Master Thesis

# Inconsistencies between Review Sites and Facebook Comments about Web Shop Experiences

Moritz Höfter

University of Twente

M.Sc Business Administration

27-Feb-2017

Supervisors:

Dr. A.B.J.M. Wijnhoven (University of Twente), a.b.j.m.wijnhoven@utwente.nl

Dr. E. Constantinides (University of Twente), e.constantinides@utwente.nl

Dr. S. Perner (TU Berlin), susanne.perner@tu-berlin.de

**Abstract**

The Internet provides multiple sources of opinions about online shops and services. Among these, review sites may be the most popular. However, social media and social networks like Facebook catch up and generate more and more reviews. Different sources often lead to contradicting information about the trustworthiness of an e-vendor and force the consumer to make trade-offs. The present thesis examines how consumers deal with information inconsistencies between Facebook and review sites. A comprehensive analysis on web shops in the UK revealed that comments on Facebook are more negative than reviews on the review site Trustpilot. In a survey among 95 students, the particular importance of these two sources of opinions was examined through a conjoint analysis. The findings of this survey indicate that Facebook comments have less impact on the purchase decision than Trustpilot content. However, there is a large group that relies more heavily on Facebook when inconsistencies occur. Knowing how review sources differ in their information and their effect on consumers' decision, shop owners can react accordingly and increase sales figures. Furthermore, the findings could be beneficial in the development of multi-platform reputation systems. Knowing the impact of different platforms can enhance the calculation of cross-platform trust scores.

**Key words:** E-commerce, Social Media, Social Networks, Reviews, Reputation, Conjoint Analysis, Sentiment Mining, Semantria

# Table of Contents

# Index of Figures

## Index of Tables

# 1 Introduction

According to the American Internet Crime Complaint Center, in 2008 over 270.000 users became victims of online frauds. The total amount of dollars lost in all these cases exceeds 260 million US Dollar (Xiao & Bensabat, 2008, p. 169). Bhasker (2006) argues that the Internet has a “fundamentally insecure infrastructure”, which makes it difficult to defend it from attacks (Bhasker, 2006, p. 202). Online transactions occur often between parties that do not know each other and have no experiences with each other. Furthermore, there is often an information asymmetry in favor of the seller due to a lack of information about the provider or about the purchased goods and services. A physical inspection as in offline trading is not possible (McKnight, Choudhury, & Kacmar, 2002). Additionally, the vendor has more control over the transaction as soon as the money is paid to him or her. The consumer cannot test product or service quality in advance and needs to rely on the trustworthiness of the counterpart, which can be exploited by the marketer. In order to mitigate these information asymmetries, trust and reputation indicators in the form of reviews, trust seals and third party certificates were introduced in e-commerce (Dellarocas, 2000; Jøsang, Ismail, & Boyd, 2007). About 75% of the consumers take advice from online reviews into consideration when they are shopping for something they haven’t purchased in the past (Smith & Anderson, 2016). Messages from peers have a different influence on the perception of potential customers than messages directly from the company due to the different level of authenticity (Huang, Yoon, & Benyoucef, 2012). A study from 2007 emphasizes that customers pay up to 20% more for a product with a five star rating compared to a product with four stars (comScore Inc., 2007). Thus, electronic word of mouth is not only important for customers, but also for web shop providers.

According to a market research, 79% of all Americans are shopping online. About one-third is purchasing online up to a few times per month. Half of all Americans use customer reviews and ratings in advance regularly, about the same number at least sometimes. Nevertheless, many Americans mention concerns about the trustworthiness of online reviews. While one-half states that reviews give an accurate impression about the business in question, the other half doubts this (Smith & Anderson, 2016). This is a clear indication that making shopping on the Internet safer is still an up to date issue.

Since the turn of the millennium, social media has become ever more important in practice as well as in research. The online social network Facebook is the biggest online network in the world. Millions of social interactions happen here every day, which has an enormous potential for scientists and social scientists in particular (Wilson, Gosling, & Graham, 2012). But not only in social science research, also in e-commerce research social media is a major topic (Thelwall, Buckley, & Paltoglou, 2011; Thelwall, Wouters, & Fry, 2008). One of the core benefits of social media for e-commerce is the formation of a new communication channel between customer and businesses. As a result, customer relationships are becoming closer, the traffic on websites is increasing, and new products and brands are being developed through the newly created social environment (Huang et al., 2012). Through the emergence of social media, e-commerce changed from being product-centered to becoming more customer-centered. Listening and responding to the customer via social media is crucial for companies in order to create or maintain an advantage over their competitors (Heydari, Tavakoli, Salim, &

Heydari, 2015). Most of this information is publicly available so that companies can tap into their consumers' preferences, behavior and networks. On the other hand, social media create a communication channel from customer to customer and shift power to the consumer. This new type of interaction allows customers to share their shopping experiences directly with their social network and support each other in decision-making processes. However, the motives to use Facebook instead of review sites in order to publish word of mouth differ (Kreis & Gottschalk, 2015), resulting in inconsistencies in the information. In other words, there could be contradicting information about the reliability and trustworthiness of a shop on different platforms. While one source shows predominantly positive reviews about the shop, another could present mostly negative statements about experiences with the same company. While consistent information, meaning that both sources contain roughly the same information, is easy to process, inconsistent information can unsettle consumers. As a result, they have to weigh both of the sources and build their own opinion upon the information they have.

Many studies in the information system field deal with the impact of reviews on the purchase intention of the consumer (Zhang, Cheung, & Lee, 2014). While one party states that negative reviews have a stronger impact on consumer's choice than positive reviews (Lee, Park, & Han, 2008; Park & Lee, 2009; Smith & Anderson, 2016), others claim that even negative reviews can result in positive outcomes (Berger, Sorensen, & Rasmussen, 2010). Another study revealed that inconsistent reviews, meaning a mixture of positive and negative reviews, have an indirectly positive influence on the purchase intention (Zhang et al., 2014). Concluding, there is a great potential for research dealing with reviews and review inconsistencies in e-commerce.

Whether Facebook comments with review character are used in consumers' purchase decisions and what impact they have has not been studied so far. In comparison to traditional reviews, which are explicit, Facebook comments are highly implicit. But there are also other differences between the two review sources. More information about the author, the ability to interact with other users in the context of a product or an online shop, and the proximity cues in the network structure (such as mutual friends, shared interests) might influence the perceived trustworthiness of reviews on Facebook compared to anonymous reviews on review sites. Furthermore, the risk of review spam is mitigated. On Facebook, every comment can be traced back to the author's profile, making it difficult to build up a positive or negative reputation through fake reviews (Wilson et al., 2012). Nevertheless, there are aspects of Facebook that might narrow the value of Facebook comments as review source. On the one hand, firms, such as online shops, can edit the content on their page. Thus, they can remove negative comments in order to increase their online reputation. Furthermore, they can encourage loyal customers or employees to push their reputation through positive comments. But in doubtful situations, consumers can get in touch with the authors of the respective comments and verify their authenticity. Another aspect is the implicit character of Facebook comments. Facebook is not supposed to be a review platform while review sites are developed explicitly for the use as reputation information source. Thus, many consumers might not consider Facebook as an appropriate information source.

To the best of my knowledge, the literature on trust and reputation systems ignored the reviews on Facebook to a great extent. So far, there are only a few studies about motives and the usage of social networks for the publication of reviews (e.g. Kreis & Gottschalk, 2015). Different motives and uses could lead to inconsistent information between different sources, which has not been studied before.

Furthermore, the impact of Facebook reviews on the purchase intention is not observed yet. Due to the implicit character and several previously stated differences, a different impact is expected. The present study will concentrate on the impact of different review channels especially in situations where inconsistencies occur.

**This study aims to determine differences in the content of different review sources and to evaluate the particular impact of these sources on the purchase intention of the consumer.**

We have the following research questions:

**RQ1: Does the information about web shop experiences in Facebook comments differ from information on review sites?**

**RQ2: Which source is more influential for purchasers in inconsistent situations?**

To answer these questions, a literature review on trust and reputation systems, trust determinants and Facebook features is conducted. Based on this, a method for evaluating the sentiment of Facebook reviews and turning them into calculable ratings is developed. Then, a comparison of a broad selection of reviews on Facebook and a review site is conducted in order to determine the differences in the information direction (positive/negative). Finally, a conjoint analysis reveals the impact of different sources in inconsistent information situations.

By revealing the differences between reviews on Facebook and review sites and the extent to that the impact on the purchase intention varies, this study contributes to several fields in research and provides major implications for practice.

For researchers, this study provides a comprehensive literature review about trust antecedents and trust indicators in the e-commerce environment and enhances these by adding new indicators from the social network Facebook. Thus, this work enlarges the pool of trust indicators for trust building models (e.g. McKnight et al., 2002; Zhang et al., 2014). This thesis delineates the method of sentiment mining and the applicability to make Facebook reviews accessible and calculable. This can be relevant for researchers and practitioners likewise. Both can use this approach to include reviews on Facebook into their studies and reputation systems respectively. Implications as well as advantages and disadvantages in comparison to traditional trust indicators are discussed. Furthermore, a conjoint analysis is used to detect the impact of different information sources. Conjoint analyses are based on simulations of different situations and force the participant to trade-offs. Trade-offs are a major aspect of decision-making and therefore crucial in online shopping. An extensive description of the process can serve as a guideline for other researchers who intend to conduct a conjoint analysis in the information service domain.

This study stresses the importance of Facebook for trust building in e-commerce and reputation systems in particular. Facebook is adding new components to enhance the overall quality of such systems. Nevertheless, as Huang et al. (2012) stated, the specific characteristics of these components must be understood in order to improve the performance of e-trading.

The implications of this work could also help to avoid cold start problems of new shops and startups. Often, such companies have not enough referrals or reviews causing start problems, so called cold start problems. Social networks add a new source of opinion exchange. Social network buzz can create a vast amount of information in a relatively short amount of time and therefore be beneficial for startups and new e-commerce pages (Gottschlich, Heimbach, & Hinz, 2013). Further, the results of this study could help online consumers to assess comments about web shops on social media. Sometimes there is an overwhelming amount of reviews on different platforms, which makes it difficult for the consumer to evaluate the reliability of a shop. Decision-making becomes even more complex when inconsistencies between reviews or between different review channels occur. An inter-platform reputation score based on the results of this study could help to overcome the problem of information explosion. There are services that aggregate reviews and ratings from different platforms in order to calculate a cross-platform trust score and make it easier for consumers to evaluate the reputation of a web shop. Whether Facebook comments should be add to these sources will be examined in this study. However, Facebook features such as reviews in Facebook comments are barely used presumably due to the complexity of the translation in a computable value. To find a way to overcome these issues is also part of this study.

The present work is structured as follows: **Chapter 2** provides basic concepts as well as relevant theories by reviewing existing literature. Following this, the premises for the sentiment analysis are outlined. Furthermore, the sentiment-mining tool Semantria is explained in more detail. Presenting actual studies and use cases provides a description about the processes and features of this tool. The subsequent **chapter 3** describes the research design, including methodology as well as the sample and delineates the process of the observations. Due to the fact that the two sub-questions are not directly connected, the chapter is divided in two parts. This applies also for **chapter 4,** were the results are outlined. Afterwards, the results are combined to derive the key findings of this study. The following **chapter 5** discusses the key findings under consideration of limitations. At the very end, theoretical as well as practical implications are highlighted.

# 2 Theoretical Foundation

This chapter should give the reader a brief but explicit idea about the concepts this thesis is built upon.

## 2.1 Literature Search

In order to structure the present study, the reference manager and organizer Mendeley was used. For searching appropriate literature, the search engine google scholar was used due to its comprehensive search options like the limitation to certain journals or time periods. In some cases, Web of Science was used in order to find back- and forward citations of relevant studies. Furthermore, the online tool "Mendeley Suggest" was employed to identify additional relevant articles. Based on a selection of articles, this tool suggests related studies and topics.

## 2.2 Trust and Reputation in E-commerce

### 2.2.1 Conceptualization of Trust and Online Trust in Particular

Trust is relevant in many aspects of our daily life. As in love or friendship, trust is the basis for meaningful relationships (Carliner, 2004). In e-commerce, trust is especially important when someone is interacting with a person or an entity for the first time. In the literature, there is a huge variety of trust interpretations. This is due to the multitude of fields where trust plays a role in. McKnight and Chervany argue metaphorically that "psychologists analyzed the personality side, sociologists interviewed the social structural side, and economics calculated the rational choice side of the trust elephant" (McKnight and Chervany, 2000, p.29). Furthermore, trust is often replaced by the terms credibility, reliability and confidence in order to circumvent the difficult definition of trust (Carliner, 2004). Sometimes, trust is seen as a multi-faceted concept, including different dimensions such as emotions, cognition and behavior (e.g. Lewis & Weigert, 1985). So far, there is no established definition of trust. As Bigley and Pearce argue, a general definition is not possible without getting to complex or abstract for research (Bigley & Pearce, 1998). According to Kim et al. (2001), research on trust tends to be disconnected, case-specific and barely integrated (Kim, Song, Braynov, & Rao, 2001). In order to provide a structured menu of defined trust concepts for other researchers, McKnight and Chervany (2000) provide an interdisciplinary model of trust constructs (see Figure 1), which contains five types of trust. These concepts can be used to measure trust in empirical research.

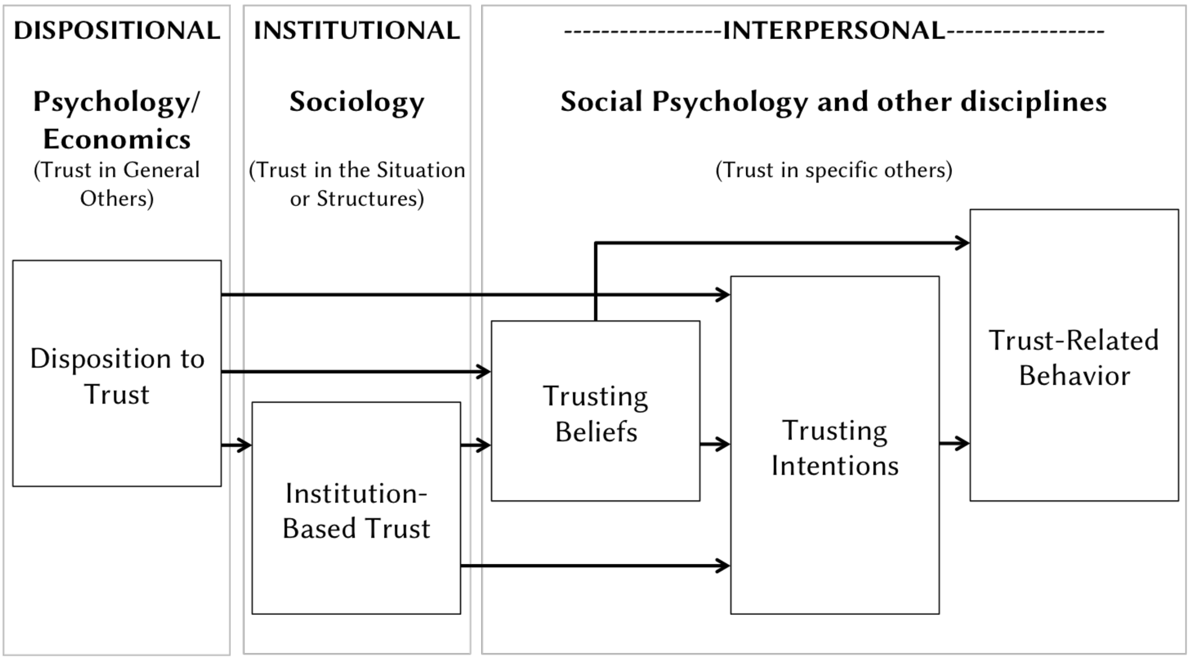


**Figure 1: Interdisciplinary model of trust constructs (McKnight and Chervany, 2000).**

McKnight and Chervany's model differentiates *dispositional trust*, which is mostly applied in psychology and economics, *institutional trust,* which is mostly used by sociology researchers, and *interpersonal trust,* used in social psychology and other disciplines. They list the five trust constructs *Disposition to Trust, Institution Based Trust, Trusting Beliefs, Trusting Intentions* and *Trust Related Behavior,* which will be briefly explained in the following.

*Disposition to Trust* is the tendency of a person to depend on general others. *Disposition to trust* differs from *Trusting Intentions* in so far that disposition describes the propensity of a person to trust others while *Trusting Intentions* is tied to a specific situation. This propensity can be traced back to the sub constructs "faith in humanity" and "trusting stance", with the assumption that trusting leads always to better outcomes than not to trust at all.

*Institution Based Trust* is – extremely simplified - the belief that there are higher conditions that will lead to a positive outcome. This can be laws as well as other regulations that will prevent from negative outcomes and mitigate the risk of the current situation. This must be distinguished from the concept of reputation. Reputation, which will be explained in the subsequent chapter, is the opinion someone (or a group of people) has about someone or something.

*Trusting Beliefs* are a person's beliefs that the other person's characteristics will be beneficial for oneself. This construct is of particular relevance for this thesis. The major sub-constructs as defined by McKnight and Chervany (2000) are "trusting belief-competence", "trusting belief-benevolence", "trusting belief-integrity" and "trusting belief-predictability".

*Trusting belief-competence* is the belief in a counterpart's ability and power to fulfill one's needs. In online transactions, this can be the belief that a merchant has the required infrastructure to deliver an order.

*Trusting belief-benevolence* is the belief in the counterpart's willingness to work in one's interest. In online trading, this could be the expectation that the merchant is not driven by opportunistic motives. According to a study by Kim, Xu and Koh (2004), trust building in e-commerce means the shaping of trust beliefs.

*Trusting belief-integrity* can be compared to reliability, meaning that the shop is honesty and keeping its promises, as to comply with the purchase agreements or delivery dates.

*Trusting Belief-predictability* is the belief that the other person's action can be predicted upon past interactions. This requires consistency in the behavior of the counterpart.

*Trusting Intentions* is a construct of trust that deals with the willingness to depend or the intention to depend on the counterpart. This means that one is willing to expose oneself to a situation that is out of his influence under consideration of all incorporated consequences. In the definition of the construct by McKnight and Chervany (2000), trust is not depending on a certain situation but on the other person. Sub-constructs are the "willingness to depend" and the "subjective probability of depending". The first is the readiness to get vulnerable to the other person; the second describes how likely a person would give himself into a dependency on another party.

The last construct is *Trust-related Behavior*. This construct is about the trusting behavior, meaning that there is an active action based on acceptance of the potential risks of fraud. This can be information sharing, informal agreements or commitment on business transactions. The latter is especially relevant for the present study.

The arrows in Figure 1 between trusting believes, trusting intentions and trust related behavior are due to the "theory of reasoned action" while all others are rather intuitive. They should present the possibility that all the constructs can relate to each other so that they can be implemented in one model of trust building (McKnight & Chervany, 2000). The theory of reasoned action was developed by Fishbein and Ayzen (1975). The theory implies that there is a causal relationship between behavioral intention, attitude, beliefs and subjective norms. The behavior of a person is determined by his or her behavioral intentions. Behavioral intention is influenced by a persons subjective norms and attitude while attitude is a function of beliefs (Fishbein, M. & Ajzen, 1975). Projected to the depicted model, this relationship corresponds to the chain of constructs from *Trusting Beliefs* to *Trust-related Behavior* (McKnight & Chervany, 2000).

A good example for the application of trust constructs in e-commerce research was conducted by McKnight, Cloudhury and Kacmar (2002). According to their study, trust is a "multi-dimensional construct with two inter-related components – trusting beliefs […] and trusting intentions-willingness to depend […]" (McKnight, Choudhury, & Kacmarc, 2002, p. 297). Trusting beliefs means the personal perception of trustworthiness**.** According to McKnight, Cummings and Chervany, the trusting beliefs consist of the following three beliefs: the belief in integrity, the belief in benevolence and the belief in competence. Integrity can be compared to reliability, meaning that the shop is honesty and keeps its promises; benevolence is the assumption that a merchant is working in the buyer's interests, and competence describes a vendor's ability of meeting the buyer's needs. Some other studies include also the additional belief in predictability (Gefen, 2002; Mcknight, Cummings, & Chervany, 1998). It is to mention that McKnight, Cloudhury and Kacmar deal with initial trust, which is about trust in unfamiliar web merchants. With the first transaction, the relationship becomes an ongoing interaction; the two counterparts have a personal opinion about each other based on a personal experience and may not need to rely on third party information. This is why this belief in predictability is excluded from the model. Willingness to depend – also called trusting intention - means the willingness to voluntary expose oneself to vulnerability. Hereby the authors mention the difference between trusting intention and the probability of depending. The higher the willingness, the more likely becomes the final action. The trust building model proposes three factors that build trust in online merchants: structural assurance, vendor reputation and web site quality. Structural assurance is the perceived safety of the web environment. The study revealed that perceived internet risk is negatively associated to consumers intentions which are split in three behaviors: First, to follow the vendors' advice, second, to share information with the vendor and last, to purchase from the shop. All other factors had a positive influence on the behavioral intentions of the consumer (McKnight et al., 2002).

The above-described model posits that several antecedents of trust, like the reputation of a vendor, are influencing the consumer's trusting beliefs. According to the model, this in turn affects the consumer's intention to trust the vendor.

Another study observes the moderating effect of inconsistent reviews on the online shopping decisions of consumers. In a laboratory experiment, different groups were formed to reveal this effect. It turned out that consumers who are exposed to inconsistent reviews tend to buy more likely than those who are not (Zhang et al., 2014). Nevertheless, the two groups lack some kind of comparability due to the fact that the control group was not exposed to consistent reviews instead, but to no reviews at all. The present study pursues a similar goal as the one presented by Zhang, Cheung and Lee (2014). **This study observes the influence of Facebook comments on the purchase intention, compared to the effect of traditional reviews in a situation where inconsistencies occur between the two sources. Furthermore, the occurrence of such inconsistencies is examined.** However, the research method and design of the present study are deviating from the two mentioned studies. Jones and George (1998) describe the evolution of trust as initial trust, trust stabilization and trust dissolution. **The focus of the present thesis is on initial trust**. In the situation of facing an unfamiliar shop, third party opinions might be one of the most important factors influencing the evaluation of a shop's trustworthiness. In the situation of a repeat purchase, the past experiences with service quality may have more impact (Jones & George, 1998). Nevertheless, there are studies which indicate that even repeat customers evaluate trustworthiness upon the current reputation instead of their own experience (Doney & Cannon, 1997). A possible explanation could be, that there is always a chance that the last experience was a lucky coincidence or that the situation has changed in the meantime.

### 2.2.2 Trust Determinants and Indicators in E-commerce

Hung, Wu and Chen (2014) argue that trust plays a major role in online environments than in the real world (Hung, Wu, & Chen, 2014). One of the reasons is that it became pretty easy to develop a serious looking shop on the Internet. Fraudulent online shops are mostly not distinguishable from serious businesses while dubious brick-and-mortar shops are easily to detect (Palmer, Bailey, & Faraj, 2000). Dellarocas (2001) argued that the risk of transactions gets greater the wider the two sides "are separated in time and space" (Dellarocas, 2001, p. 98). Thus, academics and the e-commerce industry developed several concepts and technical solutions in order to offer the consumer a wider range of cues or indicators related to trustworthiness. Such indicators are reviews, ratings, trust seals and third party certificates. Elements that promote trust are the key for enhancing consumers' trust in e-commerce. Such elements are often referred to as antecedents, underlying dimensions, determinants or principles of online trust (Wang & Emurian, 2005, p. 112).

Literature on such antecedents is divided in interpersonal trust related research and research about trust in organizations, such as electronic services. The present study deals with the latter and trust antecedents in e-commerce in particular. **After presenting the antecedents of trust, online indicators are assigned to every category of antecedents.** Beldad, de Jong and Steehouder (2010) conducted a literature review and clustered the findings in three types of antecedents: (1) Customer/client-based antecedents, (2) website-based antecedents and (3) company-based antecedents. Most literature was found in the field of e-commerce. (1) The first type of antecedents is including the consumer's propensity to trust and his or her experience and proficiency in Internet usage. (2) Website-based trust antecedents are the perceived ease of use, the quality of the information, graphical characteristics as well as social presence cues. Other antecedents in this matter are customization capacities, privacy assurance and security features as well as third-party guarantees. (3) In the third group of antecedents are an organization's reputation, its size, its offline presences as well as prior experiences with the online organization. Regarding the perceived ease of use as a website-based trust antecedent, there are several findings in the review of Beldad et al. (2010). Grabner-Kraeuter (2002) found that a website's navigation is among the best ways to convey trustworthiness. Whenever consumers feel comfortable in handling a web shop, there is a higher probability that these rely on the shop. Other findings, for instance that website content should be free from errors and present accurate as well as complete information, seem almost evident.

Online transactions lack a crucial detail compared to traditional transactions: there is no face-to-face contact with the counterpart. This makes it difficult to evaluate the trustworthiness of the seller. In order to deal with that issue, social presence cues were implemented in web-shops. Such social presence cues interfere the impression that there is another human being involved in the transaction. The feeling of interpersonal interaction leads to higher trust in a website. In order to create social presence cues, pictures of the merchant are presented on the web page of online shops (Riegelsberger, Sasse, & McCarthy, 2003). Even though they are not mentioned in this context, integrated chat features may also be social presence cues. Providing the opportunity to communicate in real-time with a person of the sales team might convey trustworthiness. Nevertheless, it might depend on the functionality of the feature and the professionalism of the contact person. Privacy assurance and security features are the next trust antecedent category presented by Beldad et al. (2010). Referring to several studies, the major findings are that security concerns have a higher impact on trust than privacy concerns. The authors encourage to implement strong privacy statements and security features on web shops. Third-party guarantees are the last website-based antecedent that is presented. The approach is based on Doney and Cannon (1998)**.** Third-party guarantees build upon the idea that trust can be transferred from a proven source to another entity. Thus, a third party recommends the transaction partner who therefore gets more trustworthiness. This trust transfer is often presented in form of seals or banners (Cheung & Lee, 2006). These are especially important in the aforementioned privacy and security features. However, there are contradictory research results about the effects of trust seals. A possible explanation is that the guarantees or the organization behind them is also unfamiliar to the trustee (Beldad, Jong, & Steehouder, 2010).

Coming to the company/organization based trust antecedents, an organization's reputation plays a major role. Positive reputation leads to more openness and higher trust while in the opposite case a negative reputation is lowering the perceived trustworthiness.

Reputation as defined by Sabater and Sierra (2001) is the "opinion or view of one about something" (Sabater & Sierra, 2001, p. 1). In their interaction, people exchange opinions about an entity, often based on past experiences or narratives. Nevertheless, the personal opinion about an entity depends on different variables. Reputation has an individual dimension, a social dimension as well as an ontological dimension. Ontology is referred to as the individual point of view every person has. Therefore, the perception of reputation differs between individuals (Sabater & Sierra, 2001). Others describe reputation as the expectation someone has about the behavior of his or her counterpart, depending on information about past experiences. These could also be narrated opinions of others (Abdul-rahman, Hailes, Street, & Kingdom, 1999). Misztal (1996) argues that reputation is helpful in order to "manage the complexity of social life" and particularly in the selection of trustworthy interaction partners. In her article, Misztal mentions that reputation can be manipulated and stereotyped, which makes the concept more ambiguous than the one of trust (Misztal, 1996). However, Grishchenko defines reputation as the expectation that the level of compliance is "near to an average compliance level of past events by the same responsible entities" (Grishchenko, 2004, p. 2), adding a responsibility component.

Wang et al. base their definition on recommendations received from peers, implying one agent's belief in the others agent's capability, honesty and reliability. The authors mention two ways to gather these recommendations from others. One is to ask a third party that is accumulating these, as for instance online reputation services like review sites. However, it is to consider that such services often do not care about the identification of the sources while relying on the amount of data to make possible bias insignificant. The other way is to collect such evaluations decentralized by ones own (Wang & Vassileva, 2003).
Referring to Hussain et al. (2007), some of the reputation definitions lack a component of context and dynamic nature (Hussain, Hussain, & Chang, 2007). Taking this into consideration and combining it with aforementioned definitions, reputation is considered as the recurring aggregation of opinions from peers about an entity in order to support personal decision-making processes.

Online review sites precede exactly these aggregation and combination processes. Their mission is to aggregate different opinions on their sites, weigh them out and offer the information to potential online customers (peers) that are uncertain about their counterpart. Some of the platforms even have specific algorithms that consider the dynamic characteristic by weighing reviews upon the time of their emergence (e.g. Trustpilot) and other criteria (Trustpilot, n.d.). In contrast, reviews on social networks are less controlled by any authority but may have other interesting implications for online trust and reputation. In order to explain the creation of online reputation in more detail, the subsequent chapter will summarize literature on reputation generation systems.

Coming back to online trust, another antecedent for trust in organizations or companies is the perceived size of the organization. A study from Jarvenpaa and Tractinsky (1999) revealed that the size of an organization has a positive impact on trust. This is partly explained by the price (Beldad et al., 2010; Jarvenpaa, Tractinsky, & Vitale, 2000). However, the impact of the price on trust could be doubted. People tend to choose the cheapest opportunity. Another assumption could be that the perceived size (which is tough to be manipulated as one could expect that bigger merchants also have reached some reputation in offline context) of a shop speaks for a longer existence and therefore for some kind of right to exist since markets would probably separate out the weakest participants.

The last trust antecedent referring to a website's characteristics is prior experience or familiarity with the company. Whenever a first interaction was successful, people tend to trust the party again. Thus, familiarity with a shop is an important indicator for the trustworthiness of a shop. However, as many other studies, this thesis deals with initial trust such as situations where the consumer has no prior experience with the merchant.

In the following table (Table 1), all determinants for website and organization based trust provided by the study from Beldad et al. (2010) are listed and respective trust indicators and features in web-shops are provided (i.e. features and indicators that are implemented on the website of a merchant. Offline indicators are not relevant in this study. Recurrent indicators are not listed repeatedly.)

The list is supposed to show the different trust antecedents and technical features that can enhance trust. Furthermore, some indicators that were found are eluded due to the assumption of weak association to trust ("minimalistic design" was not included in the antecedent of "perceived ease of use". Although there might be an influence on the usability of a website which can influence trust, the direct association to trust in an e-vendor was expected to be weak.)

**Table 1: Trust antecedents and indicators. Based on Beldad et al. (2010)**

| | Antecedents | Indicators | Study |
|---|---|---|---|
| Website based trust antecedents | Perceived ease of use of the website | Navigational features<br>Search functions<br>Product indices<br>Product comparison features | (Lohse & Spiller, 1998) |
| | | Effective navigation | (Grabner-Kraeuter, 2002) |
| | | Navigation and presentation<br>Absence of errors | (Bart, Shankar, Sultan, & Urban, 2005) |
| | | Language | (Lederer, Maupin, Sena, & Zhuang, 2000) |
| | | Short loading time | (Koehn, 2003) |
| | Information quality | Correct spelling, grammar and syntax | (Koehn, 2003) |
| | | Content quality | (Liao, Palvia, & Lin, 2006) |
| | Graphical characteristics | Clip Arts<br>Colors | (Kim & Moon, 1998) |
| | Social presence cues | Photographs* | (Riegelsberger et al., 2003) |
| | | Live help interfaces | (Qiu & Benbasat, 2005) |
| | Customization and personalization capacity | Willingness to customize products | (Koufaris & Hampton-Sosa, 2004) |
| | | Provision of customized services** | (Nusair & Kandampully, 2008) |
| | Privacy assurances and security features | Third party privacy seals<br>Privacy statements<br>Third party security seals<br>Security features | (Belanger, Hiller, & Smith, 2002) |
| | Third-party guarantees | Privacy assurance (seals)<br>Process assurance (seals)<br>Technology assurance (seals) | (Kimery, Mccord, & Kimery, 2002) |
| | | Awards<br>Page rankings<br>Website links to the page | (Toms & Taves, 2004) |
| Company/organization-based trust antecedents | Organizational reputation | Referrals<br>**Ratings**<br>**Reviews** | (Jøsang et al., 2007) |
| | | Website links to the page<br>Page rankings<br>**Reviews and Ratings**<br>Awards | (Toms & Taves, 2004) |
| | Perceived size of the organization | Number of items available | (Jarvenpaa et al., 2000) |
| | Offline presence | List of offline shops | Newly added |
| | Experience and familiarity with the online company | Consensus information**<br>Brand familiarity** | (Benedicktus, Brady, Darke, & Voorhees, 2010) |
| | Intermediaries** | Escrow services**<br>Credit card payment** | (Pavlou & Gefen, 2004) |
| * | Also important on review sites (reviewer profile) (Riegelsberger et al., 2003) | | |
| ** | Newly added. Not included in the literature review by Beldad et al. (2010) | | |

The outlined list has a collection of measures for future research projects. Based on these antecedents and the previously discussed trust constructs, the effect of single trust determinants or combinations of trust indicators on trust-related behavior (such as purchase decisions) can be tested (e.g. McKnight et al., 2002).

Later, Facebook features and their use as trust indicators will be discussed.

### 2.2.3 Reputation Systems

While trust systems are showing a score of someone's trustworthiness seen from a person's subjective view, reputation systems produce a score reflecting a party's reputation as evaluated by the community. Thus, reputation systems can be seen as a medium to transfer trust from one party to another (Jøsang et al., 2007). The basic idea of reputation systems is that parties can rate each other. The aggregation of all ratings is a good indicator for the trustworthiness of the counterpart (Jøsang et al., 2007). Sometimes, reputation systems are called "collaborative sanctioning systems, referring to the characteristic that reputation systems sanction a bad service provision and encourage others to act in an appropriate manner (Mui, Mohtashemi, & Ang, 2001).

An approach to classify reputation systems is provided by Wang and Vassileva (2007). Their typology includes three classification criteria, grouping reputation systems in (1) centralized and decentralized, (2) person/agent and resource, and (3) global and personalized. (1) Centralized systems mean a central entity that aggregates and organizes all reputations. In decentralized systems, members organize the distribution by themselves. The second criteria (2) differentiate between the units that are reviewed. Persons or agent reputation is about the actors in the system while resource is related to items or services that are sold. The last criteria (3) refer to the availability or degree of personalization of reputation. In global systems, reputation is available for everyone in the system while personalized reputation is tailored for single members (such as the system developed by Cho, Kwon, & Park, 2009). The classification by Wang and Vassileva is illustrated in Table 2.

**Table 2: Trust and reputation system classification. Based on Wang & Vassileva (2007).**

| **Trust and Reputation Systems** | | | | | | | |
|---|---|---|---|---|---|---|---|
| Centralized | | | | Decentralized | | | |
| Personal/agent | | Resources | | Personal/agent | | Resources | |
| Global | Personalized | Global | Personalized | Global | Personalized | Global | Personalized |

Using this approach, review sites are seen as centralized, belonging to personal/agent class, and global. Trust indicators on social networks are not considered explicitly in their article. Nevertheless, these can be classified with the respective categories. Regarding the centrality, reviewing comments on Facebook cannot be attributed to either of both classes explicitly. If there is a review without any technical tag to which agent it concerns, there is no central node that aggregates these opinions. If there is a tag (e.g. a post on a public page of a web shop), there is a central link to all other comments. On public pages, these reviews are also available for everyone in the network. But if someone is not looking for information explicitly, it may depend on the network proximity and similarity to the author whether or not the post is streamed to him or her.

According to Jøsang et al. (2007), reputation systems are based on a network that represents the way and the directions in which the reputation is distributed. Distributed systems have no central authority and participants have to collect their own information from other members. Such systems are also referred to peer-to-peer networks like discussion forums. In contrast to distributed systems, centralized systems collect the ratings from all users in a central place that can be accessed by others. Some systems allow their users to rate each other mutually; others add additional information from other sources. Such systems are called reputation computation engines, which are the subject of the following lines.

Reputation computation engines take information from private and public sources and calculate a reputation score that should represent the trustworthiness of the evaluated entity. The simplest idea is to subtract the sum of negative ratings from the sum of positive ratings and present the result as the reputation score. Probably the most used method is to compute the ratings of a shop and calculate an average reputation score. These models are often enhanced by different factors like the trustworthiness of the rater or the time bygone since the publication. More advanced models use complex algorithms in order to optimize the reliability of the reputation scores (Jøsang et al., 2007).

Cho, Kwon and Park (2009) differentiate two types of reputation systems: "user reputation generating mechanism" and "item rating aggregation mechanism" (Cho et al., 2009, p. 2). The first type can be split in two different approaches, depending on the directional orientation. In C2C models, bidirectional mechanisms are used that allow both parties to review or rate the others. In B2C relationships, the approach is unidirectional, meaning that only one side can evaluate the counterpart. Another classification base is the source of information. There are systems that aggregate reputation implicitly, e.g. by analyzing social networks. Other systems use explicit procedures where reviews and ratings are provided directly by the consumer - mostly voluntarily (Cho et al., 2009). In the course of time, many studies have been conducted on the credibility of reviews, ratings and raters, striving to increase reputation system reliability and make them more resistant against fraudulent purposes (e.g. Cho et al., 2009). Basic systems are designed to aggregate simple evaluations from consumers and make them accessible for interested consumers. Cho et al. (2009) created their source-credibility-theory-based Q-rater recommender, a mechanism for B2C models with unidirectional ratings that is based on the past ratings of users. This adds an implicit component to the model. Applying measurements for expertise, trustworthiness and co-orientation (a measure for the similarity between rater and information seeker), their method evaluates raters' characteristics in order to find the most reliable reputation sources. Based on a group of qualified raters and their traits, the system estimates a rating for specific items and proposes it to the seeker. Thus, the amount of ratings should be limited but best suited to the interests of the seeker. A classification of the different models is given in Table 3.

**Table 3: Classification of user reputation mechanisms. Based on Cho et al. (2009).**

| Business model | Information Source | |
|---|---|---|
| | Explicit | Implicit |
| **C2C (Bidirectional)** | Online auction site (Ebay)<br>P2P services | Social network analysis |
| **B2C (Unidirectional)** | Online shopping (Amazon)<br>Online evaluation (Epinion.com, Trustpilot.com) | Proposed "Q-rater" model by Cho et al. (2009) |

The presented model has two major implications for the present study. First, it shows how different information can be integrated in one reputation system. In this case, ratings about other users and information about the reliability of other user's ratings. Second, the work of Cho et al. (2009) introduces implicit information sources like social network analyses. In their Q-rater model, they use an implicit approach by adding implicit information like the similarity between informant and information seeker or the expertise of the informant. In the present study, the value of implicit Facebook reviews and ratings should be evaluated.

## 2.3 Facebook Features for Trust Building in E-commerce

Reviews are published on different channels like company websites, social networks and product review sites (Fang, 2014; Kreis & Gottschalk, 2015). While many studies deal with general reviews and other trust indicators, not much is known about Facebook features and their use for trust creation in e-commerce. Facebook features can be seen as a technical tool on Facebook that facilitates multiple activities. Every year, new features are implemented to Facebook (Celebi, 2015). Current features are Likes, Comments, Wall, Friend, News Feed, Message, Photo and Chat (Lee, Kim, & Ahn, 2014). Less popular features are Check in or Facebook Places. This section presents literature that deals with Facebook features in the broader sense of e-commerce.

Ahmed and Abulaish's study about spam detection in social media provides a good overview and description of Facebook. A major feature are "wall posts". A Facebook wall is the place where interactions are happening. People can leave massages, pictures or links on their wall or the wall of others, called wall posts. "Fan pages" are specific pages for celebrities or business owners. Owners of fan pages can share information with their followers on this site. "Tags" are a feature that allows the members to tag others in their posts. Such posts are also displayed automatically on the page of the tagged person or organization (Ahmed & Abulaish, 2013).

In a study about usefulness of social media in reputation management, Hong and Kim (2015) observed the relationship between a company's reputation and the interaction on its Facebook page. Using the theory of dialogic communication and the halo effect, the researchers could not find a direct correlation between the number of Facebook likes and a company's reputation, indicating that there is no halo effect in social media (Hong & Kim, 2015). Dialogic communication is a "negotiated exchange" of information (Kent & Taylor, 1998, p. 325). Halo effect means that individual attributes are judged based on a "global evaluation" (Nisbett & Wilson, 1977, p. 250). In other words, someone is drawing conclusions about specific attributes from his evaluation of the whole. This effect is also present in brand images. People have a specific image about a brand, built on the limited information they have (Reynolds, 1965).

Phelan, Chen and Haney (2013) studied the extent to that hotels use social networks as a marketing tool by conducting a content analysis on their Facebook pages. The interaction between hotels and customers was measured by the number of Facebook fans (likes), number of people talking about the hotel, amount of guests checking in in a hotel (using the check in feature), customer feedback on the Facebook page as well as replies on comments by the hotel administration. The results of the study where ambiguous. While some hotels use Facebook only for information sharing, others use the social network to engage consumers in interaction (Phelan, Chen, & Haney, 2013). This study shows very well the utility range of Facebook features in different industries.

Bushelow (2012) observed the relationship between Likes and brand loyalty and purchase intention but found no significant association (Bushelow, 2012). However, Zhang and Pennacchiotti (2013) suggested that brands which someone likes on Facebook are more likely to be bought by this person later (Zhang & Pennacchiotti, 2013). This relationship was tested on a large sample of ebay users.

Harris and Dennis (2011) observed how consumers incorporate recommendations and purchase experiences into social networks. Their study combined social networking and online shopping. In focus groups, the use of Facebook, especially in the context of providing and looking for recommendations in purchasing activities, was discussed. Furthermore, the relevance of different information sources and their influence on trust was discoursed, leading to the following findings: Very few participants use Facebook for information search purposes. Additionally, the participants had almost no interest in interacting with fan pages or people they do not know in reality. Nevertheless, the discussion revealed that many participants posted indirect or direct recommendations on their Facebook wall or asked friends for their recommendations in the social network. The major finding was a hierarchical order of trust sources. First, the participants trust real friends, then Facebook friends. After that, they trust expert blogs before independent review sites. Reviews on a merchants website were the least trustworthy recommendation source. Furthermore, the authors mention that the like button is a quick way to share interests with friends in the network. Nevertheless, the findings indicate that the like button is rather used for someone's own benefit instead of recommending the respective brand (Harris & Dennis, 2011).

An experiment conducted by Wang and Chang (2012) examined the effect of social ties in Facebook and product risks on the purchase intention. They revealed that product information and recommendations provided by strong-tie friends increase the probability of purchasing due to the higher level of diagnosticity. Diagnosticity is seen as the consumers' perception of the information source's capacity to help them in the judgment of quality. In other words, it can be seen as the capacity and helpfulness of recommendations. Thus, this study supports the findings of Harris and Dennis (2011), arguing that friend's recommendations have the highest influence on trust. However, their study was limited by the experimental design and validity.

This section presented a few studies about Facebook features and their relationship to a firm's reputation. However, no relevant studies were found about the influence of Facebook comments on online trust. Including Facebook features in trust-building models could be beneficial for research and practice.

## 2.4 Sentiment Analysis

Sentiment analysis aims to analyze sentiments, opinions, attitudes and emotions. Thus, it can be used in order to understand and interpret the vast amount of data in social media. Within this thesis, sentiment analysis is used **to interpret the differences in comments on Facebook and reviews on review sites.** The focus is on reviews on web-shops.

However, sentiment mining can also be applied to all kinds of opinions towards topics, products, individuals, organizations and services. Liu (2010) distinguishes two types of textual information which are facts and opinions. While information retrieval techniques focus on mining and processing more objective data and factual information, sentiment analysis methods assess opinions, sentiments, attitudes and emotions. Understanding the different concepts of opinions, sentiments and emotions is crucial for understanding the functionality of different sentiment analysis applications. While facts are objective measures of entities, **an opinion is a positive or negative sentiment, view, attitude, emotion, or appraisal about something or someone**. This can be, for instance, a product or a service as well as specific aspects of the same. There are regular and comparative opinions. Regular opinions are either direct or indirect, while comparative opinions are based on a comparison of different features or aspects. Furthermore, opinions can be distinguished in **implicit and explicit** (Serrano-Guerrero, Olivas, Romero, & Herrera-Viedma, 2015)**.** An explicit opinion expresses someone's subjective view while an implicit opinion is wrapped in an objective fact. Liu (2010, p. 7) provides an example of an implicit opinion with the sentence: "the earphone broke in two days". This is a factual statement, containing the opinion that the earphones are bad quality. According to Liu, subjective sentences express personal feelings, views, or beliefs but not necessarily a sentiment (Liu, 2010). In contrast, objective statements bear factual information. Even though it is not necessary, subjective sentences can also include sentiments, especially in case of judgments or agreements (Serrano-Guerrero et al., 2015).

The last concept that should be introduced is the one of emotions. Serrano-Guerrero et al. (2015) argue that emotions are the expression of one's subjective feelings and thoughts. Emotions are love, joy, surprise and anger. Emotions are closely related to opinions, which are often measured by divers emotions.

To sum it up, an opinion can be a sentiment, an attitude or an appraisal of something or someone. In contrast, subjectivity expresses feelings, beliefs or emotions. This work focuses on opinion expressions that contain positive or negative sentiments.

There are multiple tasks of sentiment analysis. These are sentiment classification, subjectivity classification, opinion summarization, opinion retrieval, detection of irony or sarcasm and opinion spam detection. In the present work, the applied function is sentiment classification. This is also known under the terms "opinion orientation" or "sentiment polarity". It is best described as the evaluation of a statement in order to predict the sentiment of a person towards something or someone. The basic classification of sentiments differentiates the categories of negative, neutral and positive. Nevertheless, sentiments can be measured on different scales, also called polarity rating. These scales can range from -1 to 1 or a scale from 1 to 5, as often used in online ratings.

While research on text interpretation has a longer history, sentiment mining and opinion mining research is booming since 2001. Several factors for the high interest in sentiment analysis are the rise of machine learning, the availability of large datasets as well as the challenges of the area and the usability in intelligence application (Pang & Lee, 2008).

Sentiment mining has lots of different applications. One of them is in the collection of information from different sources across the Internet in order to aggregate and summarize reviews and opinions about products and services. Another application is the use in review related websites, where ratings can be assigned upon the review text. But sentiment analysis is also used within recommendation systems. Sentiment scores can be helpful in order to eliminate items that bear inconsistencies between review content and the respective rating. Pang and Lee (2008) mention the use of sentiment mining for the correction of user ratings. They state that there are reviews that indicate a much higher rating than the one given by the author. Furthermore, there might be ratings that are biased which can be adjusted by the use of sentiment analysis. Another field is online marketing. Sentiment mining application can indicate when to display adds and when not, based on the sentiment of the website content. Nevertheless, tapping what is generally said about products and services might be the most important task of sentiment analysis (Pang & Lee, 2008; Serrano-Guerrero et al., 2015).

A lot of studies employing sentiment mining come from the information service field. There are several studies that use sentiment mining in order to predict the future. For instance, a model by Asur and Huberman (2010) predicts box-office revenues for movies based on the rate of tweets about the movies. They found a strong correlation between the attention a topic is given online and its success in the future (in case of the mentioned study a forthcoming movie and its later ranking). In the same study, the authors revealed the efficacy of sentiment analysis, improving the prediction after a movie was released (Asur & Huberman, 2010). Martin-Valdivia et al. (2013) developed a sentiment polarity detector based on different approaches. Their classifier included reviews from several languages as well as different classification methodologies. They used a set of Spanish movie reviews, which were translated into English. Both datasets, the Spanish and English, were assessed differently, the English data by supervised and unsupervised systems, the Spanish data only by supervised systems. Through the combination of different methods, the classification system outperformed the individual models when used alone (Martín-Valdivia, Martínez-Cámara, Perea-Ortega, & Ureña-López, 2013). A study by Bollen, Mao and Zeng (2010) derived a prediction model for stock market prices based on the Twitter mood. Based on the fact that emotions can affect decision-making and behavior, they measured the mood states of a large Twitter dataset in order to find relationships to the stock market prices. Starting from the assumption that stock market prices are rather influenced by news than by price movements, they used a timeframe that included a presidential election as well as thanksgiving in order to capture significant effects on public mood. The accuracy of their model was 86.7% (Bollen & Mao, 2011).

There are several tools for the execution of sentiment analyses. Popular ones are the freeware Social Mention or the software from Lexalytics, called Semantria. After taking all advantages and disadvantages into consideration, Semantria was used in this thesis. One of the major disadvantages was the fact that a license for Semantria is very costly. Using a 1-month free trial version, this disadvantage could be avoided. The crucial advantages were the relatively low restrictions in the trial version, the excel add-in, the categorization function, the ease of use of the tool as well as prior experiences with the software. The decision was supported by the comparative analysis of sentiment analysis tools on different datasets from Serrano-Guerrero et al. (2015) where Semantria showed a good performance in sentiment classification and polarity rating compared to considerable alternatives. However, the quality of sentiment analysis results depends highly on the quality of the input data. For this thesis, the input data is gathered from public websites. Crawling information from such sites can be complex. Often, the data is unstructured or in an unfavorable format. Furthermore, and especially in the case of this thesis, there is additional data that is not usable and has to be removed in advance (Chaovalit & Zhou, 2005). Facebook is full of information, however, not all content is about products or shopping experiences. Thus, data must be selected carefully and prepared for the following observation.

## 2.5 Hypotheses

Throughout the Internet, there are different places where consumers can publish word of mouth. Kreis and Gottschalk (2015) differentiate two elementary types of channels: channels that address content gratification and channels that offer social and process gratifications. Hereby, social networking pages appeared to provide more social and process gratification. On the other hand, product review pages seem to offer more content gratification. Therefore, there might be a difference in the usage of the two platforms. In their hypotheses, they state that consumers who strive to gain content related gratification should prefer the respective channels like review sites. In their opinion, content-related channels "facilitate serious, fact-driven argumentation and assist the provision of high-quality content that makes it easy to assist the firm or other consumers" (Kreis & Gottschalk, 2015, p. 416). Others should prefer channels with more social and process gratifications such as social networks like Facebook. Those are better to communicate with each other, for self-portraiture and other personal benefits. Due to the different motives, review content on Facebook might be much more emotional and less factual than reviews on review sites since their motives are for instance "venting negative feelings" and "maximize personal advantages" (Hennig-Thurau, Gwinner, Walsh, & Gremler, 2004; Kreis & Gottschalk, 2015). For that reason, Facebook reviews might be formulated more extreme in terms of their sentiment polarity. Furthermore, it can be assumed that reviews on the social network Facebook are on average more negative as the ones on review sites. These assumptions are leading to two hypotheses about the first sub-question:

**H1: Review sentiment on Facebook is more negative than the review sentiment on review sites.**

**H2: Review sentiment on Facebook is more extreme than review sentiment on review sites.**

To **answer the first sub-question**, reviews on the two platforms are compared and observed for inconsistencies in their polarity. Two different platforms will be subject of this study: a review site and the social network Facebook. The first hypothesis (H1) is verified once the average ratings and the average sentiment values differ significantly from each other. In the best case, a consistent difference in the sentiment is detected, meaning that one of the both sources is in most cases better or worse than the other, which would lead to a confirmation or rejection of H1. In order to test H2, standard deviations and means of review sentiments are compared to each other. It is expected that the standard deviation of Facebook reviews is higher, due to a wider spread of the sentiment scores.

**The hypothesis for the second sub-question** (which sources are more important in inconsistent situations?) are based on the assumption that the different motives to publish word of mouth on a certain platform lead to a different attitude towards the content on these platforms. Therefore, the relevance of the divers platforms is assumed to depend on the personal opinion about these platforms. This is expected to come into effect in situations were the consumers face inconsistent information on different platforms. Many studies in the information system field deal with the impact of reviews on the purchase intention of the consumer (Zhang et al., 2014). A study by Park and Lee (2009) revealed that negative reviews have a stronger impact on consumer's choice than positive reviews (Park & Lee, 2009). Adding the findings by Lee, Park and Han (2008), this effect can be traced back to the fact that negative statements are more diagnostic as well as informative (Lee et al., 2008). This is also supported by impression formation research, indicating that negative information gets more emphasis than positive ones (Simon, 2001)**.** Other studies point out that even negative reviews can result in positive outcomes through improvements in product awareness and subsequently lead to higher sales figures (Berger et al., 2010). Another study revealed that inconsistent reviews, meaning a mixture of positive and negative reviews, has an indirectly positive influence on the purchase intention (Zhang et al., 2014). The results of a market research by Smith and Anderson (2016) indicate that highly negative reviews have more influence than highly positive ones, supporting the findings of Park and Lee (2009). Nevertheless, the findings indicate that there are opposing behaviors. Some pay more attention to the positive reviews, some to the negative ones (Smith & Anderson, 2016).

As presented, many studies observed the impact of positive and negative reviews. Nevertheless, no study was found that observed the differences in the impact of **different review channels** on the purchase intention of consumers. It is unknown, whether and to what extent the consumers listen to opinions published on Facebook. Since more and more reviews are published on the social network, it is assumed that Facebook has reached some recognition as review source. Due to the implicit character of Facebook as a review channel, the impact is expected to be minor than the on of review sites. Thus, two further hypotheses are added:

**H3: For consumers, the social network Facebook is an important information source in purchase decisions.**

**H4: Facebook reviews have in general less impact on the purchase intention than reviews on review sites.**

In combination with the results from research question 1 about inconsistencies between Facebook comments and review sites, knowledge about variances in the impact can lead to meaningful implications for researchers and practitioners.

# 3 Research Design

This chapter outlines the research process and all relevant methods that are used in order to find answers for the previously stated research questions. Due to the fact that the two sub-questions are not conjunct, the research design sections as well as the results are split in two parts. Later, in the discussion section as well as the theoretical and practical implications, the findings from both parts are combined. The aim of **part 1** is to examine **differences in the review content between Facebook and the review site Trustpilot. Part 2** aims to reveal **the varying impact of review sources on the purchase intention under inconsistent information.**

## 3.1 Researched differences in the Review Sentiment on Facebook and Trustpilot

### 3.1.1 Scope

The review site under observation is called "Trustpilot". It was chosen since it is one of the most popular review sites in Europe and provides a solid amount of reviews on most of the online retailers. Other review sites that were considered did not deliver a similar amount of reviews. Facebook was chosen since it is the biggest online social network in the world. The amount of data that is generated in the network is tremendous. A preceding observation of different social networks and comparable social media revealed that Facebook has already reached some sort of review site status. Comparable platforms (like Twitter) showed insufficient amounts of review-like content and were therefore excluded from this study.

In total, 20 online retailers ware observed, which was expected to deliver enough data for a reliable comparison. Considering the two subgroups per shop, namely review site reviews on the one hand and Facebook comments on the other, the combined sample size is 40. Furthermore, up to 100 reviews and comments were extracted for each shop to conduct tests about differences and similarities between the two review sources, resulting in a much larger sample size on the shop level. The shops under observation were all from the UK. Due to the fact that Semantria was firstly developed for English language and since this study is written in English, it seemed most convenient to use shops from an Anglophone country. It was also assumed that shops have their main market in the country where they originated. As such, they might get the most reviews from within this country. Several concerns appeared about the domain and the size of the shops of subject. It was expected that the core business of online retailers is the distribution of goods. Even if there are retailers who have their own product lines, their main business is in service provision. Quality features like delivery time, reliability and service provision might be in common for all shops, no matter which domain they are operating in. In terms of size, it was expected that big retailers have more reviews than their smaller peers. Thus, medium size to big shops that have a certain kind of online reputation were chosen for this study. Also noteworthy is that only pure online retailers were subject of this study. These are shops that have no physical salesrooms (there might be flagship stores in certain cases, but the main business of the observed companies is in online retailing). It should be avoided that reviews about experiences with physical stores are among the data set. Furthermore, it ensures a better comparability between the chosen retailers. To ensure that the chosen shops provide a sufficient amount of data on both platforms, the availability of comments on Facebook and reviews on the respective review sites were examined manually in advance (see Appendix B). A list of all shops that were subject of this thesis can be found in Appendix A. The shop selection was based on a list of the top pure-play ecommerce retailers in the UK (Internet Retailing, 2016). It should be noticed that the sample of shops does not represent the population of online retailers. Nevertheless, it is not considered as relevant since the association between the sentiments of different review channels is not assumed to be influenced by the domain or the size of the retailer. In other words, the main focus is on the differences between the two sources and their impact on consumers' purchase behavior. A segmentation in domains (e.g. in clothing store and electronics store) or in size categories was not intended since the differences are assumed to originate in the characteristics of the review channel and not of the shop.

### 3.1.2 Data Collection and Processing

Data from the review site Trustpilot was extracted through the web tool "import.io"[1]. With import.io, data can be gathered automatically resulting in massive savings of time. Once the webpage from which data should be extracted is inserted, the system provides several elements of the chosen webpage. In this case, different elements like the review content, the star rating as well as the publication date of the single reviews were collected. Furthermore, a direct link to the review was stored in order to have a direct access to the single sources in case of complications. In some cases information is hidden due to the structure of the webpage. In such cases HTML-codes can be selected and extracted. In the concrete case of this study, the star rating was presented in different pictures from 1 to 5 stars, which were not automatically detected by the service. By extracting the link to the picture, the respective rating could be identified. The free version of import.io allows to extract up to 500 single pages per month. In a first step 100 reviews per shop were extracted in the browser application. For this, the import.io URL generator was used. Since the web-service extracts data page after page, a URL generator allows adding several pages at once. Trustpilot presents 20 reviews per page. Thus, 5 pages were necessary to extract 100 reviews. Trustpilot web links are structured in the following way:

https://www.trustpilot.com/review/**shopname**?page=**1**

Determining a set of numbers for the end of the link allows adding several links at once and a simple way to extract data from multiple sites in one step. Afterwards, the Data can be downloaded in CSV format. The data was processed in Excel and prepared for the reprocessing through Semantria.

There are several methods to gather data from Facebook. So-called APIs (application programming interface) enable the interaction between web applications. Via Facebook graph API, applications can receive and send data directly to the social network. For this thesis, Facebook's graph API explorer was used.[2] Within this explorer, several tasks can be chosen. One can either get, post or delete certain content via API. First of all, one of the three tasks must be chosen. Then, a specific API version must be picked. In this case, version 2.4 was chosen since the data output was presented in the most appropriate way. Different versions are due to the continuous changes in Facebook.

---

[1] www.import.io

[2] For more information go to https://developers.facebook.com/docs/graph-api?locale=en_US

In Facebook's social graph (the presentation of information) there are nodes, edges and fields. Nodes are things that are presented in Facebook such as users, pictures and pages. Edges are connections between these things like comments or wall posts on a specific page. Fields contain information about these things such as the name or the birth date of a user. In the present study, the node is the Facebook ID of the single shops. The edges are the wall posts by consumers on the respective Facebook pages. These are retrieved by the command "tagged". Fields are not used in the present API call. At the end of the API call, a user token must be inserted. This token is used to identify and authenticate the user who is requesting.

Concluding, the following API-call was applied:

https://graph.facebook.com/v2.4/**Facebook-ID**/tagged?access_token=XXX

Inserting the link, the browser presents the requested data. A sample of the 100 most recent comments per page was requested, copied to an Excel file and processed to proper format for further observations. A critical point is the identification of comments with review character. There are several intentions for Facebook comments. Not every comment left on a company wall is necessarily a review. There are also informative links, advertisements and entertainment posts - to mention just a view. Thus, it was necessary to extract reviews from the vast mass of comments. By doing that, other comments could be eliminated from the sample. Semantria allows to assign documents to a certain category. Therefore, the user can determine groups per keywords. Whenever a text contains one of these keywords, the document is assigned to the respective category. In this study, keywords that are associated with online shopping were used. Among these were keywords like "order", "received", "bought" and many more. A manual check of 100 reviews showed that the chosen keywords selected reviewing comments to an adequate extent. The result was a table of reviews from Facebook and Trustpilot with the respective sentiment value.

In order to examine Semantria's applicability to evaluate reviews, a preliminary test was conducted. Therefore, the reviews on Trustpilot were compared to the respective star rating (1-5). It was expected that the rating given by the review author would be highly correlated to the sentiment of the respective textual review. A Pearson correlation test was conducted in order to measure the strength and the direction of the relationship between the variable "star rating" and "sentiment". Person's correlation was used because all requirements were met: the two variables are continuous and paired. A scatterplot graph showed linearity between the two variables as well as only a few outliers. Those were neglected due to the sample size. Both variables were normally distributed as assessed by a visual inspection of Normal Q-Q plots. **The association was assumed to be positive linear and higher than 0.5.**

**Table 4: Correlations between star rating and sentiment score per shop.**

| Shop | Correlation between star rating and sentiment score* | Shop | Correlation between star rating and sentiment score* |
|---|---|---|---|
| ao.com | 0.554 | laredoute.co.uk | 0.778 |
| bonprix.co.uk | 0.546 | lightinthebox.com | 0.581 |
| boohoo.com | 0.563 | littlewoods.com | 0.567 |
| jdwilliams.co.uk | 0.620 | marisota.co.uk | 0.568 |
| ebuyer.com | 0.662 | mobilefun.co.uk | 0.662 |
| firebox.com | 0.503 | moonpig.com/uk | 0.589 |
| getthelabel.com | 0.736 | notonthehighstreet.com | 0.714 |
| gettingpersonal.co.uk | 0.740 | photobox.co.uk | 0.364 |
| justfab.co.uk | 0.744 | viking-direct.co.uk | 0.754 |
| kitbag.com | 0.671 | zalando.co.uk | 0.848 |

* All correlations are significant at the 0,01 level (2-tailed).

**Table 5: Correlation between star ratings and sentiment scores.**

| | | Trustpilot stars | Semantria sentiment |
|---|---|---|---|
| Trustpilot stars | Pearson Correlation | 1 | 0.741** |
| | Sig. (2-tailed) | | 0.000 |
| | N | 1909 | 1908 |
| Semantria sentiment | Pearson Correlation | 0.741** | 1 |
| | Sig. (2-tailed) | 0.000 | |
| | N | 1908 | 1908 |

** Correlation is significant at the 0.01 level (2-tailed).

As expected, there is a strong correlation between the sentiment of a review and the star rating a customer assigned to his or her evaluation. This correlation could be found on the individual level as well as in the total set of reviews (see Table 4 and Table 5). On the one hand, these results show that Semantria is capable to assess reviews appropriately. Beyond that, **this score shows that Semantria can be used to estimate the sentiment polarity of reviews on Facebook**. There is no such star rating in comments on public Facebook walls of companies. Thus, a valid tool is required to make these reviews accessible. In order to make these reviews comparable and computable, a numeric value should be assigned to them. As the above correlation analysis revealed, Semantria can be used to do this.

During the analyses another noteworthy relationship was found (see Table 6). The distribution of the sentiment means for every rating category (1-5 stars) showed that 3 star ratings had a sentiment mean of about zero, meaning that 3 stars seems to be neither negative nor positive in the mind of the customer, while 1 or 2 stars seem to be negative and 4 and 5 stars are positively assigned. While the latter fact seems logical, the others are not necessarily. Star ratings are difficult to assess. A one star restaurant is already of good quality. So why would someone assign a star to a service he is really disappointed about? Most review sites call for a minimum of one star. Thus, the results indicate that there is a clear distinction between the values of the star rating. Another interesting fact is that 4 and 5 star ratings show a much more positive review sentiment than the 1 and 2 star ratings a negative. A possible explanation could be that 4 and 5 stars are only awarded in cases of absolute satisfaction while lower ratings are already given when the consumer is only slightly disappointed.

**Table 6: Descriptive statistics of different star ratings and the respective sentiment means.**

| Descriptive Statistics | | | | | | |
|---|---|---|---|---|---|---|
| **Stars** | | **N** | **Minimum** | **Maximum** | **Mean** | **Std. Deviation** |
| 1 | Semantria sentiment | 658 | -1.525 | 0.976 | -0.0251 | 0.336 |
| | Valid N (listwise) | 658 | | | | |
| 2 | Semantria sentiment | 122 | -1.032 | 0.550 | -0.144 | 0.307 |
| | Valid N (listwise) | 122 | | | | |
| 3 | Semantria sentiment | 72 | -1.067 | 0.761 | 0.008 | 0.395 |
| | Valid N (listwise) | 72 | | | | |
| 4 | Semantria sentiment | 185 | -0.649 | 1.600 | 0.349 | 0.299 |
| | Valid N (listwise) | 185 | | | | |
| 5 | Semantria sentiment | 871 | -0.698 | 2.349 | 0.478 | 0.268 |
| | Valid N (listwise) | 871 | | | | |

The minimum and maximum values represent the minimum and maximum sentiment scores per rating group (1-5). This is an indication, that there are severe outliers within the groups. Taking a closer look to row 1, the maximum sentiment score for a review with a 1-star rating was 0.976. Compared to the second and third line, the maximum score is even higher than the respective values from following lines. Therefore, there might be errors in this evaluation. There are two potential reasons for that: Either, Semantria had issues to evaluate this review correctly and assigned a much higher value or the author of the respective review gave only one star although he or she wrote a very positive review. Interpretation errors are a major issue of sentiment mining. However, also incorrect ratings are a well-known problem in reputation research.

### 3.1.3 Analysis and Research Methods

To stress H1, a t-test was chosen to examine the named association between the two platforms. An independent t-test compares the mean of the same variable in two independent groups. Furthermore, the t-test calculates the significance for the found difference between the means, which is the probability that the means are equal even though a difference was found. In order to find out whether the samples meet the requirements of the t-test method, the samples were tested for outliers and normal distribution in advance. There were single outliers, but they were neglected due to the big sample size. As assessed by a visual inspection of normal Q-Q plots, both review channel sentiment scores deviate slightly from a normal distribution. First, the t-test was run for all reviews from both channels. To test H2 the sample was grouped into the single shops in order to show differences between the sub-samples. At the end, a further t-test was conducted in order to compare the standard deviations between the two review channels and to stress hypothesis 2.

## 3.2 Researched Impact of Review Sources on Purchase Intentions under Inconsistent Information

### 3.2.1 Scope

The second part of the present thesis is a survey. The participants of the survey were students. Referring to Zhang et al. (2014), several studies are evidence that online consumers are often young adults who have been enrolled in an university for at least a year (Kim, Ferrin, & Raghav Rao, 2009; Lim, Sia, Lee, & Benbasat, 2006). Furthermore, students are considered to be good representatives of online shoppers and their decision-making patterns. In order to reach a sufficient mass of data, a minimum of 100 participants was intended. Since the sample was not expected to be split in different groups, the sample size was enough to gather significant results**.** For the sample selection, convenience sampling was chosen. According to Saunders, Lewis and Thornhill (2009), convenience sampling (also called haphazard sampling) is especially appropriate in cases where a representation of the total population is not necessarily required. This is mostly the case when the population shows very little variations. In this thesis, online shoppers are represented by students, due to the similar characteristics. Thus, a representation of the total population of students was not pursued. A broad selection of students was expected to represent the general behavioral characteristics of online consumers to a high extent. However, Saunders et al. suggest that findings within such samples should be reconsidered by more structures samples (Saunders, Lewis, & Thornhill, 2009).

The survey was distributed per personal invitation via E-mail and Facebook. Thus, mostly friends and fellow students participated at this study. In the end, 103 students participated at the study. Nevertheless, some cases were eliminated due to missing data in the conjoint analysis. The final data set contained 95 cases.

### 3.2.2 Data Collection

The survey was developed on the platform Qualtrics. The faculty of Behavioral, Management and Social Sciences at the University of Twente provides free access to the tool. Qualtrics is a self-explanatory tool, which provides a multitude of data collection methods and a direct SPSS data export. Following the approach of Zhang, Cheung and Lee (2014), several questions about characteristics of the participants were asked in the beginning of the survey. These questions regarded mainly the demographic structure of the sample as well as the online shopping behavior and Facebook consumption of the participants. Based on these, students that have little or no experience with online shopping or online reviews could be eliminated from the sample. The results are stated in table 7.

Table 7: Characteristics of the participants and e-commerce behavior.

| **Characteristic** | |
|---|---|
| **Gender** | |
| Female | 61% |
| Male | 39% |
| **Age** | |
| Mean | 24.9 |
| Standard deviation | 3.4 |
| **How much experience do you have with online shopping*** | |
| Mean | 5.5 |
| Standard deviation | 1.5 |
| **Have you bought anything online in the last month** | |
| Yes | 96% |
| No | 4% |
| **Have you bought anything online in the last year** | |
| Yes | 100% |
| No | - |
| **Do you use Facebook** | |
| Yes | 99% |
| No | 1% |
| **Did you seek or search for other consumers' reviews or comments online regarding what you wanted to buy?** | |
| Yes | 88% |
| No | 12% |
| **How many different channels/platforms do you use when you seek information about shops?** | |
| None | 4% |
| Mostly only one | 26% |
| 2-3 | **59%** |
| More than 3 | 11% |
| | |
| *measured on a 7 point Likert scale from 1 = "not at all" to 7 = "a lot" | |
| **N=95** | |

More females participated at this survey. Nevertheless, the differences are not severe since a group analysis based on gender was not intended. The average age of the students was 25 years. Compared to other studies, the average age of the sample was high. However, a full representation of the student population was not pursued. Furthermore, the average age was expected to play a minor role in online shopping behavior. All but one of the participants use Facebook. Most of the participants spent between 15 and 60 minutes per day on Facebook (see Figure 2). With regard to online shopping behavior, all of the participants indicated that they have bought something online in the last month or at least in the last year. Thus, it can be assumed, that the sample has a lot of Internet shopping experiences. This is supported by the mean of 5.5 out of 7 on a Likert scale about online shopping experience, as evaluated by the participants. About 90% of the participants use other consumer's reviews in their online shopping processes. It is interesting that about 60% use between two and three different sources when they seek for information about an online shop.

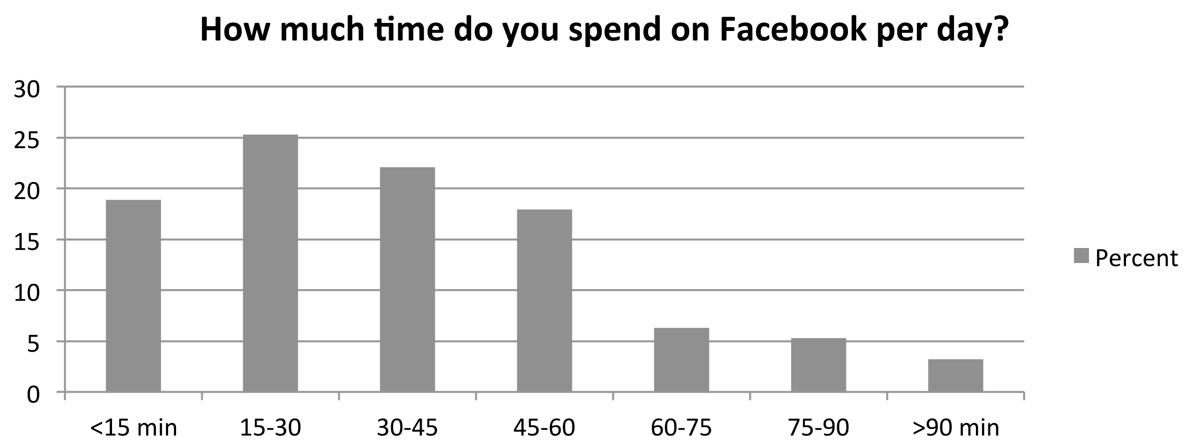


**Figure 2: Daily Facebook consumption.**

In the following questions, the importance of different trust antecedents and sources was examined in order to test the hypotheses H3 and H4. First, the participants had to group the single trust antecedents from the list on page 12 into "more important" and "less important". The item "social network information" was added in order to reveal the importance of this trust indicator compared to other indicators. On a 7-point Likert-scale, the consumers were asked to express the level of agreement on statements about the importance of Facebook in purchase decisions and the quality of Facebook reviews. Furthermore, the participants were asked whether they have ever gone to Facebook with the intent to gather information about a shop and to support their purchase decision. Students who agreed where further bidden to explain their reasons to do that and the advantages and disadvantages. This should make it possible to draw conclusions about motives and the current usage of Facebook in trust building.

The main part of the survey, the conjoint analysis, will be explained in the next section.

### 3.2.3 Conjoint Analysis

Conjoint analysis is a method that is widely used in market research. Its popularity is easy to explain: Every good that is produced can be seen as a combination of different attributes, which can be product features, functions or benefits. For instance, attributes of smartphones are display size, battery life and weight. Consumers value these attributes differently, according to their preferences and needs. For one consumer display size might be the decisive factor, for the other battery life. The single attributes can be broken down in different attribute-levels – for instance in 4-inch display, 4.7-inch display or 5.5-inch display. A combination of single attribute-levels gives a product profile (e.g. a phone with 4.7-inch display, 48 hours battery life and 300g weight). The participants of conjoint analysis are asked to state their preference for every possible combination of given attributes (or a statistically reduced set of combinations). Here, they must weigh of the advantages and disadvantages and make a trade-off between the different combinations. Participants must consider jointly and look at the combinations from a holistic view. Conjoint analysis provides several methods how consumers can express their preference. Participants can rank or rate the alternatives according to their preference. Another method is a paired comparison between two combinations per comparison. All methods lead to the same results: they measure the impact of the single attributes and the effect of the respective attribute-levels on the consumer's choice. Several conclusions can be drawn from the results. On the one hand, the importance of the single product features (attributes) can be determined. For instance, whether smartphone display size is a major decision criterion or not (or more important as battery life). On the other hand, it can be observed whether bigger or smaller screens are more preferred. However, it can be assumed that preferences for some attribute-levels are pre-determined. For instance, longer battery life is certainly always preferred over short battery life. Nevertheless, conjoint analysis could evaluate how much the consumer is willing to pay for longer battery life by adding the attribute price. Based on such knowledge, companies can develop products or services according to the preferences of the consumers and thereby outperform competition (Klein, 2002).

But the method of conjoint analysis is not only applicable in market research. Conjoint analysis can be applied whenever preferences for objects with multiple attributes are under observation**.** However, the method is not very popular in social sciences, even though there might be multiple application fields (Klein, 2002).

In online shopping, consumers face information from different sources, which could present inconsistent information. **Therefore, a combination of different attributes and attribute-levels is given and the consumer must decide which source to trust more.** While other methods often imply abstract values (see for example the study by McKnight et al. 2002, outlined in chapter 2.2.1), conjoint analysis is particularly suited for the evaluation of realistic scenarios. **Exposing the participants to the scenarios of inconsistent information between different sources forces them to weigh the single attributes (the different review sources) and make decisions upon these weightings. Based on these decisions, the importance of the different attributes (the different review sources) can be assessed** (Klein, 2002).

Transferred to the present thesis, the **attributes are the information on review sites and information on Facebook**. Attribute-levels are the information directions on these sources, which can be negative, positive or neutral. As an example, a consumer could read negative comments about a shop on Facebook, while the average rating on a review site is positive. Now, the consumer has to decide which side he trusts more. **Providing different combinations of positive and negative reviews from both sources simulates the scenario of inconsistent information.** The attributes and attribute-levels are outlined in table 8.

**Table 8: Attributes and attribute-levels in the conjoint analysis.**

| Attribute | Facebook reviews | Review site rating |
|---|---|---|
| Attribute-levels | Predominantly negative comments | Negative rating (<2 stars) |
| | Mixed comments | Neutral rating (ca. 3 stars) |
| | Predominantly positive comments | Positive rating (>4 stars) |

Several conditions are required for the attributes and their attribute-levels. First, the attributes must be relevant, interfering and independent. Independent means that there is no overlapping with other attributes. For example, a small phone with a large display would not be possible, since the display requires a frame of similar size. Second, the attribute-levels must compensate each other in order to enable trade-offs and there may be no exclusion criteria that make trade-offs obsolete. All these requirements are met in the present study. Another criterion is that the attributes and the combinations are realistic. Since inconsistencies between different review sources are common, this last requirement for the attributes and the respective levels is also met (Baltes-Götz, 2006; Klein, 2002).

As mentioned before, every combination of attribute-levels must be observed. In order to compute the importance of attributes and values, a specific order of combinations must be adhered. SPSS provides a tool to create an orthogonal design, which contains all relevant combinations and the respective order. Using this function, the orthogonal design as presented in table 9 was created (Klein, 2002).

**Table 9: Orthogonal design of situations as determined by SPSS**

| Situation | Facebook | Review Site |
|---|---|---|
| 1 | Negative | Neutral |
| 2 | Negative | Positive |
| 3 | Neutral | Negative |
| 4 | Neutral | Positive |
| 5 | Positive | Negative |
| 6 | Neutral | Neutral |
| 7 | Positive | Positive |
| 8 | Negative | Negative |
| 9 | Positive | Neutral |

Within the survey, the participants were asked to imagine a situation where they intend to buy a product that is only available on a shop that is not familiar to them (initial trust situation). Further they should imagine that online information showed inconsistent information about the trustworthiness of the shop. Then, the different situations were presented. For instance, situation 2 was described as follows: “There is a positive rating of more than 4/5 stars on a review site and predominantly negative comments on Facebook about the shop. How likely would you buy the product at this shop?” The likelihood of buying from the respective shop was measured on a 7-point Likert-scale. **The ratings reveal the consumer’s personal weights for the single review sources.** To avoid confusion, the same phrasing was used for all situations as presented in table 9.

From a statistical perspective, the scientist describes the independent variables in the orthogonal design. Then, the participants determine the values of the dependent variables through their judgment. In general, so-called “utilities” for the single attribute-levels (4-inch, 4.7-inch or 5.5-inch screen) are calculated based upon these judgments. Negative utility values have a negative effect on the total utility of a product combination, positive ones a positive effect. As a second output, the importance of the attributes itself (display size, weight, battery life) is examined. Since the method is mostly used in market research, the results are presented as “utility” (e.g. in SPSS).

However, since negative reviews or comments are assumed to have always a negative impact on the purchase likelihood, the utility values for the single attribute-levels are not very informative in the present study. In other cases like the smartphone example, the utility values of different screen size options can have major implications for product developments. The relevant conclusions for this study are provided by the importance values, which are calculated for the single attributes (review site and Facebook). The higher the importance values of one source, the more rely the participants on this information source when inconsistencies occur. The results will be presented in section 4.2.

Since SPSS has no graphical interface for conjoint analysis, a syntax commando (see Figure 3) was formulated according to the guideline by Baltes-Götz (2006).

```
conjoint
plan         = ‘*file’
/data        = ‘*file’
/score       = ‘*variable1 to *variable
/factors     = ‘facebook (discrete more) ‘review_site (discrete more)
/print       = all.
```

**Figure 3: Conjoint analysis syntax**

The following lines describe the meaning of the different commandos:

**In the plan commando**, the orthogonal design file is determined. Thus, the conjoint syntax uses the design as the basis for the calculations.

**The data commando** coordinates in which file the collected data can be found.

**The score commando** presents the respective variables within the data file. In this case, 9 different combinations were subject to the observation. In conclusion, there must be 9 variables; one for each combination. Furthermore, the score commando shows that the rating method was used. Every variable contains the score that was assigned to the single combination of attribute-levels according on a 7-point Likert-scale. Thus, the variables contain numbers from 1 to 7. In the rating method, the preference for a product increases with growing numbers. Thus, most preferred combinations get higher numbers.

**The factor commando** describes the attribute types. Discrete means that the attribute is categorical. In this example, there are three categories for the respective rating: “positive”, “neutral” and “negative”. “More” shows the direction in which the factors are expected to be related to the ratings. This refers to the numeric value that was assigned to the different values during the creation of the design. While the numeric value -1 was assigned to “negative”, “positive” was assigned to the numeric value 1. By doing this, the different values were also centered which means that neutral values are expected to be 0.

**The commando print** simply indicates which results are printed in the output window. In this case, all results are presented. The results are presented in section 4.2.

# 4 Results

## 4.1 Differences in the Review Sentiment between Facebook and Trustpilot

The following two tables (Table 10 and Table 11) show the t-test results of all reviews on both channels. As expected, the mean sentiment is lower for Facebook reviews. Furthermore, the mean of Facebook reviews is slightly negative while the average sentiment of Trustpilot reviews is positive. The standard deviation of both groups (Facebook and Trustpilot) is similar. As can be seen in Table 11, there was no homogeneity of variances for the sentiment scores as presented in the results of Levene's test (p value small). Thus, equal variances cannot be assumed and the results need to be corrected. The corrected results are presented in the second row of Table 11. As this row shows, the sentiment mean of Facebook reviews is about 0.21 +- 0.02 lower than the sentiment mean of Trustpilot reviews. The associated p-value is $< 0,001$, indicating that the mean difference in sentiment scores between Facebook and Trustpilot is significant.

**Table 10: Group statistics for the sentiments of Facebook and Trustpilot reviews.**

| **Group Statistics** | | | | | |
|---|---|---|---|---|---|
| | **Source** | **N** | **Mean** | **Std. Deviation** | **Std. Error Mean** |
| DocumentSentiment | Facebook | 969 | -0.0568 | 0.4771 | 0.0153 |
| | Trustpilot | 967 | 0.1572 | 0.4615 | 0.0148 |

**Table 11: T-test results for the comparison between sentiments of Facebook and Trustpilot.**

| **Independent Samples Test** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **Levene's Test for Equality of Variances** | | **t-test for Equality of Means** | | | | | |
| | | **F** | **Sig.** | **t** | **df** | **Sig. (2-tailed)** | **Mean Differ-ence** | **Std. Error Difference** |
| DocumentSentiment | Equal variances assumed | 1.065 | 0.302 | -10.029 | 1934.000 | 0.000 | -0.2140 | 0.0213 |
| | Equal variances not assumed | | | -10.029 | 1932.118 | 0.000 | -0.2140 | 0.0213 |

The table on the following page (Table 12) shows the t-test results broken up into the single shops. **A comparison based on every single shop seems useful since the average means could balance out each other. A shop with a very positive rating could compensate the negative ratings of other shops and therefore influence the results.** 14 shops show the predicted association that Facebook reviews were more negative than Trustpilot reviews. It should be mentioned that the primary sample size was 100 reviews per channel for each shop. Due to a limitation on Facebook reviews, the sample size of all subgroups was adjusted in order to guarantee a reliable measurement. Differing sample sizes would have led to higher errors. Nevertheless, the samples were big enough to gain reliable results (minimum sample size was 20). Minor differences in single comparisons are due to missing values in the SPSS data. However, this procedure had no direct influence on the results. The complete table can be found in Appendix C.

Table 12: Overview of t-test results, grouped by shop name.

| T-test results | | | | | | |
|---|---|---|---|---|---|---|
| **Shopname** | **Source** | **Mean*** | **Mean dif.** | **Significance** | **Std. Deviation** | **Std. Error Mean** |
| AO | Facebook | 0.3805 | -0.0021 | 0.970 | 0.4141 | 0.0463 |
| | Trustpilot | 0.3825 | | | 0.2708 | 0.0303 |
| Bonprix | Facebook | 0.1738 | -0.1883 | 0.046 | 0.4133 | 0.0720 |
| | Trustpilot | 0.3621 | | | 0.3352 | 0.0583 |
| Boohoo | Facebook | **-0.1624** | 0.0602 | 0.400 | 0.3453 | 0.0442 |
| | Trustpilot | **-0.2226** | | | 0.4409 | 0.0555 |
| Ebuyer | Facebook | -0.1396 | -0.5338 | 0.000 | 0.3724 | 0.0455 |
| | Trustpilot | 0.3943 | | | 0.3475 | 0.0428 |
| Firebox | Facebook | 0.0426 | -0.5042 | 0.000 | 0.4185 | 0.0892 |
| | Trustpilot | 0.5469 | | | 0.2976 | 0.0634 |
| Getthelabel | Facebook | -0.0254 | -0.2875 | 0.001 | 0.3901 | 0.0526 |
| | Trustpilot | 0.2621 | | | 0.4719 | 0.0636 |
| Gettingpersonal | Facebook | -0.0012 | -0.0978 | 0.274 | 0.5716 | 0.0714 |
| | Trustpilot | 0.0965 | | | 0.4235 | 0.0529 |
| Jdwilliams | Facebook | -0.0395 | -0.0516 | 0.596 | 0.4805 | 0.0733 |
| | Trustpilot | 0.0121 | | | 0.4149 | 0.0633 |
| JustFab | Facebook | **-0.0398** | 0.0072 | 0.945 | 0.4475 | 0.0726 |
| | Trustpilot | **-0.0470** | | | 0.4623 | 0.0750 |
| Kitbag | Facebook | -0.0277 | -0.1460 | 0.088 | 0.4351 | 0.0603 |
| | Trustpilot | 0.1183 | | | 0.4289 | 0.0595 |
| Laredoute | Facebook | -0.1233 | -0.3728 | 0.000 | 0.4376 | 0.0539 |
| | Trustpilot | 0.2495 | | | 0.4853 | 0.0597 |
| Lightinthebox | Facebook | -0.2492 | -0.6204 | 0.000 | 0.4654 | 0.0755 |
| | Trustpilot | 0.3712 | | | 0.3932 | 0.0638 |
| Littlewoods | Facebook | -0.3629 | -0.1432 | 0.110 | 0.5133 | 0.0712 |
| | Trustpilot | -0.2197 | | | 0.3813 | 0.0529 |
| Marisota | Facebook | **-0.0137** | 0.1651 | 0.070 | 0.4185 | 0.0638 |
| | Trustpilot | **-0.1789** | | | 0.4107 | 0.0634 |
| Mobilefun | Facebook | 0.0053 | -0.2532 | 0.001 | 0.3776 | 0.0545 |
| | Trustpilot | 0.2584 | | | 0.3655 | 0.0528 |
| Moonpig | Facebook | -0.1882 | -0.1705 | 0.081 | 0.5002 | 0.0715 |
| | Trustpilot | -0.0176 | | | 0.4564 | 0.0652 |
| Notinthehighstr | Facebook | **-0.0662** | 0.0674 | 0.678 | 0.5513 | 0.1233 |
| | Trustpilot | **-0.1336** | | | 0.4634 | 0.1036 |
| Photobox | Facebook | -0.0020 | -0.5091 | 0.000 | 0.5279 | 0.0739 |
| | Trustpilot | 0.5071 | | | 0.3478 | 0.0487 |
| Viking | Facebook | -0.2005 | -0.4544 | 0.000 | 0.4361 | 0.0623 |
| | Trustpilot | 0.2539 | | | 0.4510 | 0.0644 |
| Zalando | Facebook | -0.2492 | -0.2805 | 0.006 | 0.4654 | 0.0755 |
| | Trustpilot | 0.0313 | | | 0.3720 | 0.0620 |

*The values labeled in red are the samples that do not meet the assumption of hypothesis 1, meaning that Facebook reviews were more positive in these cases.

Hypothesis 2 is supported by the standard deviations of the sentiments as shown in Table 13 and Table 14. The mean standard deviation of sentiments is 0.449 on Facebook and 0.401 on Trustpilot. According to the t-test, the standard deviation of the sentiment scores on Facebook is about 0.0481 +-0,0191 higher than the respective value of Trustpilot. Both groups are normally distributed and had no outliers, leading to valid results. This is showing that reviews on Facebook have a higher spread than reviews on the review site, supporting the hypothesis that review sentiments on Facebook are more extreme than those on Trustpilot.

**Table 13: Group statistics for the comparison of standard deviation means.**

| **Group Statistics** | | | | | |
|---|---|---|---|---|---|
| | **Source** | **N** | **Mean** | **Std. Deviation** | **Std. Error Mean** |
| Standard deviation | Facebook | 20 | 0.449 | 0.061 | 0.0136 |
| | Trustpilot | 20 | 0.401 | 0.060 | 0.0134 |

**Table 14: T-test results for the comparison of standard deviation means.**

| **Independent Samples Test** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **Levene's Test for Equality of Variances** | | **t-test for Equality of Means** | | | | | |
| | | **F** | **Sig.** | **t** | **df** | **Sig. (2-tailed)** | **Mean Difference** | **Std. Error Difference** |
| Standard deviation | Equal variances assumed | 0.021 | 0.886 | 2.516 | 38 | 0.016 | 0.0481 | 0.0191 |

## 4.2 Varying Impact of Review Sources on the Purchase Intention under Inconsistent Information

In the beginning of this section, descriptive reports about the opinion on Facebook as a trust indicator in e-commerce will be presented. The findings could help to evaluate the awareness about Facebook as a source of online reviews and the assessment of hypothesis 3. Later, the results of the conjoint analysis are outlined. About 35% of the participants agree that Facebook is a good indicator for a shop's trustworthiness. Nevertheless, most of them (in total 25%) only agree to a low level. 30% of the sample agree that Facebook reviews are mostly very negative. However, about the same amount disagreed on this statement and 35% did neither of both. Thus, there is no common attitude about Facebook reviews and their general information direction. Nevertheless, the previous results showed that Facebook comments with review character are indeed more negative than the reviews on a review site.

About 23% of the respondents indicate that they have gone to Facebook at least once to gather information about a shop and to support their purchase decision. However, only 20% of those who did, do this on a regular basis. One of the reasons is that people can check the profile of the author. Others stated, that they used Facebook in order to double-check the results from review sites. However, some participants stated that Facebook comments about shops and products are mostly negative statements.

Furthermore, the survey revealed that the participants had no common opinion about whether Facebook reviews are more or less manipulated than the reviews on review sites. The results were normally distributed, meaning that most of the people neither agreed nor disagreed on the statement about the manipulation.

Asking for the importance of different trust indicators showed the following distribution of the criteria: Most important was the information quality, mentioned by more than 90% of the students. Second- and third-most important criteria are the ease of use as well as the online reputation of a shop. Far behind, after intermediaries, privacy and security features, design of the shop and third party guarantees, social media content was listed. The most unimportant criteria were the perceived size, offline presence and social presence cues.

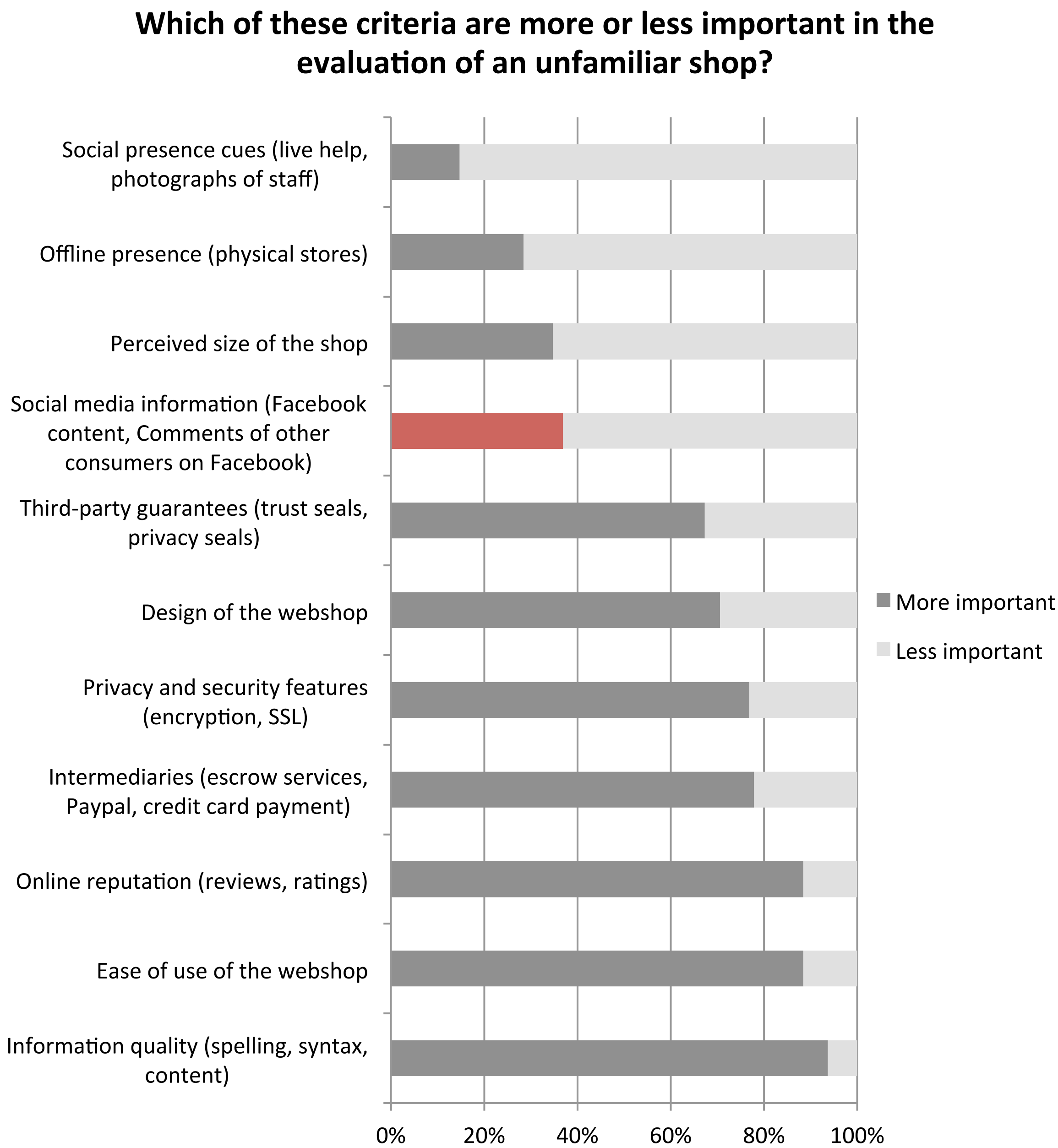


**Figure 4: Importance of trust indicators in the evaluation of unfamiliar shops.**

In the following, some information about the interpretation of the conjoint analysis results will be repeated. SPSS reports one table for utilities and one for importance values. While importance values show how relevant the single attributes are (i.e. display size, weight, battery life), utilities refer to the impact of the single values (e.g. 4-inch, 4.7-inch or 5.5-inch screen). However, for this study, the term utility values is not appropriate. Since the values show which impact the information (negative, neutral or positive) from the respective source (Facebook or review site) has on the purchase likelihood, these values will be called impact values in this study. In addition to the impact values, importance values show, which source is trusted more.

It should be mentioned that the values are not absolute. Thus, they do not present the absolute impact of the single information sources. An interpretation of the results is therefore only possible in relation to the other values. The difference between the minimum and the maximum of an attribute is no indication for the absolute relevance of the respective attribute. Therefore, also attributes with little dispersion can have a major influence. It only shows that variations within the attribute do not lead to major changes in the preference of the participant (Backhaus, Erichson, Plinke, & Weiber, 2015; Hahn, 1997).

Looking at the impact values of the total sample in table 15, there is a negative impact for negative information for both sources. As explained in the research design section, the values indicate how much the purchase likelihood for a specific combination changes when the attribute values are changed. In this example, the likelihood changes to a negative value when the information is negative. However, this was expected since negative statements about a shop should not influence the purchase likelihood positively. Whether a consumer buys at a shop depends on the sum of all impact estimates of the respective attribute levels (negative, neutral or positive) and the constant value. Applied to this study, the impact estimates as well as the constant do not have a direct relevance. Nevertheless, they could have indicated whether negative comments on Facebook have less effect on the purchase likelihood than positive comments or the other way around. Nevertheless, this could not be found in the data. The standard error column shows how precise the impact values could be estimated. Overall, the standard errors were acceptable. More important for the present study are the importance values in the second part of Table 15. These values show the relative importance of the single attributes - in this case - the information from Facebook or review sites. In the total sample, the respective values were relatively close, indicating, that there is no extreme difference regarding the two channels. However, according to the results, review site information has slightly more impact on the purchase likelihood of the consumer than Facebook comments.

**Table 15: Impact estimates and importance values. Total sample.**

| Impact | | | |
|---|---|---|---|
| | | Impact Estimate | Std. Error |
| facebook | positive | 0.975 | 0.125 |
| | neutral | 0.042 | 0.125 |
| | negative | -1.018 | 0.125 |
| review_site | positive | 1.326 | 0.125 |
| | neutral | 0.070 | 0.125 |
| | negative | -1.396 | 0.125 |
| (Constant) | | 3.761 | 0.089 |

| Importance Values | |
|---|---|
| facebook | 44.281 |
| review_site | 55.719 |
| Averaged Importance Score | |

Unfortunately, the evaluations per individual participants showed no clear homogeneity. Therefore, it can be assumed that there are subgroups for which these results are not applicable. Through segmentation, further interpretation is possible. There are two ways to segment a sample. The first one is to split the total sample on the basis of population criteria. As an example, the sample can be split in male and female. This type of segmentation is called a-priori-segmentation. However, this type of segmentation is only appropriate if differences in the population are observed. If a highly selective segmentation should be reached, an a-posteriori-segmentation is preferable (Klein, 2002).

In a-posteriori-segmentation, groups are built on the results of the preceding analysis. This is done by a cluster analysis. Therefore, clusters are developed on basis of preceding results. In this thesis, a cluster analysis is conducted in order to evaluate whether there are subgroups in the sample that conform in their judgement. In the framework of an explorative approach, further insights were expected.

Conducting a K-means cluster analysis, two major groups were differentiated. These are the results: For the first group, consisting of about 50% of the sample, review sites have a higher influence on the purchase likelihood. The other group showed a contradicting relationship. Nevertheless, the differences in the impact were higher for the group in which review sites have a higher effect. Thus, it can be assumed that there is a group, which will be called "traditional" (see Table 16), that focuses much more on review site ratings and does not much consider Facebook comments. The other group, called "modern" (see Table 17), seems to accept Facebook comments as a good indicator and pays even more attention to this source of information. Nevertheless, the difference is minor in the second group. A possible reason could be described by the statement of a participant. He uses Facebook comments as a secondary source to double-check the findings from the review site. Thus, review site information is important but Facebook comments are the clincher for the final decision. Nevertheless, the actual reasons for that relationship should be evaluated in further studies.

It is interesting, that both groups have about the same size. Therefore, the total scores are mainly influenced by the higher importance of review site ratings in the "traditional" group.

**Table 16: Impact estimates and importance values. Group "traditional". N=45**

| | | Impact | |
|---|---|---|---|
| | | Impact Estimate | Std. Error |
| facebook | positive | 0.679 | 0.058 |
| | neutral | 0.005 | 0.058 |
| | negative | -0.684 | 0.058 |
| review_site | positive | 1.753 | 0.058 |
| | neutral | 0.020 | 0.058 |
| | negative | -1.773 | 0.058 |
| (Constant) | | 3.877 | 0.041 |

| Importance Values | |
|---|---|
| facebook | 29.675 |
| review_site | 70.325 |
| Averaged Importance Score | |

**Table 17: Impact estimates and importance values. Group "modern". N=50**

| Impact | | | |
|---|---|---|---|
| | | Impact Estimate | Std. Error |
| facebook | positive | 1.242 | 0.186 |
| | neutral | 0.076 | 0.186 |
| | negative | -1.318 | 0.186 |
| review_site | positive | 0.942 | 0.186 |
| | neutral | 0.116 | 0.186 |
| | negative | -1.058 | 0.186 |
| (Constant) | | 3.658 | 0.132 |

| Importance Values | |
|---|---|
| facebook | 57.426 |
| review_site | 42.574 |
| Averaged Importance Score | |

# 5 Conclusion

## 5.1 Findings

The Internet provides multiple sources of opinions about online shops and services. Among these, review sites may be the most popular one. However, social media and social networks like Facebook catch up and generate more and more reviews. The variety of information sources bears a great opportunity for consumers to verify the trustworthiness of unfamiliar shops. However, different sources can lead to inconsistent information. While there is a lot of literature about inconsistent reviews, less is known about inconsistencies between different platforms and especially about how consumers deal with these. The main purpose of this study was to observe how consumers deal with different review sources – concretely, the review site Trustpilot and the social network Facebook - in case of inconsistent information. In order to test the four hypotheses, a method was created to compare reviews from Facebook to those on the review sites. The sentiment-mining tool Semantria was employed to assign reviews from both sources a sentiment score, which makes them comparable. Next, several tests were conducted in order to reveal the differences in the information direction between both channels. Finally, a conjoint analysis was conducted to examine the impact of each channel in situations were contradicting information is prevalent.

First, reviews on Trustpilot were compared to Facebook comments in order to examine whether the information direction of reviews (negative or positive) differs between Facebook and the review site. A sample of over 1900 comments and reviews was extracted from Facebook and Trustpilot. Then, the means per shop were compared to each other in order to reveal major inconsistencies between the two sources. In 70% of the cases, the average sentiment of Facebook reviews was lower than the sentiment of the respective Trustpilot reviews, supporting hypothesis 1. This was expected due to the findings of Kreis and Gottschalk (2015), who surveyed different review channels and their uses and gratifications. Their findings showed that Facebook serves more the social- and process-related gratifications through venting negative feelings (Kreis & Gottschalk, 2015). This thesis can be seen as a further argument for their findings. Furthermore, for some of the shops the Facebook sentiment was negative while the Trustpilot sentiment was positive. This is a further indication that there are coherences between the choice of a channel for review writing and the review sentiment. Further observations revealed that the standard deviations in sentiment scores on Facebook were higher than those on Trustpilot, indicating that Facebook comments are formulated stronger. While the means of the sentiment scores represent the average sentiments, standard deviations show how much the single scores differ from this average. Facebook reviews have a much wider distribution compared to reviews on Trustpilot, which is in line with the assumptions of hypothesis 2.

A survey among 95 students revealed that Facebook is not as important as review sites in their decision-making processes. While 35% of the students state that Facebook comments are a good indicator for the trustworthiness of an online shop, only 20% go regularly to Facebook in order to search for information about web shops. It is noteworthy that about 30% agreed that comments on Facebook pages of online shops are mostly rather negative, indicating that consumers are aware of the coherences that were found with hypothesis 1. In a rating of trust indicators as outlined in chapter 2.2.2, Facebook comments were among the less important indicators (ranked 8$^{th}$ from 11). Only 35%

assigned them to the more important criteria. Thus, hypothesis 3 can be supported partly: There is a group of consumers who use Facebook for trust building in online shopping - nevertheless, the general acceptance of Facebook comments as a trust indicator is low.

Hypothesis 4 was supported by the conjoint analysis. The findings indicate that Facebook comments have in general less impact on the purchase likelihood than ratings on review sites. This is particularly important in cases where the information is inconsistent between the two channels. However, the results on the participant level were not homogeneous. Thus, there are consumers who classify Facebook comments higher than the ratings from a review site. Indeed, conducting an a-posteriori segmentation of the sample revealed two groups of equal size. For one group, ratings of review sites were more important, for the other Facebook comments. However, the difference in the importance values was higher in the first group leading to a higher importance of review site ratings in the total sample.

## 5.2 Implications and Discussion

This thesis is the first one that compared Facebook reviews with regular reviews on a review site. Although there are several limitations, this study has major implications for theory and practice. In the literature review, an extensive list of trust indicators was elaborated. These were ranked by students in a survey. The ranking presents the order of different trust indicators as well as the perceived relevance of Facebook features on trust. For one thing, this list can be helpful for other researchers who want to observe trust indicators. For practitioners like web-shop owners, the list is a good checklist for features they should have implemented in their shops and features that could be neglected. Furthermore, shop owners can use the results to evaluate reviews and comments on Facebook. The findings show that there are two groups, which should be considered by the shop owners. There is a large group for which Facebook comments have a high relevance. Since Facebook reviews are more often negative, negative feedback on Facebook can have a large effect on sales. Shop owners should not underestimate this effect and enforce more positive comments on their Facebook wall. Encouraging satisfied consumers to leave comments on Facebook could be a potential means to build up a positive reputation. Nevertheless, both channels are important and should be treated with the same attention.

As initially stated, there are services that aggregate information about web shops from different channels throughout the Internet in order to support the consumer and facilitate the information search. For this, a way to extract information from Facebook and turn them into calculable values was presented in this study. Thus, Facebook comments can be integrated in their reputation systems, which enlarges the pool of information sources. Integrating social network reviews in multi-platform reputation systems has several advantages. First of all, it reduces the risk of bad decisions. Ignoring a great source of information could lead to negative outcomes. Since this study uncovered that Facebook comments are a valuable source of information for some consumers, their integration could increase the benefits of reputation systems. Furthermore, review sites often lack information about the authors while social network comments reveal much more data about the author and his or her character, which enhances the informative value of the respective reviews. However, differences in the importance of different sources should be considered. It turned out that Facebook comments have in general less impact than reviews on Trustpilot. Presuming that this relationship is also existing to other review sites, multi-platform aggregators should consider these differences in their reputation systems.

For instance, these services could take the observed differences into account when an overall reputation score based on different ratings is calculated. Multiplying the ratings from different sources with factors according to their relevance could lead to better recommendations.

## 5.3 Limitations and further Research

First of all, this study was restricted by the limitations of the Semantria trial version. Researchers who have unlimited access to the sentiment-mining tool should enhance this study by a greater sample and data set in order to increase the validity of the presented results. Regarding the review extraction from Facebook, a keyword query was used to filter reviews from other comments. Even if this method was highly successful in this study, it might not be appropriate for other studies. Developing a comprehensive list of review and e-commerce related keywords could be useful for a precise extraction of reviews from the social networks Facebook. Facebook is available in over 70 languages (Wilson et al., 2012) and the platform is used by over 1.7 billion people (Statista, n.d.). Utilizing Facebook as a place to extract opinions about web shops bears a high potential for e-commerce.

A major limitation is that the present thesis only examined whether there are differences in reviews on Facebook and a review site as well as in the importance for the consumer's judgment of trustworthiness while the reasons and coherences were neglected. Thus, taking different product features or shopping domains into account could enhance the implications of this study. The results may vary among these and detect further insights about relevant relationships. A technology freak might value Tweets much more than opinions on review sites. A teenager looking for apparel might hear much more on the voice of peers on Facebook. In their paper, Sen and Lerman (2007) mention that negative reviews are less useful for hedonic products than for utilitarian goods. Thus, consumers tend to evaluate negative reviews on luxury goods less than such for frequently bought necessaries (Khan, Dhar, & Wertenbroch, 2005; Sen & Lerman, 2007). Further segmentations of the sample could lead to more insights. Adding more sources like Twitter or other social media might be useful since different platforms serve a different audience with other needs and motives. Examining these differences might enhance the implications and lead to further improvements in reputation systems and their effect on trust. Furthermore, in-depth interviews with operators of public Facebook pages could add value to the understanding of social network comments related to e-commerce experiences. Their opinion could open up new perspectives that might be worth an observation. Furthermore, future studies should look for personality traits in consumers and compare these to the differences in review source importance. In the present study, an a-posteriori segmentation was conducted. Comparing a-posteriori segmented samples with characteristics of the participants would reveal further relationships. Especially the technology acceptance model could be a good starting point. Many studies that deal with e-commerce behavior lean on the technology acceptance model (e.g. Lederer et al., 2000). Furthermore, different social network uses and gratifications could provide an explanation for the differences in the relevance of the two sources.

Finally, this study was conducted mainly with German participants. The results may differ across cultures. Hofstedes cultural dimensions could be used to develop an intercultural observation and explain differences in review writing and perception (Minkov & Hofstede, 2011). The present work examined the tool Semantria and its capabilities regarding the processing of review information in a social network context. Semantria can process different languages. Different cultural dimensions may also be incorporated in the natural language. It might be an interesting research question to what extant the sentiment polarity differs between cultures with shared language and whether cultural dimensions could add to the precision of sentiment mining and the accuracy of review assessment. The more we know about the sources of information, the authors and their motives as well as the consumers and their expectations, the more can reputation systems be tailored to their needs.

## References


Abdul-rahman, A., Hailes, S., Street, G., & Kingdom, U. (1999). Relying on trust to find reliable information. In *Proceedings 1999 International Symposium on Database, Web and Cooperative Systems (DWACOS'99)* (pp. 1–6). Baden-Baden, Germany.

Ahmed, F., & Abulaish, M. (2013). A generic statistical approach for spam detection in Online Social Networks. *Computer Communications*, *36*(10-11), 1120–1129. http://doi.org/10.1016/j.comcom.2013.04.004

Asur, S., & Huberman, B. (2010). Predicting the Future with Social Media. *Computing*, *1*, 492–499. http://doi.org/10.1109/WI-IAT.2010.63

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2015). *Multivariate Analysemethoden: Eine Anwendungsorientierte Einführung*. Springer-Verlag.

Baltes-Götz, B. (2006). *Conjoint-Analyse mit SPSS*.

Bart, Y., Shankar, V., Sultan, F., & Urban, G. L. (2005). Are the Drivers and Role of Online Trust the Same for All Web Sites and Consumers? A Large-Scale Exploratory Empirical Study. *Journal of Marketing*, *69*(4), 133–152. http://doi.org/10.1509/jmkg.2005.69.4.133

Belanger, F., Hiller, J. S., & Smith, W. J. (2002). Trustworthiness in Electronic Commerce : The Role of Privacy , Security , and Site Attributes. *The Journal of Strategic Information*, *11*(July 2016), 245–270. http://doi.org/10.1016/S0963-8687(02)00018-5

Beldad, A., de Jong, M., & Steehouder, M. (2010). Computers in Human Behavior How shall I trust the faceless and the intangible ? A literature review on the antecedents of online trust. *Computers in Human Behavior*, *26*, 857–869. http://doi.org/10.1016/j.chb.2010.03.013

Benedicktus, R. L., Brady, M. K., Darke, P. R., & Voorhees, C. M. (2010). Conveying Trustworthiness to Online Consumers: Reactions to Consensus, Physical Store Presence, Brand Familiarity, and Generalized Suspicion. *Journal of Retailing*, *86*(4), 322–335. http://doi.org/10.1016/j.jretai.2010.04.002

Berger, J., Sorensen, A. T., & Rasmussen, S. J. (2010). Positive Effects of Negative Publicity: When Negative Reviews Increase Sales. *Marketing Science*, *29*(5), 815–827. http://doi.org/10.1287/mksc.1090.0557

Bhasker, B. (2006). *Electronic commerce: framework, technologies and applications.* (4th ed.). New Delhi: McGraw Hill Education (India) Private Limited.

Bigley, G., & Pearce, J. L. (1998). Straining for shared meanings in organizational science: Problems of trust and distrust. *Academy of Management Review*, *23*(3), 405–421.

Bollen, J., & Mao, H. (2011). Twitter Mood Predicts The Stock Market. *Journal of Computational Science*.

Bushelow, E. E. (2012). Facebook Pages and Benefits to Brands. *The Elon Journal of Undergraduate Research in Communications*, *3*(December 2011), 5–20.

Carliner, S. (2004). *An Overview of Online Learning*. *Concepts, elements and implications* (Vol. 21). http://doi.org/10.1016/j.chb.2003.11.008

Celebi, S. I. (2015). How do motives affect attitudes and behaviors toward internet advertising and Facebook advertising? *Computers in Human Behavior*, *51*, 312–324. http://doi.org/10.1016/j.chb.2015.05.011

Chaovalit, P., & Zhou, L. (2005). Movie review mining: A comparison between supervised and unsupervised classification approaches. *, 2005. HICSS'05. Proceedings of the 38th Hawaii International Conference on System Schiences*

Cheung, C. M. K., & Lee, M. K. O. (2006). Understanding Consumer Trust in Internet Shopping : A Multidisciplinary Approach. *Journal of the American Society for Information Science and Technology*, *57*(4), 479–492. http://doi.org/10.1002/asi

Cho, J., Kwon, K., & Park, Y. (2009). Q-rater: A collaborative reputation system based on source credibility theory. *Expert Systems with Applications*, *36*, 3751–3760. http://doi.org/10.1016/j.eswa.2008.02.034

comScore Inc. (2007). Online Consumer-Generated Reviews Have Significant Impact on Offline Purchase Behavior. Retrieved November 19, 2016, from http://www.comscore.com/Insights/Press-Releases/2007/11/Online-Consumer-Reviews-Impact-Offline-Purchasing-Behavior?cs_edgescape_cc=US

Dellarocas, C. N. (2000). Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. In *Proceedings of the 2nd ACM conference on Electronic commerce* (pp. 150–157). http://doi.org/http://doi.acm.org/10.1145/352871.352889

Dellarocas, C. N. (2001). Building Trust On-Line: The Design of Reliable Reputation Reporting Mechanisms for Online Trading Communities. *Social Science Research Network*, *2000*(July), 95–113. http://doi.org/10.2139/ssrn.289967

Doney, P. M., Cannon, J., & Mullen, M. (1998). Understanding the influence of national culture on the development of trust. *Academy of Management Review*, (July), 23.

Doney, P. M., & Cannon, J. P. (1997). An Examination of the Nature of Trust in Buyer-Seller Relationships. *Journal of Marketing*, *61*(April), 35–51. http://doi.org/10.2307/1251829

Fishbein, M. & Ajzen, I. (1975). Belief, attitude, attitude, intention and behavior: An introduction to theory of research. *Reading, MA : Addison-Wesley Addison-Wesley*, 578.

Gefen, D. (2002). Reflections on the dimensions of trust and trustworthiness among online consumers. *ACM Sigmis Database*, *33*(3), 38–53. http://doi.org/10.1145/569905.569910

Gottschlich, J., Heimbach, I., & Hinz, O. (2013). the Value of Users' Facebook Profile Data – Generating Product Recommendations for Online Social Shopping Sites. In *European Conference on Information Systems* (pp. 1–12).

Grabner-Kraeuter, S. (2002). The role of consumers' trust in online-shopping. *Journal of Business Ethics*.

Grishchenko, V. S. (2004). A fuzzy model for context-dependent reputation. In *Proceedings of the 2004 International Conference on Trust, Security, and Reputation on the Semantic Web-Volume 127* (pp. 72–77).

Hahn, C. (1997). *Conjoint- und discrete Choice-Analyse als Verfahren zur Abbildung von Präferenzstrukturen und Produktauswahlentscheidungen: ein theoretischer und computergestützter empirischer Vergleich*. Lit-Verlag.

Harris, L., & Dennis, C. (2011). Engaging customers on facebook: Challenges for e-retailers. *Journal of Consumer Behaviour*, *10*(6), 338–346. http://doi.org/10.1002/cb.375

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*. http://doi.org/10.1002/dir.10073

Heydari, A., Tavakoli, M., Salim, N., & Heydari, Z. (2015). Detection of review spam: A survey. *Expert Systems with Applications*. http://doi.org/10.1016/j.eswa.2014.12.029

Hong, J. Y. J., & Kim, J. H. (2015). *Are Social Media Useful for Managing Reputation Online? Comparing User Interactions Online with Reputation Indicators*. Springer International Publishing. http://doi.org/10.1007/978-3-319-20367-6

Huang, Z., Yoon, S. Y., & Benyoucef, M. (2012). Adding Social Features to E-commerce. *2012 Proceedings of the Conference on Information Systems Applied Research*, 1–11.

Hung, Y. H., Wu, L. F., & Chen, C. C. (2014). Building trust in hospitality and culture exchange travel sites: Lessons from heuristic evaluation of CouchSurfing. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8528 LNCS, pp. 630–642). Springer International Publishing. http://doi.org/10.1007/978-3-319-07308-8_60

Hussain, F. K., Hussain, O. K., & Chang, E. (2007). An overview of the interpretations of trust and reputation. In *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA* (pp. 826–830). http://doi.org/10.1109/EFTA.2007.4416865

Internet Retailing (2016). Pureplay Top 100: The best online-only retailers in the UK 2016 - InternetRetailing. Retrieved from http://internetretailing.net/iruk/pureplay-top100-best-online-retailers-uk-2016/

Jarvenpaa, S. L., Tractinsky, N., & Vitale, M. (2000). Consumer Trust in an Internet Store: A cross cultural validation. *Information Technology and Management*, *1*(1), 45–71. http://doi.org/10.1111/j.1083- 6101.1999.tb00337.x

Jones, G. R., & George, J. M. (1998). the Experience and Evolution of Trust: Implications for Cooperation and Teamwork. *Academy of Management Review*, *23*(3), 531–546. http://doi.org/10.5465/AMR.1998.926625

Jøsang, A., Ismail, R., & Boyd, C. (2007). A survey of trust and reputation systems for online service provision. *Decision Support Systems*, *43*(2), 618–644. http://doi.org/10.1016/j.dss.2005.05.019

Kent, M., & Taylor, M. (1998). Building dialogic relationships through the World Wide Web. *Public Relations Review*, 321–334.

Khan, U., Dhar, R., & Wertenbroch, K. (2005). A behavioral decision theory perspective on hedonic and utilitarian choice. In S. Ratneshwar & D. G. Mick (Eds.), *Inside consumption: Frontiers of Research on Consumer Motives, Goals, and Desires* (pp. 144–165).

Kim, D. J., Ferrin, D. L., & Raghav Rao, H. (2009). Trust and satisfaction, two stepping stones for successful e-commerce relationships: A longitudinal exploration. *Information Systems Research*, *20*(2), 237–257. http://doi.org/10.1287/isre.1080.0188

Kim, D., Song, Y., Braynov, S., & Rao, R. (2001). A B-to-C trust model for on-line exchange. *AMCIS 2001 Proceedings*.

Kim, H. W., Xu, Y., & Koh, J. (2004). A Comparison of Online Trust Building Factors between Potential Customers and Repeat Customers. *Journal of the Association for Information Systems*, *5*(10), 392–420. http://doi.org/10.1016/j.ijhcs.2003.06.002

Kim, J., & Moon, J. Y. (1998). Designing towards emotional usability in customer interfaces—trustworthiness of cyber-banking system interfaces. *Interacting with Computers*, *10*(97), 1–29. http://doi.org/10.1016/S0953-5438(97)00037-4

Kimery, K. M., Mccord, M., & Kimery, K. (2002). Third-party assurances: mapping road to trust in E-retailing. *Journal of Information Technology Theory and Application*, *4*(2), 63–82.

Klein, M. (2002). Die Conjoint-Analyse: Eine Einführung in das Verfahren mit einem Ausblick auf mögliche sozialwissenschaftliche Anwendungen. *ZA-Information / Zentralarchiv Für Empirische Sozialforschung*, *50*, 7–45.

Koufaris, M., & Hampton-Sosa, W. (2004). The development of initial trust in an online company by new customers. *Information & Management*.

Koehn, D. (2003). The nature of and conditions for online trust. *Journal of Business Ethics*.

Kreis, H., & Gottschalk, S. A. (2015). Relating eWOM Motives to eWOM Channel Choice – Why do We Post Where We Do? *Schmalenbach Business Review (SBR)*, (October), 406–429.

Lederer, A. L., Maupin, D. J., Sena, M. P., & Zhuang, Y. (2000). The technology acceptance model and the World Wide Web. *Decision Support Systems*, *29*(3), 269–282. http://doi.org/10.1016/s0167-9236(00)00076-2

Lee, E., Kim, Y., & Ahn, J. (2014). How do people use Facebook features to manage social capital? *Computers in Human Behavior*.

Lee, J., Park, D. H., & Han, I. (2008). The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic Commerce Research and Applications*, *7*(3), 341–352. http://doi.org/10.1016/j.elerap.2007.05.004

Lewis, D., & Weigert, A. (1985). Trust as a Social Reality. *Social Forces*, *63*(4), 967–985. http://doi.org/10.1093/sf/63.4.967

Liao, C., Palvia, P., & Lin, H. (2006). The roles of habit and web site quality in e-commerce. *International Journal of Information Management*.

Lim, K., Sia, C., Lee, M., & Benbasat, I. (2006). Do I Trust You Online, and If So, Will I Buy? An Empirical Study of Two Trust-Building Strategies. *Journal of Management Information Systems*, *23*(2), 233–266. http://doi.org/10.2753/MIS0742-1222230210

Liu, B. (2010). Sentiment Analysis and Subjectivity. *Handbook of Natural Language Processing*, 433–436.

Lohse, G., & Spiller, P. (1998). Electronic shopping. *Communications of the ACM*.

Luo, Q., & Zhong, D. (2015). Using social network analysis to explain communication characteristics of travel-related electronic word-of-mouth on social networking sites. *Tourism Management*, *46*, 274–282. http://doi.org/10.1016/j.tourman.2014.07.007

Martín-Valdivia, M. T., Martínez-Cámara, E., Perea-Ortega, J. M., & Ureña-López, L. A. (2013). Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches. *Expert Systems with Applications*, *40*(10), 3934–3942. http://doi.org/10.1016/j.eswa.2012.12.084

McKnight, D. H., & Chervany, N. L. (2000). Trust and Distrust Definitions : One Bite at a Time. *Proceedings of the Workshop on Deception, Fraud, and Trust in Agent Societies. Held during the Autonomous Agents Conference: Trust in Cyber-Societies, Integrating the Human and Artificial Perspectives*, 27–54. http://doi.org/10.1007/3-540-45547-7_3

McKnight, D. H., Choudhury, V., & Kacmarc, C. (2002). The impact of initial consumer trust on intentions to transact with a web site: A strust building model. *Journal of Strategic Information Systems*, *11*, 297–323.

McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Trust formation in new organizational relationships. *Academy of Management Review*, *23*(3), 473–490. http://doi.org/10.5465/AMR.1998.926622

Minkov, M., & Hofstede, G. (2011). The evolution of Hofstede's doctrine. *Cross Cultural Management: An International Journal*, *18*(1), 10–20. http://doi.org/10.1108/13527601111104269

Misztal, B. A. (1996). *Trust in Modern Societies: The Search for the Bases of Social Order*. Polity Press. http://doi.org/10.1521/bumc.2013.77.1.94

Mui, L., Mohtashemi, M., & Ang, C. (2001). A probabilistic rating framework for pervasive computing environments. *Proceedings of the MIT Student Oxygen Workshop*.

Nisbett, R. E., & Wilson, T. D. (1977). The Halo Effect: Evidence for Unconscious Alteration of Judgments. *Journal of Personality and Social Psychology*, *35*(4), 250–256. http://doi.org/http://dx.doi.org.ezproxy.snhu.edu/10.1037/0022-3514.35.4.250

Nusair, K., & Kandampully, J. (2008). The antecedents of customer satisfaction with online travel services: a conceptual model. *European Business Review*.

Palmer, J. W., Bailey, J. P., & Faraj, S. (2000). The Role of Intermediaries in the Development of Trust on the WWW: The Use and Prominence of Trusted Third Parties and Privacy Statements. *Journal of Computer-Mediated Communication*, *5*(3). http://doi.org/10.1111/j.1083-6101.2000.tb00342.x

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, *2*(1–2), 1–135. http://doi.org/10.1561/1500000011

Park, C., & Lee, T. (2009). Information direction, website reputation and eWOM effect: A moderating role of product type. *Journal of Business Research*.

Pavlou, P., & Gefen, D. (2004). Building effective online marketplaces with institution-based trust. *Information Systems Research*.

Phelan, K. V., Chen, H. T., & Haney, M. (2013). "Like" and "Check-in": how hotels utilize Facebook as an effective marketing tool. *Journal of Hospitality and Tourism Technology*, *4*(2), 134–154. http://doi.org/10.1108/JHTT-Jul-2012-0020

Qiu, L., & Benbasat, I. (2005). Online consumer trust and live help interfaces: The effects of text-to-speech voice and three-dimensional avatars. *International Journal of Human-Computer Interaction*, *19*(1), 75–94. http://doi.org/10.1207/s15327590ijhc1901_6

Reynolds, W. (1965). The role of the consumer in image building. *California Management Review*.

Riegelsberger, J., Sasse, M., & McCarthy, J. (2003). Shiny happy people building trust?: photos on e-commerce websites and consumer trust. *Proceedings of the SIGCHI*.

Sabater, J., & Sierra, C. (2001). REGRET: A reputation model for gregarious societies. In *Proceedings of the 4th Int. Workshop on Deception, Fraud and Trust in Agent Societies* (Vol. 73, pp. 194–195). http://doi.org/10.1145/375735.376110

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research Methods for Business Students*. *Research methods for business students*. http://doi.org/10.13140/RG.2.1.1419.3126

Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the web. *Journal of Interactive Marketing*, *21*(4), 76–94. http://doi.org/10.1002/dir.20090

Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*, *311*, 18–38. http://doi.org/10.1016/j.ins.2015.03.040

Simon, S. J. (2001). The Impact of Culture and Gender on Web Sites: An Empirical Study. *Advances in Informaiton Systems*, *32*(1).

Smith, A., & Anderson, M. (2016). Online Shopping and E-Commerce. *Pew Research Center.* Retrieved February 12, 2017 from http://www.pewinternet.org/2016/12/19/online-shopping-and-e-commerce/

Statista. (n.d.). Facebook users worldwide 2016. Retrieved December 12, 2016, from https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, *62*(2), 406–418. http://doi.org/10.1002/asi.21462

Thelwall, M., Wouters, P., & Fry, J. (2008). Information-centered research for large-scale analyses of new information sources. *Journal of the American Society for Information Science and Technology*, *59*(9), 1523–1527. http://doi.org/10.1002/asi.20829

Toms, E., & Taves, A. (2004). Measuring user perceptions of web site reputation. *Information Processing & Management*.

Trustpilot. (n.d.). Erläuterung zum TrustScore – Trustpilot Support-Center. Retrieved November 19, 2016, from https://support.trustpilot.com/hc/de/articles/201748946-Erl%C3%A4uterung-zum-TrustScore

Wang, Y. D., & Emurian, H. H. (2005). An overview of online trust: Concepts, elements, and Implications. *Computers in Human Behavior*, *21*(1), 105–125. http://doi.org/10.1016/j.chb.2003.11.008

Wang, Y., & Vassileva, J. (2003). Bayesian Network-Based Trust Model. In *Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003)*, 372-378.

Wang, Y., & Vassileva, J. (2007). A Review on Trust and Reputation for Web Service Selection. In *27th International Conference on Distributed Computing Systems Workshops* (pp. 3–8). http://doi.org/10.1109/ICDCSW.2007.16

Wilson, R. E., Gosling, S. D., & Graham, L. T. (2012). A Review of Facebook Research in the Social Sciences. *Perspectives on Psychological Science*, *7*(3), 203–220. http://doi.org/10.1177/1745691612442904

Xiao, B., & Bensabat, I. (2008). Product-related deception in e-commerce: A theoretical perspective, *32*(1), 25–33.

Zhang, K. Z. K., Cheung, C. M. K., & Lee, M. K. O. (2014). Examining the moderating effect of inconsistent reviews and its gender differences on consumers' online shopping decision. *International Journal of Information Management*, *34*(2), 89–98. http://doi.org/10.1016/j.ijinfomgt.2013.12.001

Zhang, Y., & Pennacchiotti, M. (2013). Recommending branded products from social media. *Proceedings of the 7th ACM Conference on Recommender Systems - RecSys '13*, 77–84. http://doi.org/10.1145/2507157.2507170

# Appendix A

**Full list of all shops selected for this thesis based on a list of the top 100 online-only retailers** (Internet Retailing, 2016)**.**

| Shop name | Trading website | https://de.trustpilot.com/review/... | Facebook ID |
|---|---|---|---|
| AO | ao.com | www.ao.com | 108506697584 |
| BonPrix | bonprix.co.uk | www.bonprix.co.uk | 140257922695181 |
| Bohoo | boohoo.com | www.boohoo.com | 112850788732826 |
| JD Williams | jdwilliams.co.uk | www.jdwilliams.co.uk | 503270076404440 |
| Ebuyer | ebuyer.com | www.ebuyer.com | 25406059349 |
| Firebox | firebox.com | www.firebox.com | 81798513556 |
| Get The Label | getthelabel.com | getthelabel.com | 152068310849 |
| Getting Personal | gettingpersonal.co.uk | www.gettingpersonal.co.uk | 12715085660 |
| Just Fab | Justfab.co.uk | www.justfab.co.uk | 351894691532965 |
| Kitbag | kitbag.com | www.kitbag.com | 6672735668 |
| La Redoute | laredoute.co.uk | www.laredoute.co.uk | 126718768641 |
| Light in the box | lightinthebox.com | www.lightinthebox.com | 181245226552 |
| Littlewoods | littlewoods.com | www.littlewoods.com | 236783176807 |
| Marisota | marisota.co.uk | www.marisota.co.uk | 191703790869555 |
| Mobile Fun | mobilefun.co.uk | www.mobilefun.co.uk | 112779212099877 |
| Moonpig | moonpig.com/uk | moonpig.co.uk | 127708073462 |
| Not on the high street | notonthehighstreet.com | www.notonthehighstreet.com | 35351148881 |
| Photobox | photobox.co.uk | www.photobox.co.uk | 8411712370 |
| Viking | viking-direct.co.uk | www.viking-direct.co.uk | 159713114110991 |
| Zalando | zalando.co.uk | www.zalando.co.uk | 107018652710311 |

# Appendix B

**List of shops, review amount, average star rating and trust score. Obtained from trustpilot.com on 03.11.2016.**

| Shop | Reviews[3] | Average star rating | Trust Score[4] | Inviting[5] |
|---|---|---|---|---|
| ao.com | 73429 | 5 | 9.5 | X |
| bonprix.co.uk | 777 | 5 | 9.1 | X |
| boohoo.com | 116891 | 4 | 7.0 | X |
| jdwilliams.co.uk | 100 | 1 | 2.5 | |
| ebuyer.com | 18027 | 5 | 9.1 | |
| firebox.com | 626 | 5 | 9.2 | |
| getthelabel.com | 115 | 3 | 6.9 | |
| gettingpersonal.co.uk | 228 | 2 | 3.6 | |
| justfab.co.uk | 843 | 3 | 6.2 | |
| kitbag.com | 1798 | 3 | 6.6 | |
| laredoute.co.uk | 706 | 3 | 6.9 | |
| lightinthebox.com | 6749 | 4 | 7.6 | |
| littlewoods.com | 2493 | 2 | 3.9 | |
| marisota.co.uk | 87 | 1 | 2.6 | |
| mobilefun.co.uk | 1658 | 4 | 8.6 | |
| moonpig.com/uk | 157 | 2 | 3.4 | |
| notonthehighstreet.com | 82 | 1 | 2.0 | |
| photobox.co.uk | 13902 | 4 | 8.4 | |
| viking-direct.co.uk | 230 | 2 | 4.9 | |
| zalando.co.uk | 258 | 2 | 4.6 | |

[3] The number of reviews varies substantially between the different shops. However, this should not affect the thesis. This study aims to observe differences in the content of different review channels and the impact of the different sources on the purchase intention.
[4] Trustpilot Trust Score is a computed value based on several variables like amount of ratings, age of the single ratings as well as the single ratings itself.
[5] Some companies invite their customers to review them on specific review sites. Thus, the results and average ratings might be influenced.

# Appendix C

**Table of all shops and the mean difference between the average sentiment of reviews and Facebook comments.**

| Independent Samples Test | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | |
| Shopname | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference * | Std. Error Difference |
| AO | Equal variances assumed | 5.43 | 0.021 | -0.038 | 158 | 0.970 | -0.0021 | 0.0553 |
| | Equal variances not assumed | | | -0.038 | 136.123 | 0.970 | -0.0021 | 0.0553 |
| Bonprix | Equal variances assumed | 1.917 | 0.171 | -2.032 | 64 | 0.046 | -0.1883 | 0.0926 |
| | Equal variances not assumed | | | -2.032 | 61.381 | 0.046 | -0.1883 | 0.0926 |
| Boohoo | Equal variances assumed | 0.252 | 0.617 | 0.845 | 122 | 0.400 | 0.0602* | 0.0713 |
| | Equal variances not assumed | | | 0.848 | 116.94 | 0.398 | 0.0602 | 0.0710 |
| Ebuyer | Equal variances assumed | 0.539 | 0.464 | -8.545 | 131 | 0.000 | -0.5338 | 0.0625 |
| | Equal variances not assumed | | | -8.550 | 130,618 | 0.000 | -0.5338 | 0.0624 |
| Firebox | Equal variances assumed | 0.913 | 0.345 | -4.606 | 42 | 0.000 | -0.5042 | 0.1095 |
| | Equal variances not assumed | | | -4.606 | 37.912 | 0.000 | -0.5042 | 0.1095 |
| Getthelabel | Equal variances assumed | 5.012 | 0.027 | -3.482 | 108 | 0.001 | -0.2875 | 0.0826 |
| | Equal variances not assumed | | | -3.482 | 104.305 | 0.001 | -0.2875 | 0.0826 |
| Gettingpersonal | Equal variances assumed | 3.415 | 0.067 | -1.099 | 126 | 0.274 | -0.0978 | 0.0889 |
| | Equal variances not assumed | | | -1.099 | 116.155 | 0.274 | -0.0978 | 0.0889 |
| Jdwilliams | Equal variances assumed | 0.702 | 0.405 | -0.533 | 84 | 0.596 | -0.0516 | 0.0968 |
| | Equal variances not assumed | | | -0.533 | 82.254 | 0.596 | -0.0516 | 0.0968 |
| JustFab | Equal variances assumed | 1.005 | 0.319 | 0.069 | 74 | 0.945 | 0.0072* | 0.1044 |
| | Equal variances not assumed | | | 0.069 | 73.922 | 0.945 | 0.0072 | 0.1044 |
| Kitbag | Equal variances assumed | 0.107 | 0.744 | -1.723 | 102 | 0.088 | -0.1460 | 0.0847 |
| | Equal variances not assumed | | | -1.723 | 101.979 | 0.088 | -0.1460 | 0.0847 |

## Appendix C continued

**Independent Samples Test**

| Shopname | | Levene's Test for Equality of Variances F | Sig. | t-test for Equality of Means t | df | Sig. (2-tailed) | Mean Difference * | Std. Error Difference |
|---|---|---|---|---|---|---|---|---|
| Laredoute | Equal variances assumed | 1.211 | 0.273 | -4.634 | 130 | 0.000 | -0.3728 | 0.0804 |
| | Equal variances not assumed | | | -4.634 | 128.633 | 0.000 | -0.3728 | 0.0804 |
| Lightinthebox | Equal variances assumed | 1.273 | 0.263 | -6.276 | 74 | 0.000 | -0.6204 | 0.0988 |
| | Equal variances not assumed | | | -6.276 | 71.993 | 0.000 | -0.6204 | 0.0988 |
| Littlewoods | Equal variances assumed | 2.961 | 0.088 | -1.615 | 102 | 0.109 | -0.1432 | 0.0887 |
| | Equal variances not assumed | | | -1.615 | 94.141 | 0.110 | -0.1432 | 0.0887 |
| Marisota | Equal variances assumed | 0.071 | 0.791 | 1.836 | 83 | 0.070 | 0.1651* | 0.0900 |
| | Equal variances not assumed | | | 1.836 | 82.998 | 0.070 | 0.1651 | 0.0899 |
| Mobilefun | Equal variances assumed | 0.275 | 0.601 | -3.338 | 94 | 0.001 | -0.2532 | 0.0758 |
| | Equal variances not assumed | | | -3.338 | 93.9 | 0.001 | -0.2532 | 0.0758 |
| Moonpig | Equal variances assumed | 1.314 | 0.255 | -1.763 | 96 | 0.081 | -0.1705 | 0.0967 |
| | Equal variances not assumed | | | -1.763 | 95.203 | 0.081 | -0.1705 | 0.0967 |
| Notinthehighstr | Equal variances assumed | 0.889 | 0.352 | 0.419 | 38 | 0.678 | 0.0674* | 0.1610 |
| | Equal variances not assumed | | | 0.419 | 36.91 | 0.678 | 0.0674 | 0.1610 |
| Photobox | Equal variances assumed | 14.556 | 0 | -5.751 | 100 | 0.000 | -0.5091 | 0.0885 |
| | Equal variances not assumed | | | -5.751 | 86.523 | 0.000 | -0.5091 | 0.0885 |
| Viking | Equal variances assumed | 0.013 | 0.911 | -5.070 | 96 | 0.000 | -0.4544 | 0.0896 |
| | Equal variances not assumed | | | -5.070 | 95.891 | 0.000 | -0.4544 | 0.0896 |
| Zalando | Equal variances assumed | 0.703 | 0.405 | -2.854 | 72 | 0.006 | -0.2805 | 0.0983 |
| | Equal variances not assumed | | | -2.871 | 70.05 | 0.005 | -0.2805 | 0.0977 |

*The values labeled in red are the samples that do not meet the assumption of hypothesis 1, meaning that Facebook comments were more positive in these cases.