	0.028	0.008	-0.017	0.079	-0.146	-0.044	-0.089	0.102	-0.020	-0.333	-0.165	0.843	-0.093
Hedonic attitude 5	-0.013	0.027	0.010	0.010	0.012	0.013	-0.041	-0.055	-0.010	-0.029	-0.121	-0.277	-0.063	-0.400	-0.093	0.878

Table B-3: Eigenvalues and explained variance of extracted factors

	Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	4.841	30.256	30.256
2	2.655	16.592	46.848
3	1.665	10.408	57.256
4	0.946	5.910	63.166
5	0.875	5.469	68.635

Figure B-1: Scree plot of extracted factors

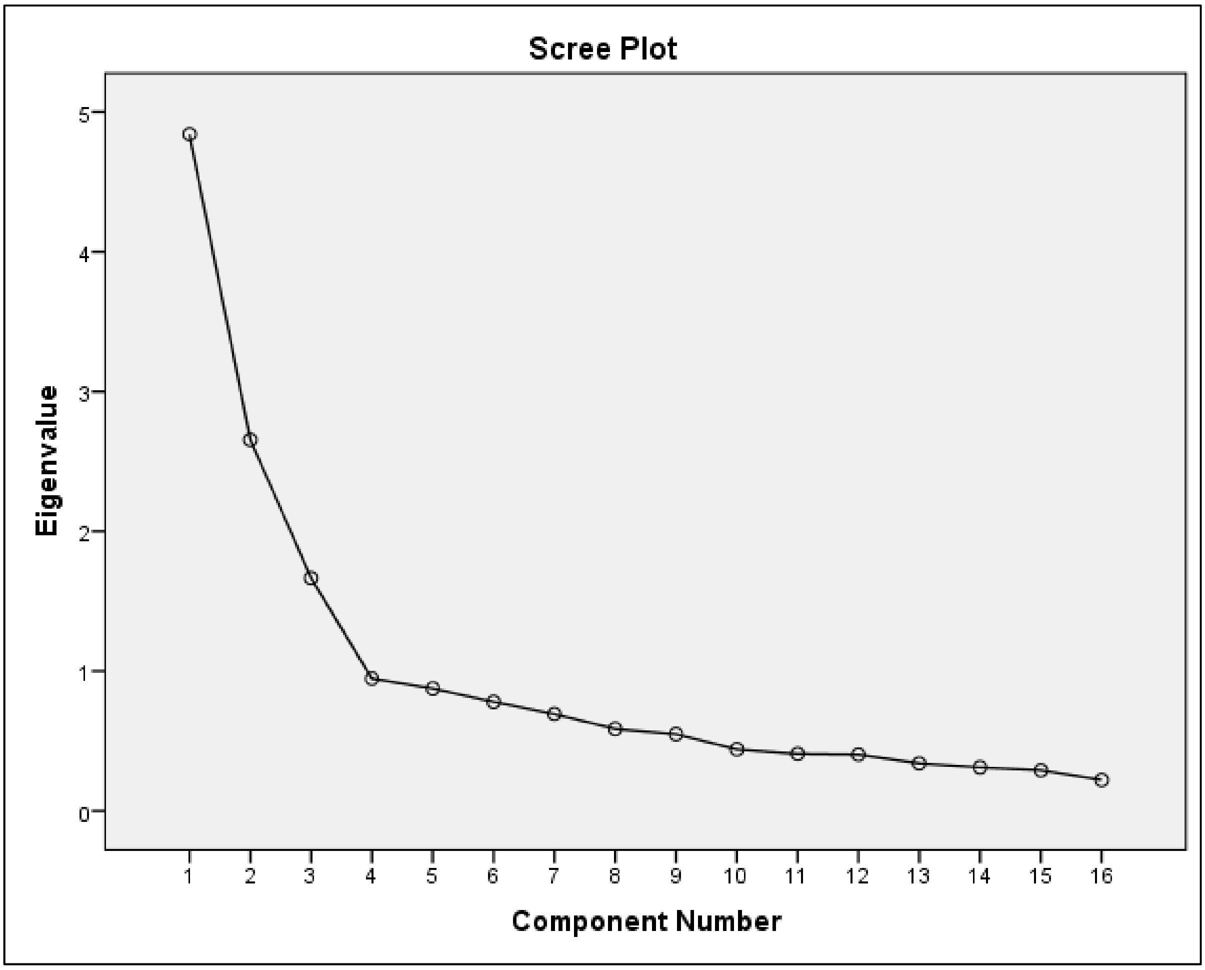


Table B-4: Unrotated and rotated factor solution

	Unrotated factor solution			Rotated factor solution		
	Factor 1 (internet skills)	Factor 2 (hedonic attitude)	Factor 3 (utilitarian attitude)	Factor 1 (internet skills)	Factor 2 (hedonic (attitude)	Factor 3 (utilitarian attitude)
Internet skills 1	-0.519	0.603	0.198	0.816	-0.047	-0.061
Internet skills 2	-0.581	0.603	0.158	0.840	-0.062	-0.129
Internet skills 3	-0.369	0.448	0.028	0.565	0.039	-0.131
Internet skills 4	-0.436	0.429	0.135	0.617	-0.069	-0.087
Internet skills 5	-0.563	0.533	0.170	0.778	-0.099	-0.119
Internet skills 6	-0.479	0.478	0.171	0.688	-0.084	-0.077
Utilitarian attitude 1	0.640	0.025	0.541	-0.191	0.098	0.810
Utilitarian attitude 2	0.487	0.236	0.264	-0.015	0.286	0.530
Utilitarian attitude 3	0.661	0.080	0.551	-0.157	0.138	0.839
Utilitarian attitude 4	0.399	0.035	0.313	-0.111	0.086	0.489
Utilitarian attitude 5	0.599	0.122	0.529	-0.096	0.139	0.790
Hedonic attitude 1	0.604	0.428	-0.188	-0.053	0.714	0.266
Hedonic attitude 2	0.536	0.457	-0.395	-0.048	0.803	0.064
Hedonic attitude 3	0.652	0.438	-0.325	-0.108	0.822	0.186
Hedonic attitude 4	0.454	0.451	-0.387	-0.005	0.748	0.022
Hedonic attitude 5	0.691	0.405	-0.183	-0.118	0.749	0.317

References

Abhishek, Vibhanshu, Peter S. Fader, and Kartik Hosanagar (2012), “Media Exposure through the Funnel: A Model of Multi-Stage Attribution,” *SSRN Electronic Journal*.

Ai, Chunrong, and Edward C. Norton (2003), “Interaction terms in logit and probit models,” *Economics Letters*, 80(1), 123–29.

Ailawadi, Kusum L., and Paul W. Farris (2017), “Managing Multi- and Omni-Channel Distribution: Metrics and Research Directions,” *Journal of Retailing*, (forthcoming).

Anderson, E. W. (1998), “Customer Satisfaction and Word of Mouth,” *Journal of Service Research*, 1(1), 5–17.

Andrews, Michelle, Xueming Luo, Zheng Fang, and Anindya Ghose (2016), “Mobile Ad Effectiveness: Hyper-Contextual Targeting with Crowdedness,” *Marketing Science*, 35(2), 218–33.

Arnold, Mark J., and Kristy E. Reynolds (2003), “Hedonic shopping motivations,” *Journal of Retailing*, 79(2), 77–95.

Atchadé, Yves F. (2006), “An adaptive version for the metropolis adjusted Langevin algorithm with a truncated drift,” *Methodology and Computing in Applied Probability*, 8(2), 235–54.

Atchadé, Yves F., and Jeffrey S Rosenthal (2005), “On adaptive Markov chain Monte Carlo algorithms,” *Bernoulli*, 11(5), 815–28.

Babić Rosario, Ana, Francesca Sotgiu, Kristine De Valck, and Tammo H.A. Bijmolt (2016), “The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors,” *Journal of Marketing Research*, 53(3), 297–318.

Barlow, Alexis K.J., Noreen Q. Siddiqui, and Mike Mannion (2004), “Developments in information and communication technologies for retail marketing channels,” *International Journal of Retail & Distribution Management*, 32(3), 157–63.

Bart, Yakov, Andrew T. Stephen, and Miklos Sarvary (2014), “Which Products Are Best Suited to Mobile Advertising? A Field Study of Mobile Display Advertising Effects on Consumer Attitudes and Intentions,” *Journal of Marketing Research*, 51(3), 270–85.

Bartlett, M. S. (1937), “Properties of Sufficiency and Statistical Tests,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 160(901), 268–82.

Becker, Sascha O, and Andrea Ichino (2002), “Estimation of average treatment effects based on propensity scores,” *The Stata Journal*, 2(4), 358–77.

Berelson, Bernard (1949), “What ‘Missing the Newspaper’ Means,” in *Communication Research*, P. F. Lazarsfeld and F. N. Stanton, eds., New York: Harper, 111–29.

Bickart, Barbara, and RM Schindler (2001), “Internet forums as influential sources of consumer information,” *Journal of Interactive Marketing*, 15(3), 31–40.

Biocca, Frank, Chad Harms, and Jenn Gregg (2001), “The Networked Minds Measure of Social Presence: Pilot Test of the Factor Structure and Concurrent Validity Co-Presence,” in *Proceedings of the 4th Annual International Workshop on Presence*.

Bleier, Alexander, and Maik Eisenbeiss (2015), “Personalized Online Advertising Effectiveness: The Interplay of What, When, and Where,” *Marketing Science*, 34(5), 669–88.

Brasel, S. Adam, and James Gips (2014), “Tablets, touchscreens, and touchpads: How varying touch interfaces trigger psychological ownership and endowment,” *Journal of Consumer Psychology*, 24(2), 226–33.

Braun, Michael, and Wendy W. Moe (2013), “Online Display Advertising: Modeling the Effects of Multiple Creatives and Individual Impression Histories,” *Marketing Science*, 32(5), 753–67.

Bruce, Norris I., B.P.S. Murthi, and Ram C. Rao (2016), “A Dynamic Model for Digital Advertising: The Effects of Creative Formats, Message Content and Targeting on Engagement,” *Journal of Marketing Research*, (forthcoming).

Caliendo, Marco, Michel Clement, Dominik Papies, and Sabine Scheel-Kopeinig (2012), “Research Note —The Cost Impact of Spam Filters: Measuring the Effect of Information System Technologies in Organizations,” *Information Systems Research*, 23(3-NaN-2), 1068–80.

Caliendo, Marco, and Sabine Kopeinig (2008), “Some Practical Guidance for the Implementation of Propensity Score Matching,” *Journal of Economic Surveys*, 22(1), 31–72.

Canrtril, Hadley, and Gordon W. Allport (1935), *The Psychology of Radio*, New York: Harper.

Cattell, Raymond B. (1966), “The Scree Test For The Number Of Factors,” *Multivariate Behavioral Research*, 1(2), 245–76.

Chaffey, Dave (2016), “Mobile Marketing Statistics compilation,” <http://www.smartinsights.com/mobile-marketing/mobile-marketing-analytics/mobile-marketing-statistics/> (Feb. 20, 2017).

Chang, Lulu (2015), “Americans spend an alarming amount of time checking social media on their phones,” <http://www.digitaltrends.com/mobile/informate-report-social-media-smartphone-use/> (Oct. 5, 2016).

Chatterjee, Patrali, Donna L Hoffman, and Thomas P Novak (2003), “Modeling the Clickstream: Implications for Web-Based Advertising Efforts,” *Marketing Science*, 22(4), 520–41.

Childers, Terry L, Christopher L Carrb, Joann Peck, and Stephen Carson (2001), “Hedonic and utilitarian motivations for online retail shopping behavior,” *Journal of Retailing*, 77(4), 511–35.

Cho, Chang-hoan, and Hongsik John Cheon (2004), “Why Do People Avoid Advertising on the Internet?,” *Journal of Advertising*, 33(4), 89–97.

Christensen, Björn, Dominik Papies, Dennis Proppe, and Michel Clement (2014), “Gütemaße der logistischen Regression bei unbalancierten Stichproben,” *WiSt - Wirtschaftswissenschaftliches Studium*, 43(4), 211–13.

Cronbach, Lee J. (1951), “Coefficient alpha and the internal structure of tests,” *Psychometrika*, 16(3), 297–334.

Danaher, Peter J, Michael S Smith, Kulan Ranasinghe, and Tracey S Danaher (2015), “Where, when, and how long: Factors that influence the redemption of mobile phone coupons.,” *Journal of Marketing Research*, 52(5), 710–25.

Darden, William R., and Michael J. Dorsch (1990), “An action strategy approach to examining shopping behavior,” *Journal of Business Research*, 21(3), 289–308.

Dehejia, Rajeev H, and Sadek Wahba (2002), “Propensity Score-Matching Methods for Nonexperimental Causal Studies,” *The Review of Economics and Statistics*, 84(1), 151–61.

Dellarocas, Chrysanthos, and Ritu Narayan (2006), “A Statistical Measure of a Population’s Propensity to Engage in Post-Purchase Online Word-of-Mouth,” *Statistical Science*, 21(2), 277–85.

Dhar, Subhankar, and Upkar Varshney (2011), “Challenges and business models for mobile location-based services and advertising,” *Communications of the ACM*, 54(5), 121–29.

DiPrete, Thomas A., and Markus Gangl (2004), “Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments,” *Sociological Methodology*, 34(1), 271–310.

Drèze, Xavier, and François-Xavier Hussherr (2003), “Internet advertising: Is anybody watching?,” *Journal of Interactive Marketing*, 17(4), 8–23.

eMarketer (2015), “Mobile to Account for More than Half of Digital Ad Spending in 2015,” <http://www.emarketer.com/Article/Mobile-Account-More-than-Half-of-Digital-Ad-Spending-2015/1012930> (Feb. 14, 2016).

——— (2016), “Worldwide Retail Ecommerce Sales Will Reach $1.915 Trillion This Year,” <https://www.emarketer.com/Article/Worldwide-Retail-Ecommerce-Sales-Will-Reach-1915-Trillion-This-Year/1014369> (Dec. 5, 2016).

Evans, David S (2009), “The Online Advertising Industry: Economics, Evolution, and Privacy,” *Journal of Economic Perspectives*, 23(3), 37–60.

Evans, Kenneth R, Tim Christiansen, and James D Gill (1996), “The Impact of Social Influence and Role Expectations on Shopping Center Patronage Intentions,” *Journal of the Academy of Marketing Science*, 24(3), 208–18.

Floyd, Kristopher, Ryan Freling, Saad Alhoqail, Hyun Young Cho, and Traci Freling (2014), “How Online Product Reviews Affect Retail Sales: A Meta-analysis,” *Journal of Retailing*, 90(2), 217–32.

Fong, Nathan M., Zheng Fang, and Xueming Luo (2015), “Geo-Conquesting: Competitive Locational Targeting of Mobile Promotions,” *Journal of Marketing Research*, 52(5), 726–35.

Fulk, Janet (1993), “Social construction of communication technology,” *Academy of Management Journal*, 36(5), 921–50.

Gensler, Sonja, Peter Leeflang, and Bernd Skiera (2012), “Impact of online channel use on customer revenues and costs to serve: Considering product portfolios and self-selection,” *International Journal of Research in Marketing*, 29(2), 192–201.

——— (2013), “Comparing methods to separate treatment from self-selection effects in an online banking setting,” *Journal of Business Research*, 66(9), 1272–78.

Ghose, Anindya, Avi Goldfarb, and Sang Pil Han (2013), “How is the mobile internet different? Search costs and local activities,” *Information Systems Research*, 24(3), 613–31.

Ghose, Anindya, Sang Pil Han, and Sunghyuk Park (2013), “Analyzing the Interdependence between Web and Mobile Advertising.”

Ghose, Anindya, and Vilma Todri (2016), “Towards a Digital Attribution Model: Measuring the Impact of Display Advertising on Online Consumer Behavior,” *MIS Quarterly*, (forthcoming).

di Girogi, Erin (2015), “Measurement Proves the Effectiveness of Mobile Advertising,” <https://www.comscore.com/Insights/Blog/Measurement-Proves-the-Effectiveness-of-Mobile-Advertising> (Sep. 1, 2016).

Goldfarb, Avi, and Catherine Tucker (2011), “Online Display Advertising: Targeting and Obtrusiveness,” *Marketing Science*, 30(3), 389–404.

——— (2015), “Standardization and the Effectiveness of Online Advertising,” *Management Science*, 61(11), 2707–19.

Goodrich, Kendall (2011), “Anarchy of effects? Exploring attention to online advertising and multiple outcomes,” *Psychology and Marketing*, 28(4), 417–40.

Google (2015a), “Guide to ad sizes,” <https://support.google.com/adsense/answer/6002621?hl=en> (Jul. 24, 2015).

——— (2015b), “Active View metrics,” <https://support.google.com/dfp_premium/answer/3154105?hl=en> (Jul. 24, 2015).

Greenberg, Edward (2008), *Introduction to Bayesian Econometrics*, New York, New York, USA: Cambridge University Press.

Grewal, Dhruv, Yakov Bart, Martin Spann, and Peter Pal Zubcsek (2016), “Mobile Advertising: A Framework and Research Agenda,” *Journal of Interactive Marketing*, 34, 3–14.

Guitart, Ivan A., and Guillaume Hervet (2016), “The impact of contextual television ads on online conversions: An application in the insurance industry,” *International Journal of Research in Marketing*, (forthcoming).

Guttman, Louis (1954), “Some necessary conditions for common-factor analysis,” *Psychometrika*, 19(2), 149–61.

Haario, Heikki, Eero Saksman, and Johanna Tamminen (2001), “An Adaptive Metropolis Algorithm,” *Bernoulli*, 7(2), 223–42.

Hamilton, James D (1989), “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle,” *Econometrica*, 57(2), 357–84.

Hart, Sandee (2014), “Mobile vs. Desktop: 10 Key Differences,” <https://www.paradoxlabs.com/blog/mobile-vs-desktop-10-key-differences/> (Feb. 19, 2017).

Heckman, James (1997), “Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations,” *The Journal of Human Resources*, 32(3), 441–62.

Heckman, James J. (1981), *Heterogeneity and State Dependence*, *Studies in Labor Markets*, Chicago: University of Chicago Press.

Heckman, James J, Hidehiko Ichimura, and Petra E Todd (1997), “Matching Evidence Job As An Econometric Estimator : Evaluating Programme a Evaluation from Training,” *The Review of Economic Studies*, 64(4), 605–54.

Heckman, James, and Salvador Navarro-Lozano (2004), "Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models," *The Review of Economics and Statistics*, 86(1), 30–57.

Heine, Christopher (2013), "The Top 7 Reasons Why Mobile Ads Don't Work," <http://www.adweek.com/digital/top-7-reasons-why-mobile-ads-dont-work-153202/> (Mar. 1, 2017).

Hennig-Thurau, Thorsten, Kevin P. Gwinner, Gianfranco Walsh, and Dwayne D. Gremler (2004), "Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet?," *Journal of Interactive Marketing*, 18(1), 38–52.

Hennig-Thurau, Thorsten, and Gianfranco Walsh (2004), "Electronic Word-of-Mouth : Motives for and Consequences of Reading Customer Articulations on the Internet," *International Journal of Electronic Commerce*, 8(2), 51–74.

Hiniker, Alexis, Shwetak N. Patel, Tadayoshi Kohno, and Julie A. Kientz (2016), "Why would you do that? Predicting the uses and gratifications behind smartphone-usage behaviors," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '16*, New York, New York, USA: ACM Press, 634–45.

Hirschman, Elizabeth C, and Morris B Holbrook (1982), "Hedonic Consumption : Emerging Concepts , Methods and," *Journal of Marketing*, 46(3), 92–101.

Hoban, Paul R., and Randolph E. Bucklin (2015), "Effects of Internet Display Advertising in the Purchase Funnel: Model-Based Insights from a Randomized Field Experiment," *Journal of Marketing Research*, 52(3), 375–93.

Hoffman, Donna L., and Thomas P. Novak (1996), "Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations," *Journal of Marketing*, 60(3), 50–68.

Hong, Weiyin, James Y.L. Thong, and Kar Yan Tam (2004), "The effects of information format and shopping task on consumers' online shopping behavior: A cognitive fit perspective," *Journal of Management Information Systems*, 21(3), 149–84.

Hoskins, Britta (2013), “Four common complaints about advertising (and how to deal with them),” <http://www.dmnews.com/marketing-strategy/four-common-complaints-about-advertising-and-how-to-deal-with-them/article/323724/> (Jan. 14, 2018).

Hosmer, Davd W., and Stanley Lemeshow (2000), *Applied Logistic Regression*, New York, New York, USA: John Wiley & Sons, Inc.

IAB (2015), “Display & Mobile Advertising Creative Format Guidelines,” <http://www.iab.com/wp-content/uploads/2015/11/IAB_Display_Mobile_Creative_Guidelines_HTML5_2015.pdf> (Feb. 18, 2017).

Kaiser, Henry F. (1958), “The varimax criterion for analytic rotation in factor analysis,” *Psychometrika*, 23(3), 187–200.

Kaiser, Henry F, and K Dickmann (1959), “Analytic determination of common factors,” *American Psychologist*, 14, 425.

Kaiser, Henry F, and J. Rice (1974), “Little Jiffy, Mark Iv,” *Educational and Psychological Measurement*, 34(1), 111–17.

Katz, Elihu (1959), “Mass communication Research and the study of culture: An Editorial Note on a Possible Future for this Journal,” *Departmental Papers (ASC)*, 1–6.

Khachaturyan, A, S Semenovsovskaya, and B Vainshtein (1981), “The thermodynamic approach to the structure analysis of crystals,” *Acta Crystallographica Section A*, 37(5), 742–54.

Kim, Hongki, Kil-Soo Suh, and Un-Kon Lee (2013), “Effects of collaborative online shopping on shopping experience through social and relational perspectives,” *Information & Management*, 50(4), 169–80.

Knox, George, and Rutger van Oest (2014), “Customer Complaints and Recovery Effectiveness: A Customer Base Approach,” *Journal of Marketing*, 78(5), 42–57.

Kuisma, Jarmo, Jaana Simola, Liisa Uusitalo, and Anssi Öörni (2010), “The Effects of Animation and Format on the Perception and Memory of Online Advertising,” *Journal of Interactive Marketing*, 24(4), 269–82.

Kumar, N., and I. Benbasat (2006), “Research Note: The Influence of Recommendations and Consumer Reviews on Evaluations of Websites,” *Information Systems Research*, 17(4), 425–39.

Kumar, V., L. Aksoy, B. Donkers, R. Venkatesan, T. Wiesel, and S. Tillmanns (2010), “Undervalued or Overvalued Customers: Capturing Total Customer Engagement Value,” *Journal of Service Research*, 13(3), 297–310.

Kumar, V., S. Sriram, Anita Luo, and Pradeep. K. Chintagunta (2011), “Assessing the Effect of Marketing Investments in a Business Marketing Context,” *Marketing Science*, 30(5), 924–40.

Lambrea, Madalina (2016), “Mobile vs Desktop: 13 Essential User Behaviors,” <https://www.appticles.com/blog/2016/03/mobile-vs-desktop-13-essential-user-behaviors/> (Feb. 19, 2017).

Lambrecht, Anja, and Catherine Tucker (2013), “When Does Retargeting Work? Information Specificity in Online Advertising,” *Journal of Marketing Research*, 50(5), 561–76.

van der Lans, Ralf, Michel Wedel, and F. G. M. (Rik) Pieters (2014), “Brand Search Benefits of Online Advertising: An Eye-Tracking Experiment,” *SSRN Electronic Journal*, Robert H. Smith School Research Paper No. RHS 2460284, 1–30.

Lechner, Michael (2001), “Identification and estimation of causal effects of multiple treatments under the conditional independence assumption,” in *Econometric Evaluation of Labour Markets Policies*, M. Lechner and F. Pfeiffer, eds., Heidelberg: Physica-Verlag, 1–18.

Leonard, Heather (2013), “Mobile Web Traffic Is Consistent Throughout The Day,” <http://www.businessinsider.com/mobile-web-browsing-traffic-is-steady-2013-33?IR=T> (Feb. 17, 2016).

Li, Hongshuang (Alice), and P.K. Kannan (2014), “Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment,” *Journal of Marketing Research*, 51(1), 40–56.

Li, Xinxin, Lorin M. Hitt, and Z. John Zhang (2011), “Product Reviews and Competition in Markets for Repeat Purchase Products,” *Journal of Management Information Systems*, 27(4), 9–42.

Lindner, Matt (2016), "Burberry ties together its online and offline stores," <https://www.digitalcommerce360.com/2016/11/14/burberry-ties-together-its-online-and-offline-stores/> (Jan. 14, 2018).

Luo, Anita, and V. Kumar (2013), "Recovering Hidden Buyer–Seller Relationship States to Measure the Return on Marketing Investment in Business-to-Business Markets," *Journal of Marketing Research*, 50(1), 143–60.

Luo, Xueming (2005), "How Does Shopping With Others Influence Impulsive Purchasing?," *Journal of Consumer Psychology*, 15(4), 288–94.

Luo, Xueming, Michelle Andrews, Zheng Fang, and Chee Wei Phang (2014), "Mobile Targeting," *Management Science*, 60(7), 1738–56.

MacDonald, Ian L-, and Walter Zucchini (1997), *Hidden Markov and other models for discrete-valued time series*, Chapman & Hall.

MacInnis, Deborah J., Ambar G. Rao, and Allen M. Weiss (2002), "Assessing When Increased Media Weight of Real-World Advertisements Helps Sales," *Journal of Marketing Research*, 39(4), 391–407.

Maclnnis, Deborah J., and Bernard J. Jaworski (1989), "Information Processing from Advertisements: Toward an Integrative Framework," *Journal of Marketing*, 53(4), 1.

Mathwick, Charla (2002), "Understanding the online consumer: A typology of online relational norms and behavior," *Journal of Interactive Marketing*, 16(1), 40–55.

McAlexander, James H., John W Schouten, and Harold F Koenig (2002), "Building Brand Community," *Journal of Marketing*, 66(1), 38–54.

McDonough, Michele (2016), "Mobile Usage Statistics: Key Facts and Findings for Publishers," <http://blog.ezoic.com/mobile-usage-statistics-key-facts-and-findings-for-publishers/> (Oct. 5, 2016).

Minnema, Alec, Tammo H.A. Bijmolt, Sonja Gensler, and Thorsten Wiesel (2016), "To Keep or Not to Keep: Effects of Online Customer Reviews on Product Returns," *Journal of Retailing*, 92(3), 253–67.

Mithas, S., and M. S. Krishnan (2009), "From Association to Causation via a Potential Outcomes Approach," *Information Systems Research*, 20(2), 295–313.

Mithas, Sunil, M. S. Krishnan, and Claes Fornell (2005), "Why Do Customer Relationship Management Applications Affect Customer Satisfaction?," *Journal of Marketing*, 69(4), 201–9.

Moe, Wendy W. (2003), "Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream," *Journal of Consumer Psychology*, 13(1–2), 29–39.

Moe, Wendy W., and David a. Schweidel (2012), "Online Product Opinions: Incidence, Evaluation, and Evolution," *Marketing Science*, 31(3), 372–86.

Molitor, Dominik, Philipp Reichhart, Martin Spann, and Anindya Ghose (2015), "Measuring the Effectiveness of Location-Based Pull Advertising: A Randomized Field Experiment," *SSRN Electronic Journal*.

Montoya, R., O. Netzer, and K. Jedidi (2010), "Dynamic Allocation of Pharaceutical Detailing and Sampling for Long-Term Profitability," *Marketing Science*, 29(5), 909–24.

Moses, Lucia (2014), "You're not hallucinating: Ads are getting more intrusive," <http://digiday.com/publishers/youre-seeing-annoying-ads/> (Jan. 9, 2017).

Nelder, J. A., and R. Mead (1965), "A Simplex Method for Function Minimization," *The Computer Journal*, 7(4), 308–13.

Netzer, Oded, James. M. Lattin, and V. Srinivasan (2008), "A Hidden Markov Model of Customer Relationship Dynamics," *Marketing Science*, 27(2), 185–204.

Norton, Edward C. EC, Hua Wang, and Chunrong Ai (2004), "Computing interaction effects and standard errors in logit and probit models," *Stata Journal*, 4(2), 154–67.

Novak, Thomas P., Donna L. Hoffman, and Yiu-Fai Yung (2000), "Measuring the Customer Experience in Online Environments: A Structural Modeling Approach," *Marketing Science*, 19(1), 22–42.

Ofcom (2015), "Adults' media use and attitudes Report 2015," <https://www.ofcom.org.uk/__data/assets/pdf_file/0014/82112/2015_adults_media_use_and_attitudes_report.pdf> (Nov. 28, 2016).

Oliver, Richard L. (2009), *Satisfaction: A behavioral perspective on the consumer*, New York: M.E. Sharpe.

Patel, Dev, Jeremy Schneider, and Kushan Surana (2013), "Cracking the Mobile Advertising Code," <https://hbr.org/2013/09/cracking-the-mobile-advertisin> (Sep. 1, 2016).

Petty, Richard E., and John T. Cacioppo (1986), *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*, New York: Springer.

Pieters, Rik, L Warlop, and Michel Wedel (2002), "Breaking through the clutter: Benefits of advertisement originality and familiarity for brand attention and memory," *Management Science*, 48(6), 765–81.

Pieters, Rik, and Michel Wedel (2004), "Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-Size Effects," *Journal of Marketing*, 68(2), 36–50.

——— (2012), "Ad Gist: Ad Communication in a Single Eye Fixation," *Marketing Science*, 31(1), 59–73.

Pieters, Rik, Michel Wedel, and Rajeev Batra (2010), "The Stopping Power of Advertising: Measures and Effects of Visual Complexity," *Journal of Marketing*, 74(5), 48–60.

Pieters, Rik, Michel Wedel, and Jie Zhang (2007), "Optimal Feature Advertising Design Under Competitive Clutter," *Management Science*, 53(11), 1815–28.

Ratcliff, Christopher (2014), "More than 40% of online adults are multi-device users: stats," <https://econsultancy.com/blog/64464-more-than-40-of-online-adults-are-multi-device-users-stats/> (Feb. 26, 2017).

Ray, Soumya, Sung S Kim, and James G Morris (2014), "The Central Role of Engagement in Online Communities," *Information Systems Research*, 25(3), 528–46.

Reichheld, Frederick F (2003), "The one number you need to grow.," *Harvard business review*, 81(12), 46–54, 124.

Reinartz, Werner, Benedict Dellaert, Manfred Krafft, V. Kumar, and Rajan Varadarajan (2011), "Retailing Innovations in a Globalizing Retail Market Environment," *Journal of Retailing*, 87(1), S53–66.

Del Rey, Jason (2012), "Will Mobile's Massive Growth Ever Equal Real Revenue?," <http://adage.com/article/digital/mobile-s-massive-growth-equal-real-revenue/237511/> (Sep. 1, 2016).

Rishika, Rishika, Ashish Kumar, Ramkumar Janakiraman, and Ram Bezawada (2013), "The Effect of Customers' Social Media Participation on Customer Visit Frequency and Profitability: An Empirical Investigation," *Information Systems Research*, 24(1), 108–27.

Roberts, G. O., A. Gelman, and W. R. Gilks (1997), "Weak convergence and optimal scaling of random walk Metropolis algorithms," *Annals of Applied Probability*, 7(1), 110–20.

Rohm, Andrew J, and Vanitha Swaminathan (2004), "A typology of online shoppers based on shopping motivations," *Journal of Business Research*, 57(7), 748–57.

Rosenbaum, Paul R. (1987), "Sensitivity analysis for certain permutation inferences in matched observational studies," *Biometrika*, 74(1), 13–26.

Rosenbaum, Paul R., and Donald B Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70(1), 41–55.

Rosenbaum, Paul R., and Donald B. Rubin (1985), "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score," *The American Statistician*, 39(1), 33.

Rosenbaum, Paul R, and Donald B Rubin (1984), "Reducing Bias in Observational Studies Using Score on the Propensity Subolassification," *Journal of the American Statistical Association*, 79(387), 516–24.

Rossi, Peter E. (2014), "Even the Rich Can Make Themselves Poor: A Critical Examination of IV Methods in Marketing Applications," *Marketing Science*, 33(5), 655–72.

Rossi, Peter E., and Greg M. Allenby (2003), "Bayesian Statistics and Marketing," *Marketing Science*, 22(3), 304–28.

Rossi, Peter E., Greg M. Allenby, and Robert E Mcculloch (2005), *Bayesian Statistics and Marketing*, New York: Wiley.

Rubin, Alan M (1984), "Ritualized and Instrumental Television Viewing," *Journal of Communication*, 34(3), 67–77.

Rutz, Oliver J, Randolph E Bucklin, and Garrett P Sonnier (2012), "A Latent Instrumental Variables Approach to Modeling Keyword Conversion in Paid Search Advertising," *Journal of Marketing Research*, 49(3), 306–19.

Schumann, Jan H., Florian von Wangenheim, and Nicole Groene (2014), "Targeted Online Advertising: Using Reciprocity Appeals to Increase Acceptance Among Users of Free Web Services," *Journal of Marketing*, 78(1), 59–75.

Schweidel, David. A., Eric. T. Bradlow, and Peter. S. Fader (2011), "Portfolio Dynamics for Customers of a Multiservice Provider," *Management Science*, 57(3), 471–86.

Shankar, Venkatesh, and Sridhar Balasubramanian (2009), "Mobile Marketing: A Synthesis and Prognosis," *Journal of Interactive Marketing*, 23(2), 118–29.

Shankar, Venkatesh, Alladi Venkatesh, Charles Hofacker, and Prasad Naik (2010), "Mobile Marketing in the Retailing Environment: Current Insights and Future Research Avenues," *Journal of Interactive Marketing*, 24(2), 111–20.

Silverman, B. W. (1986), *Density estimation for statistics and data analysis*, New York: Chapman & Hall.

Simola, Jaana, Jarmo Kuisma, Anssi Öörni, Liisa Uusitalo, and Jukka Hyönä (2011), "The impact of salient advertisements on reading and attention on web pages.," *Journal of Experimental Psychology: Applied*, 17(2), 174–90.

Stafford, Thomas F., and Mark L. Gillenson (2004), "Motivations for Mobile Devices: Uses and Gratifications for M-Commerce," in *Proceedings of the Third Annual Workshop on HCI Research in MIS*, Washington, D.C., 70–74.

Stafford, Thomas F., Marla Royne Stafford, and Lawrence L. Schkade (2004), "Determining Uses and Gratifications for the Internet," *Decision Sciences*, 35(2), 259–88.

Statista.com (2014), "Changes in retargeting budgets expected by marketing professionals worldwide as of August 2014, by channel," <https://www.statista.com/statistics/446910/retargeting-budget-change/> (Jan. 14, 2018).

——— (2017), "Digital display advertising spending in the United States from 2014 to 2019 (in billion U.S. dollars)," <https://www.statista.com/statistics/273443/online-display-advertising-revenue-in-the-united-states/> (Jan. 14, 2018).

Steuer, Jonathan (1992), "Defining Virtual Reality: Dimensions Determining Telepresence," *Journal of Communication*, 42(4), 73–93.

Sutanto, Juliana, Elia Palme, Chuan-Hoo Tan, and Chee Wei Phang (2013), "Addressing the Personalization-Privacy Paradox: An Empirical Assessment From a Field Experiment on Smartphone Users," *MIS Quarterly*, 37(4), 1141–64.

Tauber, Edward M (1972), "Why Do People Shop?," *Journal of Marketing*, 36(4), 46–59.

Trusov, Michael, Liye Ma, and Zainab Jamal (2016), "Crumbs of the Cookie: User Profiling in Customer-Base Analysis and Behavioral Targeting," *Marketing Science*, 35(3), 405–26.

Urban, Glen L., Guilherme (Gui) Liberali, Erin MacDonald, Robert Bordley, and John R. Hauser (2014), "Morphing Banner Advertising," *Marketing Science*, 33(1), 27–46.

Vakratsas, Demetrios, and Tim Ambler (1999), "How Advertising Works: What Do We Really Know?," *Journal of Marketing*, 63(1), 26–43.

Verhoef, Peter C., P.K. Kannan, and J. Jeffrey Inman (2015), "From Multi-Channel Retailing to Omni-Channel Retailing," *Journal of Retailing*, 91(2), 174–81.

Voss, KE, ER Spangenberg, and Bianca Grohmann (2003), "Measuring the Hedonic and Utilitarian Dimensions of Consumer Attitude," *Journal of Marketing Research*, 40(3), 310–20.

Wang, Rebecca Jen-Hui, Edward C. Malthouse, and Lakshman Krishnamurthi (2015), "On the Go: How Mobile Shopping Affects Customer Purchase Behavior," *Journal of Retailing*, 91(2), 217–34.

Wedel, Michel, and Rik Pieters (2015), "The Buffer Effect: The Role of Color When Advertising Exposures Are Brief and Blurred," *Marketing Science*, 34(1), 134–43.

Wooldridge, Jeffrey M. (2010), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.

Xiao, Bo, and Izak Benbasat (2007), "E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact," *MIS Quarterly*, 31(1), 137–209.

Xu, Lizhen, Jason A Duan, and Andrew Whinston (2014), "Path to Purchase: A Mutually Exciting Point Process Model for Online Advertising and Conversion," *Management Science*, 60(6), 1392–1412.

Yoo, Chan Yun (2008), "Unconscious processing of Web advertising: Effects on implicit memory, attitude toward the brand, and consideration set," *Journal of Interactive Marketing*, 22(2), 2–18.

Youden, W. J. (1950), "Index for rating diagnostic tests," *Cancer*, 3(1), 32–35.

Yuen, Euphemia F T, and Sian S L Chan (2010), "The effect of retail service quality and product quality on customer loyalty," *Journal of Database Marketing & Customer Strategy Management*, 17(3–4), 222–40.

Zhao, Zhong (2004), "Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence," *The Review of Economic Studies*, 86(1), 91–107.

Zhu, L., I. Benbasat, and Z. Jiang (2010), "Let's Shop Online Together: An Empirical Investigation of Collaborative Online Shopping Support," *Information Systems Research*, 21(4), 872–91.

Convincing online consumers to purchase

Sascha Leweling

This dissertation deals with three different research questions: (1) “Can marketers improve the effectiveness of online display advertising by serving ads that match a consumer’s latent interest in the firm’s offering?”, (2) “Can marketers improve the effectiveness of online display ads by serving ads that match the device a consumer is using?” and (3) “Do user generated content and social shopping tools facilitate purchase decisions and affect customer revenue positively?” Each research question is examined empirically in its own chapter. This thesis provides new insights for online display advertising and how it influences the consumer.

18,20 €
ISBN 978-3-8405-0184-5