Master Thesis

# Boosting Online Trust Perception

## *Does the implementation of a reputation system foster the perception of trust on B2C online marketplaces?*

A. L. Pieniazek

July 19, 2013

*Trust is a major success factor in any sort of transaction or interaction – interpersonal, business or alike. Since years, the subject of trust gained on importance in the growing field of online based activities and advanced to the leading variable for contestability and success. But the intangible and multidimensional nature of online trust complicates its measurement and the detection of its determinants. By an extensive literature review, the specific role of online trust and its influencing factors are identified. Thereby, so-called reputation systems emerge as appropriate measures to communicate and eventually foster online trust perception. An online marketplace' attempt to implement such a system, serves as testing field in order to get an up to date look on its effectiveness. The empirical research aims at identifying a reputation system as measure to foster online trust perception and at the same time assess its resistance to fraudulent and manipulative behavior. In general, the results recognized reputation systems as working trust mechanisms. Although severe violations of the systems' robustness were not identified, the danger of fraud could however not be ruled out entirely. The thesis concludes with a set of propositions for further research and implications for practice.*

**Keywords**: *Online Trust, Online Marketplace, E-Commerce, Reputation Systems, Semantic Analysis, Fraud Prevention*

## General information

*Project initiator*


| | |
|---|---|
| Name: | Andreas L. Pieniazek |
| Student-ID: | s1359932 |
| E-mail: | a.l.pieniazek@student.utwente.nl |
| Telephone: | (+49) (0) 173/3951459 |


*Supervisory committee*


| | |
|---|---|
| 1.Internal supervisor: | Dr. A. B. J. M. Wijnhoven (UTwente) |
| E-mail: | fons.wijnhoven@utwente.nl |
| Telephone: | (+31) (0) 53 489 2372 |

| | |
|---|---|
| 2.Internal supervisor: | Dr. Koray Erek (TU-Berlin) |
| E-mail: | koray.erek@tu-berlin.de |
| Telephone: | (+49) (0)30 314 78703 |

| | |
|---|---|
| External supervisor: | Dr. Robert Martignoni (Autoscout 24) |
| E-mail: | rmartignoni@autoscout24.de |
| Telephone: | (+49) (0) 89 444 56-1221 |

## List of Tables



## List of Figures

# Contents

# Chapter 1: Introduction

*Online Trust Antecedents*

The discussion on trust increased among authors from several fields of research such as philosophy, sociology, economics or management. Several definitions evolved on the topic, specifically in business contexts (Blois, 1999). Schurr & Ozanne for instance define trust "as the belief that a party's word or promise is reliable and that a party will fulfill his/her obligations in an exchange relationship" (Schurr & Ozanne, 1985, p. 940). Further definitions see "trust as one party's belief that its needs will be fulfilled in the future by actions undertaken by the other party" (E. Anderson & Weitz, 1989, p. 312). In a broader perspective trust is "a willingness to rely on an exchange partner in whom one has confidence" (Moorman, Zaltman, & Deshpande, 1992, p. 315; Schurr & Ozanne, 1985, p. 940). As indicated the problem of trust as a general concept is the absence of a commonly accepted definition (Kee & Knox, 1970; McKnight & Chervany, 2002; Rosseau, Sitkin, Burt, & Camerer, 1998). What can be derived from various definitions of trust is a multidimensional approach of trust conceptualizations (Beldad, de Jong, & Steehouder, 2010). Thereby two main trust conceptualizations are identifiable: Trust viewed as an expectation of behavior in an interaction with a partner (Barber, 1983; Rotter, 1967) and trust seen as the acceptance and the exposure of an individual to vulnerability (Mayer, Davis, & Schoorman, 1995; Rosseau et al., 1998).

Approaching trust from a less abstract perspective, Child (2001) states, "(t)rust is vital for any relationship, business or otherwise, when there is insufficient knowledge and understanding of the other person or group"(Child, 2001, p. 276). Undoubtedly it has an important effect especially in the relationship initiation process, the further development of these relationships and obviously the persistence of them (Andersen, 2001; Cova & Salle, 2000; Witkowski & Thibodeau, 1999). One outcome of successful business relationships can lead to joint value creation of organizations, triggered either by rationalization, learning or both (Andersen & Kumar, 2006). As Kumar and Nti (1998) state, a high level of psychological commitment is required in order to enable relationships among business partners. An implication of this is that the confidence towards a partner must be on a high level to enable trust (Das & Teng, 1998). A further implication is the effort that is required for maintaining and deepening such relations (Andersen & Kumar, 2006). Several scholars identified trust as a catalyzer for a higher grade of adaptability and flexibility with a positive reinforcement of the business partner relationship (Arino, de la Torre, & Ring, 2001). According to Geyskens, Steenkamp, and Kumar (1998) most studies define trust among organizations and business partners as belief of one firm in the honesty or benevolence toward its cooperation partner. J. C. Anderson and Narus (1990) further distinguish among honesty and benevolence. They argue that honesty aims at a partner's reliability and benevolence is concerned with a partner's interest in joint benefits as well as the individual welfare of the cooperation partner.

What has to be kept in mind is that trust initially can only be provided by individuals (Blois, 1999). Therefore, the trust embedded in the interaction of organizations is based on the relationship quality of the people which represent those companies (Blois, 1999; Child, 2001). As a consequence trust creates the basis for successful teamwork and the creation of joint knowledge between different departments and units of one organization. In the long run trust can as well contribute to the overall firm performance (Chang & Wong, 2010; Child, 2001; Tzafrir, 2005). Taking a broader view Chang and Wong (2010) argue that missing trustworthiness illustrated by frequent cheating and fraud in a market can lead to market failure. As a result one can claim that trust moderates the stability of a community by the right balance of trust and distrust.

*Online Trust*

In online environments any sort of activity has a rather faceless or intangible character and a connection to human interactions or relationships among individuals may appear difficult (Beldad et al., 2010). Therefore, when referring to online transactions the perception of trust is a core ingredient for any e-commerce or social commerce system in order to foster loyalty on the consumer and provider side (Atif, 2002). According to Corritore, Kracher, and Wiedenbeck (2003) trust in an online situation of risk is given by an attitude of confident expectation in which the user's vulnerabilities will not be exploited. It can as well be understood as the confidence in a company by its stakeholders regarding the company's online activities and its web presence (Shankar, Urban, & Sultan, 2002). In a definition of Bart, Shankar, Sultan, and Urban (2005) emitting confidence by a website is as well one major criterion of online trust. Essentially, trust is based upon a positive impression of an electronic entity connected to a participants' willingness to accept vulnerability (Urban, Amyx, & Lorenzon, 2009). However, research identified further criteria describing the nature of online trust. Gefen (2002) for instance argues that next to confidence two more beliefs encompass trust: competence and benevolence. These criteria were also validated by Belanger, Hiller, and Smith (2002) or Lee and Turban (2001). Further McKnight, Choudhury, and Kacmar (2000) argue that trust is the belief in another one's benevolence and competence.

Hereby, the multidimensional character of trust and the difficulty for a common definition is again identifiable (Beldad et al., 2010; McKnight & Chervany, 2002). Nevertheless, the importance of trust for the online environment is decisive. This is because the attraction of new customers, members or participants and retaining them must be seen as critical for any sort of online entities (Jarvenpaa & Toad, 1996; Reichheld & Schefter, 2000). Sultan, Urban, Shankar, and Bart (2002) add that online trust significantly affects customers' intention to act with regard to purchase or loyalty.

The question arises how to actively communicate trust in web based environments. One possibility is the implementation of reputation systems such as ratings or reviews (Ba & Pavlou, 2002; De Maeyer & Estelami, 2011; Poston & Speier, 2005). This trust mechanism is based on the basic principle of word of mouth (Dellarocas, 2003). Thereby, a reputation system aims at the collection of feedback (in

form of ratings and reviews) and the aggregation as well as distribution of the feedback information; hereinafter referred to as reputation scores (Sun & Liu, 2012). Furthermore, an aggregation of customer ratings, in general consensus information, is a reputable means for enhancing online trust beliefs (Benedicktus, Brady, Darke, & Voorhees, 2010; Pavlou & Gefen, 2004). Quality enhancement of specific content by ratings and credibility indicators could as well be identified by Poston and Speier (2005) in different fields, in particular knowledge management systems. However, it is questionable whether the implementation of such improvement measures for online trust can work without failure or fraud. Since reputation systems rely mostly on user feedback this evokes the danger of manipulation and misleading information of some actors, for instance in order to promote themselves, their services or products (Jøsang, Ismail, & Boyd, 2007; Poston & Speier, 2005; Sun & Liu, 2012).

*Research Aim and Questions*

To address this impact bias of reputation systems, this thesis aims to explore the effects of implementing a reputation system on the perception of trust in B2C relationships. A further goal is to approach a given reputation system in terms of the indicators determining its success. The basic success factors as proposed by Resnick, Kuwabara, Zeckhauser, and Friedman (2000) are:

*(1) Accuracy for long-term performance*
*(2) User incentive in order to agglomerate feedback*
*(3) Usability and smoothness*

Whereas the first refers to the longevity of online entities and the belief that online activities should follow the expectation and possibility of a future interaction *(1)*. In order to make a reputation system practical, user feedback needs to be agglomerated first and then made available. This essential property is however dependent on users' willingness to provide feedback. Therefore, reputation systems must yield some sort of incentive *(2)*. The third requirement alludes to the actual usability of a reputation system and in which way participants of online activities respond to it *(3)*. However, these requirements are regarded as fulfilled by the majority of reputation systems in today's online environment with a high internet affinity among individuals and act as rather classic preconditions. Dingledine, Freedman, and Molnar (2000) extended those requirements with a further one which must still be seen as order of the day:

*(4) Robustness against attacks*

Beyond the above mentioned properties a reputation system's capabilities to resist fraudulent or manipulative behavior is the essential challenge for such systems today (Sun & Liu, 2012).

Consequently, the scope of the thesis is to scrutinize the impact of a reputation system on online participants' trust perception of online offerings. Thereby, a reputation system should be identified as

an effective tool strengthening the perception of online trust in a B2C context. A further intention is to evaluate if contemporary fraud prevention methods for reputation systems, even though they evolved over time, still might allow potentially fraudulent and manipulative behavior. The dilution of a reputation system's validity and biased decision support for customers might be the consequence (Poston & Speier, 2005; Sun & Liu, 2012). Therefore, possible measures for the perception of trust towards a cooperation partner, product or service have to be identified.

As a result the following central research question arises: *Does the implementation of a reputation system foster the perception of trust on B2C online marketplaces?*
This question is accompanied by the following sub-questions:

(a) What are the main determinants of online trust?
(b) How are reputation systems and fraud assessed in terms of online trust?
(c) Does a reputation system measurably affect the perception of online trust?
(d) How likely is fraud and manipulation of a current reputation system?
(e) Is there a need for a higher layer of trust which evaluates the rating's trustworthiness?

Question a) and b) will be approached by an extensive literature review in order to clarify the special role of online trust for any sort of online activities and which status reputation systems hereby incorporate. The main empirical research focus of this thesis lies upon the questions (c) – (e). Question c) will be elaborated and evaluated by an experimental research design to analyze the effects of the implemented reputation system on customers' online trust perception. The collaboration with an online marketplace company enables access to required data. For the impact and danger of fraud ((d) and (e)) a second empirical design tests the current fraud prevention mechanisms and scrutinizes the reliability of the system in use. Customer ratings of a different site feature of the company at hand will thereby serve as data set for evaluation.

The structure of the thesis follows an analogical order to the research questions above. *Chapter 2* examines the determinants of online trust by means of a literature review. *Chapter 3* goes further into detail regarding the role of reputation systems and completes the theoretical part of the thesis. *Chapter 4* incorporates the research approach for both the impact of a reputation system on trust and the likeliness of fraud. In *Chapter 5* the results of the research are presented. Thereby evolving conclusions, a discussion on limitations of the study and suggestions for future research are part of the last *Chapter 6*.

# Chapter 2: Literature Review

## 2.1 Setup of the Review

The gathering of background information is based on an extensive literature review. Thereby. the triangulation method for search patterns as exercised by Fielding (2012) is applied. For the start a more unfocused keyword search led to the identification of first usable literature on the topic. Terms as “online trust”, “e-commerce” “ratings” or “reputation system” helped to emphasize the topic and get a feeling from which directions the topic of interest is approached in literature. Furthermore the search with basic keywords enables a more specified search with advanced combinations of terms of interest. For instance, after having identified literature that deals with online trust in a general way, keyword combinations enabled to filter literature for coherences in the field of interest. By using exemplary phrases as “impact of online trust on”, “online trust generation”, “effects of trust mechanisms” or separated terms as “reputation system, fraud, manipulation” the field of potentially valuable inputs reopened.

The search for reliable literature with keywords was furthermore accompanied by forward and backward analysis of citations and references of already evaluated works. Results were clustered with regard to the *mutually exclusive* and *collectively exhaustive* (MECE) principle into several categories. Those were for instance *Managing Trust*, *Online Trust Perception*, *Ratings* or *Spam Detection*. Each category was then evaluated and thereby the amount of suitable literature systematically reduced. As main source for the qualitative and quantitative secondary data, primarily internet based search engines as *Web of Knowledge*, *SCOPUS* and *EBSCOhost* were used.

The purpose of this section is to establish a general understanding for the special nature of online trust by referring to the existing knowledge provided by literature. From a general perspective on the topic of trust a link to the particular requirements for trust in an online environment are established. A slightly modified model by Urban et al. (2009) thereby serves as guideline how the specific determinants of online trust relate to each other and which impact trust incorporates for online based activities and the involved parties (Figure 1). The review will conclude by distinguishing the role of reputation systems as an important trust mechanism and thereby scrutinize the principle of those systems and how they might affect the perception of trust by online participants.

In *Chapter 1* trust was identified as a multidimensional concept with different definitions in dependence of the perspective the subject is approached from. Having established some understanding for the nature of trust in general and in business relationship contexts, what can be derived from the literature in terms of trust in an online environment?

## 2.2 Dimensions of Online Trust

According to Corritore et al. (2003) online trust is described with an individual's expectation that its vulnerabilities will not be exploited in an online situation of risk. Other scholars refer to online trust as the reliance on and confidence in an organization by its stakeholders regarding all of the company's online activities (Shankar et al., 2002). Are there than major differences in the perception and communication of online and offline trust? Due to Corritore et al. (2003) the existing literature on offline trust serves as basis and in the majority of cases is applicable to online environments. Rosseau et al. (1998) for instance argue that trust might be characterized by some sort of psychological state which compromises an intention to accept vulnerability. Hence, it can be argued that online participants must possess a certain level of confidence in each other (Urban et al., 2009). Therefore the offline surrounding serves as starting point for research to asses trust in online environments. In a definition of Bart et al. (2005) (elaborated upon the determination of Rosseau et al. (1998)) emitting confidence by a website is one major criteria of online trust. Consequently online trust is based upon a positive impression of an online entity connected to a participants' willingness to accept vulnerability (Urban et al., 2009). The both dimensions – exposure to vulnerability and an individual's expectation towards the behavior of a partner – are also assumed as valid by Beldad et al. (2010).

Beyond that research identified further criteria describing online trust. Gefen (2002) demonstrated in an experiment that next to confidence, competence and benevolence criteria encompass trust. These criteria were also validated by Belanger et al. (2002), Lee and Turban (2001) and McKnight et al. (2000). Thereby, according to the competence criterion trust is facilitated when participants demonstrate to have the competencies, characteristics and required skills to influence opinions within a specific domain (Mayer et al., 1995). The benevolence criterion again insinuates the good nature of individuals by arguing that individuals set aside egocentric motives and self-serving behavior (Mayer et al., 1995). In order to facilitate trust, online entities should be aware that online trust is assessed based predominantly on confidence, competence and benevolence criteria. Research has meanwhile reached consistency by arguing that online trust can be broken down into those three dimensions (Urban et al., 2009).

Besides the congruities literature as well illustrates several differences in the nature of online and offline trust. The major one is characterized by the object of trust (Shankar et al., 2002). In offline transactions trust is associated with a person, company or entity (Doney & Cannon, 1997). The object of trust in online transactions however is exemplified by the internet (technology) as an activity enabler and the entity deploying this technology (Boyd, 2003; Shankar et al., 2002). Jøsang et al. (2007) comprehends another difference in offline information sharing which is rather limited to local environments such as communities or organizations. Information sharing via IT systems and the internet reaches a global scope. As a consequence performing online transactions may require a more

distinct understanding of cultural differences between transaction partners and legal concerns. This again points out the rather difficult nature of online trust and demonstrates a pattern of unpredictability which spawns risks and leads to situations of uncertainty (Pavlou, 2003); especially with new exchange infrastructures and the exposure to multiple merchants and options on a global scale (Angriawan & Thakur, 2008).

Still, literature identified trust as a major success factor for any sort of online interaction or transaction (Belanger et al., 2002; Lua, Chen, & Cai, 2011; Ratnasingam & Phan, 2003). Research thereby sets the scope of importance equally on a private consumer or business background. The following model (Figure 1) serves as guiding theme for the position trust incorporates in terms of online processes and upon that how trust is actually gathered and works. The casual model is based on a slightly modified approach of Urban et al. (2009). Thereby the relationships of the variables are the same but the determinants of each variable are treated in a more generalizable manner.

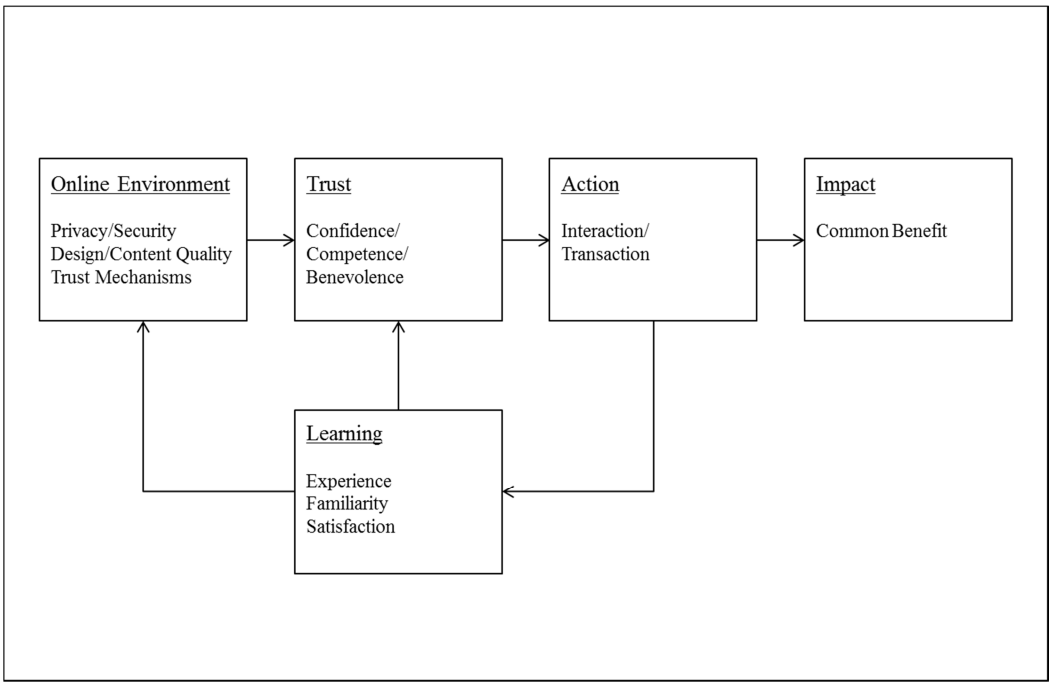



Figure 1: Modified Trust Model based on Urban et al. (2009)

In order to elaborate which position trust plays in such a correlation of different determinants, a starting point has to be specified. At first, trust is directly influenced by the online environment. As indicated by Figure 1 several attributes of the online environment do have an effect on online trust. Those features include *privacy* and *security* issues, the *design* and *content quality* of an online presence and finally special *trust mechanisms*, directly aiming at the communication and enhancement of online trust. The effect of such features may be the incidental or intentional generation of trust. Therefore trust is first of all a product of online environment features. The variable itself is based upon *confidence*, *competence* and *benevolence* criteria as stated by Beldad et al. (2010) as well as Urban et al. (2009).

Furthermore, trust mediates the relationship of perceiving and dealing with the online environment and undertaking any sort of online actions, such as *interactions* or *transactions*. Doney and Cannon (1997) for instance identified trust as a mediator which affects the decision consideration of individuals. Several studies further illustrated that trust as a mediating variable influences the purchase behavior of online participants (Jarvenpaa, Tractinsky, & Vitale, 2000; Shankar et al., 2002; Yoon, 2002). Bart et al. (2005) as well identified trust as mediator between online environment features such as website characteristics and users' behavioral intention. That means participants which intend to engage in any sort of activity are directly affected by their perception of trust. But the impact and role of trust does not stop there. If online trust can be emphasized as the mediating variable between the online environment and online activities some sort of positive impact might be addressed to the existence of this variable. This is in particular the case if online actions are perceived by its participants as favorable, based on the perception of trust. According to the model, such an impact can be described as *common benefits* for the stakeholders of online based actions. But the model does not conclude with such an outcome.

It should be realized that the building of trust during only one session is rather unlikely (Urban et al., 2009). Figure 1 accentuates this peculiarity by arguing that trust is as well generated as some sort of process. The process is presented by a feedback loop of Trust-Action-Learning with several repetitions (Urban et al., 2009), whereas the variable learning is characterized by determinants as the *experience*, *familiarity* and *satisfaction* participants perceived in past online activities. Consequently those determinants contribute in addition to the online environment features to the development of online trust.

Keeping such a constellation in mind each variable of the model will be examined in detail and illustrated how it correlates to online trust. Following Figure 1 the first step is to constitute which features of the online environment affect and shape trust.

## 2.3 Online Environment

### *2.3.1 Privacy and Security Features*

Figure 1 reveals privacy and security concerns as a first determinant affecting the perception of online trust. Privacy and security can be interpreted as basic criteria in order to assess the trustworthiness of an e-vendor or online transaction partner in any online relationship (Aiken & Boush, 2006). Especially first-time online customers are affected regarding such issues (Koufaris & Hampton-Sosa, 2004).

According to Yoon (2002), a customer's perception of online trust is significantly affected by the security of a transaction. Belanger et al. (2002) were able to come to similar outcomes in their study and registered a high rank for security features among their respondents and as well came to know that privacy statements had a strong impact on customer's perception of trust. The influence of privacy

concerns on trust perception have been already pointed out by very early studies on online trust as by Hoffman, Novak, and Peralta (1999). Such concerns can be characterized with spam mails, the tracking of the customers' internet usage history, third parties storing private and confidential information, tracking customer preferences with cookies and the exposure to organization with doubtful use of customers' private data (Wang, Lee, & Wang, 1998). Many of those concerns nowadays became illegal or at least are reduced by governmental regulations. Still other issues as for instance the storing of private information are problematic with new services evolving (like for instance cloud services) and the amount of stored private data rising (Garg, Versteeg, & Buyya, 2013).

More recent studies reveal that a company's trustworthiness is fostered by enhanced privacy policies on the organization's website (Lauer & Deng, 2007). Pan and Zinkhan (2006) support these findings. They were able to demonstrate in an experiment that internet users were positive about the trustworthiness of a company when a privacy policy was present. Thereby particularly the role of self-disclosure affected internet usage (Joinson, Reips, Buchanan, & Schofield, 2010; Nosko, Wood, & Molema, 2010) and as mentioned by Krasnova, Spiekermann, Koroleva, and Hildebrand (2010) trust impairs the perceived risks connected to the disclosure of identifiable information.

Then again, other studies revealed that many internet users are not really concerned anymore with the security and privacy features of websites by not consulting the organization's privacy statements before providing private data for online transactions (Arcand, Nantel, Arles-Dufour, & Vincent, 2007; Jensen, Potts, & Jensen, 2005). This illustrates a change of online trust perception over time and can be explained with the trust building process of Figure 1. Especially from the company perspective, online trust evolved since the origin of online transactions and e-commerce (Hoffman et al., 1999; Sultan et al., 2002). From an early focus, mostly on security issues like the handling of customers' confidential financial data, to the implementation of privacy policies in order to guarantee a professional processing with customer's personal data, the perception of trust evolved into a multidimensional and complex construct (Hoffman et al., 1999; Sultan et al., 2002).

Other studies renounce from the opinion that trust is perceived mostly through security and privacy features and detect new variables affecting online trust perception. Mesch (2012) for instance showed that the online and offline world are connected in terms of trust perception. "Offline trust (measured as trust in social institutions and trust in individuals) is associated with trust online" (Mesch, 2012, p. 1476). Those findings go in a similar direction as the implications by Gefen and Straub (2004) who identified social presence as another factor determining the perception of online trust. The authors argue that "although a Website is typically devoid of actual human interaction, nonetheless, the perception that there is a social presence does in itself increase e-Trust" (Gefen & Straub, 2004, p. 417). Consequently perceived social presence on a website can be regarded as important, since there appears to be a resemblance to an actual interpersonal interaction which consumers tend to be more familiar with (Gefen & Straub, 2004).

### *2.3.2 Design and Content Quality*

Since online trust derives from user experience, familiarity and satisfaction evolved through previous transactions (Figure 1) there must be a basis for trust for those without any prior experience (Beldad et al., 2010). General determinants to communicate trustworthiness in online interactions are participants' reputation, performance and appearance (Beldad et al., 2010). While the first two determinants refer to the online organizations in general the latter is connected to the design, usability and representation of a company's website interface (Jones & Leonard, 2008; Urban et al., 2009). Schlosser, White, and Lloyd (2006) as well identify a website's design as determinant which has an impact on consumer trust and influences their intention for an online interaction. Several studies support these findings like for instance Grabner-Kraeuter (2002), Yang, Hu, and Chen (2005) and Bart et al. (2005) arguing that the design and appearance of an entity's website affects customer's trust. While Grabner-Kraeuter (2002) thereby focuses on the functionality and reliability of an e-commerce system, Yang et al. (2005) identify the design as a basis for potential customers to form a first impression of a transaction partner's trustworthiness. Bart et al. (2005) found user friendly navigation and presentation as most important variables affecting consumer trust.

Social psychology studies have shown that the physical attractiveness of items or persons does influence their perceived trustworthiness and credibility (Berscheid & Walster, 1974). Urban et al. (2009) summarizes the impact of design on trust in online environments with two main assumptions: "(1) A good-looking website […] makes users think they are browsing in a professional environment and helps foster trustworthiness of the company behind the site; and (2) Browsing in a good-looking and user-friendly website encourages users to spend more time on the website, and, the longer they stay, the higher the probability of the site gaining the consumers' trust" (Urban et al., 2009, p. 182).

A further issue is the ease of use of a technology. Davis (1989) describes this by referring to the technology acceptance model where the perceived ease of use is an important variable. In online services, ease of use can be interpreted by the navigational structure of a website (Urban et al., 2009). According to Grabner-Kraeuter (2002) an effective navigation is even one of the best ways to communicate credibility and trustworthiness. Many empirical studies support this assumption and argue that the perceived ease of use significantly affects the formation of trust (Bart et al., 2005; Flavian, Guinaliu, & Gurrea, 2006). Flavian et al. (2006) thereby revealed that low levels of usability can be the origin for technical errors which might evolve in feelings of distrust and hinder customers to use a service again. Wijnhoven, Ehrenhard, and Alink (2012) identified a service's technical architecture and service employees' motivations, characterized by their knowledge and their behavioral repertoires when responding to incidents, as possible causes for unreliability.

The information quality offered in online environments is also related to the topic. According to Liao, Palvia, and Lin (2006) customer trust in online transactions may be increased by the content quality of a website. The authors refer to the completeness, usefulness and accuracy of the offered information.

Same goes for errors in the website. Errors like incomplete pages, missing links or other inconveniences must be seen as so called “trust busters” (Urban et al., 2009). Customers are likely to trust online, if websites are free from errors and contain complete, accurate and current information (Bart et al., 2005; Liao et al., 2006). The importance of information quality lies in the intangible character of online transactions and services. Customers are not able to previously touch or test an item they desire to buy online. Accordingly, they are highly dependent on the in-depth knowledge and clear information when for instance consulting e-health sites (Liao et al., 2006; Sillence, Briggs, Harris, & Fishwick, 2007). Additional studies show a positive correlation between the overall user satisfaction which does have an impact on customer trust (Pavlou, 2003) and the visual communication quality of a website (Lindgaard & Dudek, 2003; Tractinsky, Katz, & Ikar, 2000). Having identified the importance and different ways of online trust perception, now mechanisms as tools for the active communication of trust will be examined.

### *2.3.3 Trust Mechanisms*

Participating in online transactions demands some level of dependence and is possibly fraught with a certain amount of risk. In order to reduce these obstacles, so called trust mechanism should be employed (Salam, Iyer, Palvia, & Singh, 2005). There are various attempts to foster trustworthiness and credibility and at the same time to reduce the perceived risk of participants in online transactions. A popular method for trust communication is the presence of digital certificate technologies such as trust marks or seals (Belanger et al., 2002; Kim, Ferrin, & Rao, 2008). These certificates are predominantly provided by third parties, such as banks, accountants or consumer unions (Kim et al., 2008; Salam et al., 2005). Kim et al. (2008) conclude on the topic “(t)he purpose of trusted third-party seals is to help reduce consumers' perceived risk in electronic commerce, provide assurance to consumers that a Website discloses and follows its operating practices, that it handles payments in a secure and reliable way, that it has certain return policies, and/or that it complies with a privacy policy that says what it can and cannot do with personal data it has collected online” (Kim et al., 2008, p. 550). These assumptions apply not only for e-commerce websites, but equally for online marketplaces, social media platforms and any other kind of online presence.

However the opinions on the actual effectiveness of such a trust mechanism vary in the literature. Kim et al. (2008) for instance were able to demonstrate that third-party seals had no impact on an online user’s trust perception and thereby correspond with several other scholars arguing that assurance seals have no significant influence on neither the user’s trust nor the intention to engage in an online transaction (Belanger et al., 2002; McKnight, Kacmar, & Choudhury, 2004). Conversely their counterparts argue that the presence of third-party seals and trust marks does have a significant impact on trust in online transactions (Aiken & Boush, 2006; Wu, Hu, & Wu, 2010).

Different mechanisms, upon which lies the focus of this thesis, are so-called reputation systems. The basis of the mechanism is grounded in the ancient history of human society and can be circumscribed

with the impact of word of mouth (Dellarocas, 2003). Reputation systems are aiming at the collection of evidence regarding attributes of individual items, the aggregation of the results and the representation of these aggregated results by so called reputation scores (Sun & Liu, 2012). Mostly, those systems use the feedback of its participants in form of reviews or ratings for credibility and trustworthiness (Sun & Liu, 2012). Research has identified a significant impact of ratings and reviews on online users' behavior and thereby on the willingness to get into an online transaction (Ba & Pavlou, 2002; Y. Chen & Xie, 2005; Pavlou & Gefen, 2004; Poston & Speier, 2005). The objects of trust can be divided in products, services, businesses, users and basically any kind of digital content (Sun & Liu, 2012). According to Jøsang et al. (2007), a reputation system may refer to three types of evidences: Direct observations based on business employees' opinions; expert opinions provided either voluntarily or for a fee; and feedback provided directly by users. Thereby research comprehensively examined the impact of expert reviews and online recommendation systems (Y. Chen & Xie, 2005; Gretzel & Fesenmaier, 2006) which are both considered as reliable, but connected to more costs when used for a large number of items/services (Jøsang et al., 2007). More recent research examines the topic of reputation systems with feedback given directly by e-commerce customers or participants of online transactions (Forman, Ghose, & Wiesenfeld, 2008; Smith, Menon, & Sivakumar, 2005). These include so called star rating systems and free text user reviews (Hu, Bose, Koh, & Liu, 2012) as can be found on common online marketplaces and retailing websites like for instance *Amazon* or *Ebay* (Mudambi & Schuff, 2010). Such customer reviews and ratings already showed some positive impact on sales behavior of online customers (Chevalier & Mayzlin, 2006; Clemons, Gao, & Hitt, 2006). The researchers argue that predominantly positive ratings and reviews positively impact the interest in items and services and as a consequence enhance customers' purchase intention. Studies showed that companies intentionally provide their product information to online discussion forums to proactively animate users and customers to spread the word about their services or products (Godes & Mayzlin, 2004; Mayzlin, 2006).

At the same time this knowledge makes user feedback less reliable than the other two types of evidence defined by Jøsang et al. (2007). According to Houser and Wooders (2006) the growing influence of reputation systems on customers' purchasing decisions increase the incentive for manipulations. For this reason some authors classify user feedback as a highly influential, but also least reliable source of evidence (Sun & Liu, 2012). Especially when it comes to online reputation systems as ratings and reviews a willingness to exploit online users' trust with manipulation and fraud can be identified. With this the purchase intentions, the willingness to engage in a transaction or the own competitive advantage over the competition can be affected (Houser & Wooders, 2006; Hu et al., 2012; Jøsang et al., 2007; Sun & Liu, 2012).

Thereby, the concentration on reputation systems arises from the power of electronic word of mouth on online transactions (Benedicktus et al., 2010; De Maeyer & Estelami, 2011; Poston & Speier, 2005;

Salam et al., 2005). Many online entities identified ratings and reviews as a new tool for marketing (Dellarocas, 2003) and in various cases strategic manipulations occur with the effort to influence user behavior (Hu et al., 2012). Zhang, Bian, and Zhu (2013) identified little time and effort in order to increase an entity's online reputation as a main driver for committing fraud. Since fraudulent actors focus mostly on quick success and instant benefits they are keen on increasing their reputation extremely fast. Hu et al. (2012) for instance define "review manipulation as vendors, publishers, writers, or any third-party consistently monitoring the online reviews and posting non-authentic online reviews on behalf of customers when needed, with the goal of boosting the sales of their products" (Hu et al., 2012, p. 674). Manipulations hereby refer to posted information that doesn't reflect real customer experience. Especially online auction marketplaces are often affected by fraudulent actors which artificially improve their ratings by trading favorable reputations directly on such platforms (Dini & Spagnolo, 2009; Zhang et al., 2013). By implication, these platforms relented and the public feedback market has strongly contained, but the issue of trust fraud still exists (Zhang et al., 2013).

The potential of fraud within reputation systems can be further seen by the establishment of businesses which exclusively concentrate on the promotion or downgrading of online reputation through artificial feedback in order to gain profit (Sun & Liu, 2012; Zhang et al., 2013). Sun and Liu (2012) recognized three major approaches of reputation systems attacks. The impostors obtain information on the target which reputation score should either increase or decrease, then those companies distort the evidence collection process by the integration of manipulative feedback and in extreme cases the attackers try to rig the algorithm aggregating the evidence. The modified algorithms misclassify honest user feedback as dishonest, at the same time dishonest feedback as honest and yield at inequitable high or low scores for the target of the attack.

The effects of such measures are detrimental for the trustworthiness and credibility of any online actor or entity. Coordinated distortions of reputation scores may mislead consumers in their purchase decisions (Hu et al., 2012) and make the impact of reputation systems basically worthless by undermining "users confidence about reputations-centric-systems, and (…) eventually lead to system failure" (Sun & Liu, 2012, p. 88).

## 2.4 Action in Form of Interactions or Transactions

The possible result of perceived trustworthiness may lead an online participant to some sort of action. It can be argued that this action would not take place without the existence of the variable trust. Why trust plays such a dominant role for individuals performing online activities, is described by Ridings, Gefen, and Arinze (2002) with the absence of direct interpersonal contact and visual cues as a consequence of the virtual nature of such activities. But the targets of trust in online activities as well have to present themselves as trustworthy parties in order to minimalize the perception of risks (Haas & Deseran, 1981). There are two main sources of risk connected to online transactions or interactions;

those being the risk of monetary damages and the threat of misusing and manipulating someone's private data (Beldad et al., 2010). But in most transactions such as economic exchanges not everything that creates the situation of risk can be verified and eliminated beforehand. Therefore the necessity of trust arises (Tullberg, 2008). This refers back to trust as a mediating variable between the online environment and online activities.

In marketing research, customers need to decide to what extend they can trust a company in order to purchase its products or engage in any sort of commitments (Doney & Cannon, 1997). The same goes for activities in an online environment. Even though online activities have to be characterized as faceless or intangible, the general acceptance of online transactions and interaction constantly rises (Beldad et al., 2010). The reason for this growth can be connected to the monetary or effort reducing benefits online activities provide. But the effect equally has to be addressed to individuals' trust into online transactions or interactions and the technology behind them (Beldad et al., 2010).

Trust has been identified by research as an enabler influencing online participants' behavior (Urban et al., 2009). It is an enabler for a customer's decision to buy something online, to engage in a transaction with a website or e-vendor, to interact with an online community or basically any other sort of online activity (Urban et al., 2009).

## 2.5 Impact as Common Benefit

The lack of trust is regarded by literature as one of the biggest obstacles for customers to engage in online transactions. In order to create a competitive advantage on the internet, participants are obligated to create a climate of trust and advised to maintain this status (Gefen & Straub, 2004; Murphy & Tocher, 2011; Shankar et al., 2002; Urban et al., 2009). Without a clear understanding of the importance of trust in online business relationships, it might be difficult to stay competitive (Shankar et al., 2002). Creation and maintenance of trust determines the usability, sales revenues and profitability of an online entity. Shankar et al. (2002) further argue that with more stakeholders having access to various options and huge amounts of information on the internet it becomes critical for firms to gain and retain their current and potential customers' trust. Organizations can position themselves better and achieve superior firm performance when creating trust by knowing their stakeholders needs (Shankar et al., 2002). Urban et al. (2009) identify trust as a crucial component of an organizations e-business strategy, directly correlating with a firm's competitive advantage. Jarvenpaa et al. (2000) as well as Yoon (2002) support these findings by arguing online trust affects the risk perception, attitude and willingness to buy items online.

Upon that the customer to customer segment with social media networks like Google+, Facebook and Twitter grows on importance. Those sites are among the most favored websites on the internet (Lua et al., 2011). On such platforms new forms of transactions evolved, circumscribed as social commerce

(Haefliger, Monteiro, Foray, & von Krogh, 2011; Ji & Wang, 2011; Jones & Leonard, 2008). Online entities thereby serve as intermediary for both parties and therefore trust is not only crucial between the traders but as well in the organization that provides the website (Ji & Wang, 2011; Shen & Zhong, 2008).The lack of trust and a risky environment were recognized as the main reasons which could prevent the development of social commerce (Ji & Wang, 2011). This visualizes the importance of online trust since the attraction of new customers and in the best case retaining them is critical for the success and the major goal for any online business (Jarvenpaa & Toad, 1996; Reichheld & Schefter, 2000).

## 2.6 Learning

### *2.6.1 Experience*

One important determinant, related to the impact of trust on e-commerce websites and online marketplaces, is the role of user experience (Corbitt, Thanasankit, & Yi, 2003; Gefen, 2000; Gefen, Karahanna, & Straub, 2003). Referring to Figure 1, Boyd (2003) argues that the inevitability of a first-time situation in any sort of online activity makes trusting difficult. As a consequence the approach suggests that missing experience of users with online transactions and online companies leads to a different level of trust than for instance experienced users do achieve (Beldad et al., 2010). There are several empirical studies related to the impact of user experience on the trustworthiness of online offerings. Metzger (2006) focused on customers' trust perception in dependence of their experience with online marketplaces and commerce. Web usage competency can thereby be understood as the skills in using computer technology. The findings were compared with the experience those users have with traditional commercial exchanges. Findings showed that users with a higher grade of online experience demonstrate lower levels of perceived risks and are more likely to trust transactions in an online environment. According to Corbitt et al. (2003) customer trust levels are as well assumed to be influenced by the customer's web experience. The authors identified a positive relation between the degree of trust in a website and the level of experience the customer shows with web interfaces. As a result the experience level influences the user's willingness to trust the technology (Internet as a whole) and could as well have an impact on the perceived trust in e-commerce and online marketplaces.

Even though literature predominantly identifies this positive relationship between web experience and the level of online trust, some studies show deviating results. Aiken and Boush (2006) recognized a positive relation between internet experience and online trust for new and intermediate users, but were able to reveal a negative correlation for more advanced users. They demonstrated the experience-trust relationship with an inverted U shape. From this it follows that user's trust in online activities and entities increases in an early phase of usage when the experience as well increases. Later on trust rather declines with a higher level of experience caused by more knowledge concerning privacy and security concerns (Aiken & Boush, 2006).

### *2.6.2 Familiarity*

A different way how users subjectively reduce uncertainty and increase the level of trust is in the feature of familiarity (Komiak & Benbasat, 2006). Luhmann, Davis, Raffan, and Rooney (1979) define familiarity as a precondition for trust. According to Gefen (2000) familiarity and trust are both complexity reducing methods and therefore complement each other. This relation is based on the assumption that trust in individuals and entities evolves out of an expectation of trustees towards them. Gefen (2000) hypothesized in an experiential survey that "(i)ncreased degrees of familiarity with an E-commerce vendor and its procedures will increase trust in the vendor" (Gefen, 2000, p. 729). The survey showed a significant influence of familiarity on online trust and identified familiarity as a determinant for user intentions on the internet such as the intention to purchase a product online. Mollering (2006) supports the opinion that familiarity is essential in trust building processes.

### *2.6.3 Satisfaction*

A further determinant related to experience and familiarity is customer satisfaction. There exists a positive relationship between trust and customer satisfaction (Pavlou, 2003). This assertion derives from the observation that customers who are satisfied with their online experience are likely to trust their interaction partner for a potential second transaction. Yoon (2002) and Flavian et al. (2006) support the argument and reveal with empirical studies that customers' satisfaction in an online transaction indeed determines their trust in the entity they had the online interaction with. Yoon (2002) further adds that satisfaction does not only have an impact on customer trust, but beyond that on the familiarity and evokes greater usage.

## 2.7 Implications

Trust is a multidimensional concept in the physical world as well as in an online environment. Still the majority of scholars identified trust as crucial for any kind of interpersonal, business or online relationship or transaction. Despite the similarities, there are however some fundamental differences in the cognition of trust, regarding traditional and online environments. First the object of trust differs and consequently the traditional cues of trust and reputation in the offline environment are missing online. A second difference is the ease of information sharing and communication towards a global community, whereas in the physical world information exchanges are mostly limited to local communities.

Users' *experience*, *satisfaction* and *familiarity* were recognized as factors determining trust in an online environment. Furthermore the *design* and the *information quality* do as well affect the perceived trustworthiness towards an online entity. *Privacy* and *security* features were considered the basis for actors to engage in any sort of online interaction and refer to early works on online commerce and transactions. More important for this thesis are the insights gathered on the active communication and perception of trust by so called *trust mechanisms*. Since online trust as mediating variable

acquired a position as key driver for success and the number of internet affine users constantly grows, research identified a swap from a rather passive assurance of safety to mechanisms actively promoting trustworthiness and credibility.

Online reputation systems are discussed among scholars as sort of trust mechanism with primarily positive impact on consumers' perception of trust. These systems are based on feedback by ratings and reviews. Thereby direct user feedback was identified as highly influential, but as well endangered by fraud and therefore less reliable. Many of the reviewed scholars, even very recent publications are based on implications which are five to ten years old. User perceptions may have changed since people are much more internet affine than a few years ago and internet transactions grow constantly in significance, especially with the evolving field of social commerce.

The aim of the thesis is to dig deeper into trust mechanisms and gain an up to date look on the impact of a reputation system on the perception of trust of online users. Thereby the implementation of such a system on an online marketplace serves as a basis for research. Furthermore the intention is withal to assess the danger of trust fraud and the possible need for more sophisticated fraud prevention by evaluating the current prevention mechanisms. In the next chapter however, the focus lies upon the principles these systems are actually based on. Furthermore, fraud prevention mechanisms are approached from a structural perspective.

# Chapter 3: Reputation Systems as Online Trust Generator

According to Jøsang et al. (2007) there are two main purposes research on trust and reputation systems should focus on. First it should concentrate on the search for adequate online substitutes for the traditional approaches used in the physical world towards trust and reputation. Thereby new information elements should be detected and applied, especially geared to specific online applications, which are able to derive measures of reputation and trust. The second purpose is devoted to the creation of efficient systems for gathering such information by taking advantage of IT technology and the Internet. The resulting measures should serve to improve the overall quality of online markets and support online users' perception of trustworthiness and their decision making process.

In case of reputation systems Resnick et al. (2000) define three fundamental properties as necessity to operate:

(1) Every online interaction should follow the expectation and possibility of a future interaction. Therefore the longevity of online entities has to be guaranteed. In practice this means for instance agents should not be able to fiddle with their identity in order to erase the connections to their past (potentially fraudulent) behavior.
(2) User feedback, in form of ratings or reviews, about past online interactions has to be agglomerated and made available. This rather essential property is however dependent on the participants' willingness to provide feedback. Therefore reputation systems must yield some sort of incentive.
(3) The reputation system must guide the decision making process for current interactions based on the feedback (ratings/reviews) of past interactions. Herby the property refers to the actual usability of a reputation system and in which way interacting participants respond to it.

Additionally to those properties, Dingledine et al. (2000) extended the requirements by (4) *robustness against attacks*. Beyond the above mentioned properties the authors determined a reputation system's capabilities to resist attempts of any entities to influence or manipulate reputation scores as a further essential feature.

In what follows, the principle of *trust transitivity* upon which most reputation systems rely is illustrated in detail. Furthermore reputation systems are approached from a more technical perspective. Thereby *reputation network architectures* are presented. The chapter closes with a reflection on the technical perspective of *fraud prevention*.

## 3.1 Trust Transitivity

Since trust is rather vague and difficult to determine, the perception and communication of trust by reputation systems are based in the broadest sense on the principle of trust transitivity (Jøsang & Pope, 2005). The idea behind this simple principle is illustrated in Figure 2. Trust is thereby derived from a

transitive trust path. If an entity [A] trusts another entity [B] and this entity [B] then again trusts an entity [C], [A] will be able to derive a measure of trust in [C]. This process is based on the assumption that [B] refers [C] to [A] (Jøsang et al., 2007). The exchange order is illustrated with the numbers in the brackets in Figure 2. In order to function, there are so called semantic constrains that have to be taken into account when relying on the transitive trust deviation. This means the entities *must* trust each other in the above shown order to consider the principle as valid. It is not enough if [A] trusts [B] but [B] does not trust [C]. Such a framework can be applied as well on a multidimensional level with several parallel trust paths as demonstrated in Figure 3.

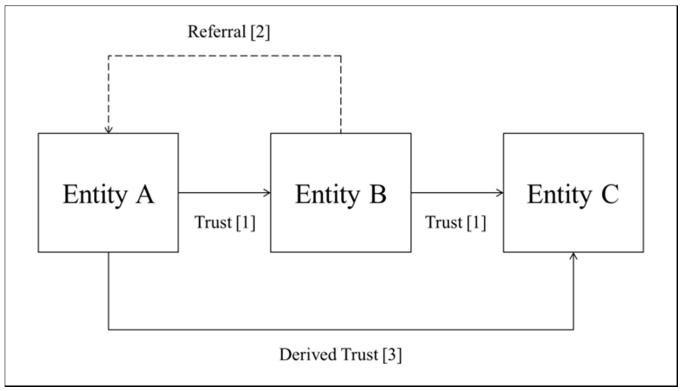



Figure 2: Trust Transitivity Principle (Own illustration based on Jøsang et al. 2007)

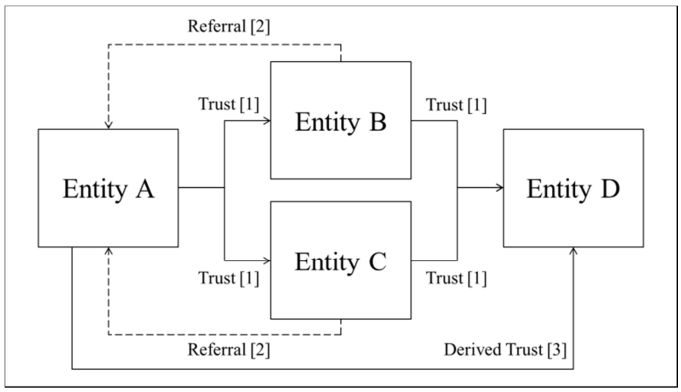



Figure 3: Parallel Transitive Trust Chains (Own illustration based on Jøsang et al. 2007)

Jøsang et al. (2007) explain the relations with a practical example. The initial situation may be the same as in Figure 2. Let's assume a person [A] needs some maintenance done in its household and asks person [B] to recommend a good craftsman. [B] recommends the craftsman [D] to [A]. In order to be sure [A] wants to get a second opinion and asks person [C] about the craftsman [D]. If both paths refer to [D] as trustworthy a strengthened perception of derived trust from person [A] to person [D] might evolve. Here again the framework is based on the assumption that both [B] and [C] refer [D] to [A] and all the direct trust paths are guaranteed. This concept colludes well with the idea of electronic word of mouth by Dellarocas (2003).

Reputation systems incorporate the idea of trust transitivity but, also rely on a broader view. Therefore they are typically based on public information to determine a community's general opinion. Thus, the

impact of trust transitivity for reputation systems can be described with one party trusting another party on behalf of the reputation score of some remote party. Consequently the system produces a party's (public) reputation score based on the opinions of a community (Jøsang & Pope, 2005).

### 3.2 Reputation Network Architecture

According to Wijnhoven (2012), the design process of information service infrastructures in general can follow the *Leibnizian* inquiring system. Thereby the focus lies upon rationalism and logic. The system further implicates that knowledge can be passed on by the representation of causal understandings such as specific predictions or explanations. If such a model of casual relations is well-outlined and consistent it allows logical reasoning about the elements it refers to.

Additionally information service architectures may be approached from a *product-oriented* and *process-oriented* design theory (Wijnhoven, 2012). Considering the first one (product-oriented), two sub-forms can be identified. The content aspect of information services distinguishes among centralized and aggregator website architectures. Centralized website architectures thereby aim at providing its users with the right information. Aggregator architectures enable the creation of own information by users. The use-value covers the second product-oriented perspective on information service architectures. Hereby again two types can be distinguished. The so called community architecture allows community building by creating ways of interacting with content. The integrated data architecture combines and integrates data from different sources. In the process-oriented design theory the focus lies on the representation of each technical layer of an information service (Wijnhoven, 2012). The approach enables the representation of the goals of each layer of an information service. Furthermore responsible actors for each service architecture layer may be identified and functionality, necessary use-features and content of an information service can be implemented.

Reputation systems in general are divided into two different reputation network architectures (Gutowska, Sloane, & Buckley, 2009; Jøsang et al., 2007; Liu, Munro, & Song, 2010). Such architectures illustrate the communication paths of ratings and reputation scores between individuals or entities in a reputation system. The so called *distributed reputation systems* rely on decentralized solutions with no central location for participants to submit ratings or obtain reputation scores (Jøsang et al., 2007). Users consequently have to store their reputation information at individual locations and for instance provide these information on request of relying entities (Liu & Munro, 2012). For this thesis however the focus lies on the *centralized reputation system* which is most widely adopted by online entities (Liu & Munro, 2012). It further correlates to the aggregator architecture, presented by Wijnhoven (2012).

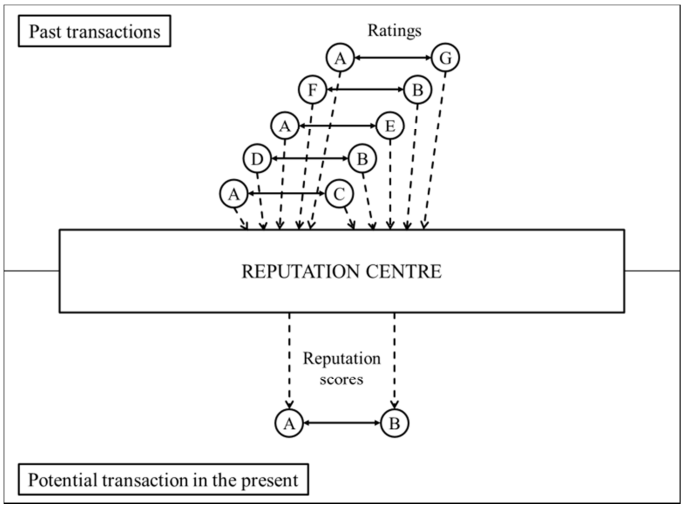



Figure 4: Centralized Reputation System Architecture (Own illustration based on Jøsang et al. 2007)

Centralized reputation systems aggregate participants' feedback of interactions or transactions and in this way, indications of the performance of a given participant are stored. The feedback is done by ratings and reviews from other members of the same community which had direct contact and experience with that particular participant. The role of the reputation center can be described with a central authority which accumulates all the user ratings and reviews, next derives the reputation scores for each of the participants and finally displays all the reputation scores publicly (in a community). The provided information of each participant is then used for instance as decision guidance whether or not to engage in a transaction with a specific party. This system furthermore implies that those transactions which occurred between reputable participants are expected to lead to more favorable outcomes than comparable ones with disreputable participants (Jøsang et al., 2007). For a better understanding the relations are clarified in Figure 4. The framework shows a possible transaction of participants [A] and [B] in the present. Both transaction partners have a history of transactions in the past. After each of those past transactions the two participants, as well as their former transaction partners, gave ratings on the performance of each other in the transaction. Those ratings of all the participants are collected in the reputation center and there (as a function of the received ratings) the reputation scores of each participant are continuously updated. Finally the scores are presented to the online audience and the participants can decide based on reputation scores if they want to cooperate with a particular transaction partner or not.

In general Jøsang et al. (2007) defines two fundamental aspects of centralized reputation systems:

(1) In order to provide reviews and ratings about partners in past transactions to a central authority, the system needs centralized communication protocols. Those protocols serve as well to obtain a potential partner's reputation scores back from the central authority.
(2) Secondly the central authority has to make use of a reputation computation engine to be able to derive reputation scores of each agent which are based on the agent's received ratings and reviews. In addition the scores could be based on further information if possible.

Going a step back, reputation systems were recognized as mechanisms communicating trust to online users. Thereby those systems differ from traditional trust mechanisms which were earlier identified as security and privacy features. It is necessary to clarify this point in order to understand the difficulties which might erase when implementing reputation systems. The difference was first described by Rasmusson and Jansson (1996) with hard and soft security. Hard security mechanisms are referring to security and privacy features which prevent unauthorized access to resources and consequently protect users from malicious practices. However many situations require protection from parties who offer resources on the internet. False or misleading information provided by deceitful actors won't be uncovered by traditional mechanisms. It is here where the so called soft security mechanisms (as reputation systems) come into place to offer protection against such types of threats.

### 3.3 Fraud Prevention

The reputation system itself however can become victim of fraud. The most obvious problem, connected to a trust communication system which is dependent on feedback, form unfair or manipulated ratings and reviews. Consequently the prevention and filtering of fraudulently positive or negative ratings is a fundamental issue when these systems serve as decision support for online users (Jøsang et al., 2007). In such a case the party which relies on the reputation scores of others cannot estimate the authenticity of feedback given on a subjective basis.

According to Wijnhoven (Wijnhoven, personal communication, March 14, 2013) the issue of credibility and trustworthiness of reputation scores can be spread into multiple layers: The first level concentrates on the authenticity and reliability of online *representations* for any kind of products, goods or services and thereby on indicators of bias, untrustworthiness or manipulations. The next layer deals with *actual rating data* and how actors evaluate those ratings and feedback. On this level mechanisms like sentiment analysis or opinion mining come into place in order to scrutinize the credibility of given feedback which may be biased. Upon that there is an even higher level of trust which refers to the actual *assessment of ratings* using all the big data and analytics on a specific market. In other words who or what guarantees that the evaluations and gathered data from the two lower levels are reliable? Is it reasonable to trust unhesitatingly reports or let's say the ratings of the ratings?

The focus of this thesis primarily lies upon the first two layers by reviewing the effectiveness of given fraud prevention mechanisms. In case of reputation systems, these mechanisms are predominantly arranged as fraud prevention filters in the participant's feedback process to avoid bias from unfair or manipulated ratings and reviews (Figure 5).

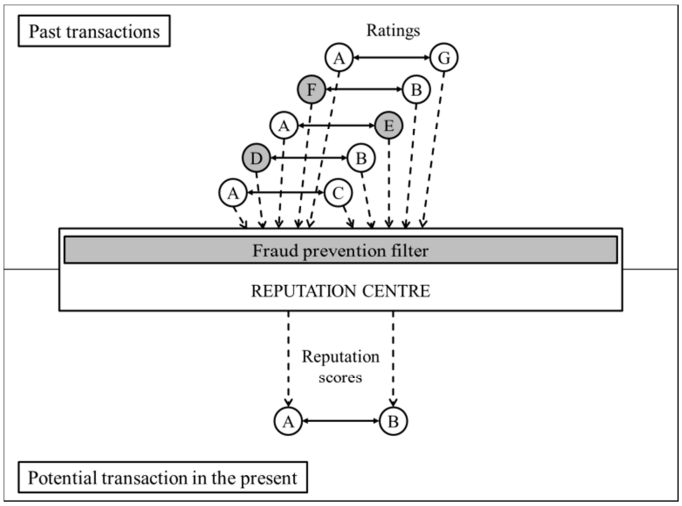



Figure 5: Fraud Prevention Filter (Own illustration based on Jøsang et al. 2007)

The fraud prevention filter (Figure 5) should detect manipulative or fraudulent feedback of participants [D], [E] and [F] in order to prevent dilutions in the reputation scores of the potential transaction partners [A] and [B]. Thereby fraud prevention methods can be grouped broadly into two categories: endogenous and exogenous discounting of fraudulent ratings. The first category which is supported by Whitby, Jøsang, and Indulska (2004) as well as M. Chen and Singh (2001) identifies fraudulent feedback by analyzing and comparing reputation scores. The assumption behind the approach is that manipulated ratings and reviews may be located by some sort of statistical properties. Supporters of the second category concentrate on methods which use a rater's externally conditioned reputation to determine the trustworthiness and credibility of the actor's ratings. The assumption behind this approach is that users with a low reputation tend to give rather unfair ratings themselves and vice versa (Jøsang et al., 2007). More recent approaches combine the two methods in order to build more robust reputation systems (Duan & Liu, 2012; Sun & Liu, 2012). Duan and Liu (2012) for instance introduce a so called restrictive reputation model to foster the robustness of a system. Their approach assigns different weights to customer and vendor ratings according to the rating frequency and the monetary value of a transaction in dependence of specific customer and vendor properties.

In order to engage the problem from those perspectives a two experiment design is chosen for the research. The first experiment will focus on the implications which can be derived from the implementation of third party dealer ratings on a marketplace website and thereby on the user perception of trust. The second one will exclusively focus on the danger of fraud and manipulations in combination with user feedback in form of ratings and free text reviews. A detailed description of both approaches is presented in the following *Methodology* section.

# Chapter 4: Methodology

## 4.1 Research Design

The purpose of this section is to give an overview of the research methods, using the so called research process onion by Saunders, Lewis, and Thornhill (2009). The approach is thereby divided into 5 layers which are: Research Philosophy, Research Approaches, Research Strategies, Time Horizons and Data Collection Methods. For this thesis the respectively chosen approaches are summarized in Table 1.

| **Research Process Onion** | | | | | |
|---|---|---|---|---|---|
| *Research Philosophy* | | | | | |
| Positivism | | Realism | | Interpretivism | |
| *Research Approaches* | | | | | |
| Inductive | | | Deductive | | |
| *Research Strategies* | | | | | |
| Experiment | Survey | Case Study | Grounded Theory | Ethnography | Action Research |
| *Time Horizons* | | | | | |
| Cross Sectional | | | Longitudinal | | |
| *Data Collection Methods* | | | | | |
| Sampling | Secondary Data | Observation | Interviews | Questionnaires | |

Table 1: Research Process based on Saunders, Lewis & Thornhill, 2003, Chapter 4, p. 83

In a positivist research philosophy the researcher takes the role of an objective analyst. The research is marked by scientific reasoning and attributes of this philosophy are highly structured methodologies and quantifiable observations such as statistical analyses (Remenyi, Williams, Money, & Swartz, 1998). Generalizations thereby serve to reduce complexity and enhance transparency. Hand in hand with the philosophy of positivism goes the deductive research approach which is based on developing a theory accompanied by hypotheses (Hussey & Hussey, 1997). Those hypotheses are then tested with appropriate research strategies. To draw conclusions, in this thesis an experiment research strategy is chosen. The experiment qualifies as a classical form of research whereby defined theoretical hypotheses are tested using a selection of samples of individuals from known populations as dataset. Since the time horizon of this study is rather limited for the period of approximately four months a cross sectional approach seems to be more promising than a longitudinal alignment. Considering the data collection methods, mostly observations according the experiments and secondary data in the case of the theoretical background (Literature Review) are used as sources.

## 4.2 Research Framework

The overall goal of this thesis is to elaborate if *the implementation of a reputation system fosters the perception of trust on online marketplaces?* From *Chapter 3* as well as the findings derived by means of an extensive literature review, reputation systems can be identified as contingencies to communicate and affect the perception of an individual's online trust. To elaborate a relation among a reputation system and the perception of users' trust in a B2C context the following research model was developed and serves as framework for the study (Figure 6). With reference to the trust model (Figure 1) presented in *Chapter 2* the research model below is located somewhere between the variables *online environment*, *trust* and *action*.

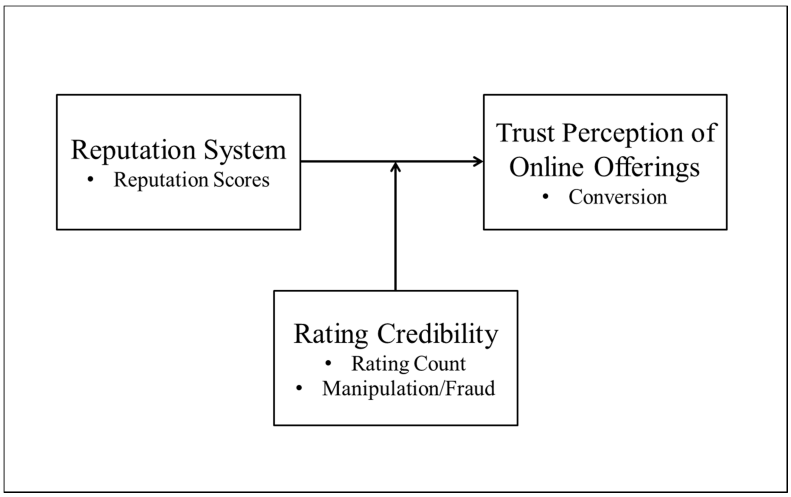



Figure 6: Research Model (Own illustration)

The basic assumption of the study describes the impact of a *reputation system* on an online user's *trust perception of online offerings*. Since a reputation system consists of several parts as indicated in *Chapter 3* the general impact of the system on trust perception is measured by the perceivable ones on an online marketplace; such as the reputation score. Direct measures for trust cannot be applied due to the multidimensional character of the variable. Therefore conversion is chosen as an appropriate proxy for the trust perception of online offerings. The relationship of the two variables is furthermore moderated by the *rating credibility*. A basic measure for this moderating variable is the rating count a reputation score consists of (Poston & Speier, 2005). Upon that, the possibility of fraud and manipulations determines the rating credibility. In what follows the research hypotheses of this study are illustrated in detail.

**H1**: The implementation of reputation scores has a positive effect on click conversion.

**H2**: The implementation of reputation scores has a positive effect on contact conversion.

Both Hypotheses focus on customers' trust perception of online offerings in dependence of the visible implementation of reputation scores on an online marketplace. Click and contact conversion is thereby used as a proxy for trust perception. As indicated in the theoretical part of the thesis, trust has a strong

impact on online activities, such as purchase intentions (Ba & Pavlou, 2002; Gefen et al., 2003; Kim et al., 2008) which can be measured by conversion rates such as Click-Through-Rates (CTRs) or Contact-Conversion-Rates (CCRs) (Conroy & Bearse, 2006; Perdikaki, Kesavan, & Swaminathan, 2012; Sismeiro & Bucklin, 2004). This is further clarified by Morys (2011) and his seven layer model of conversion (Figure 7). Conversion is thereby directly affected by seven determinants, one of which is defined as trust. The implementation of a reputation system on a website exclusively affects the trust variable whereas the remaining ones stay untouched. Differences in conversion can consequently be directly interpreted as differences in trust perception and support conversion as a valid proxy for trust perception.

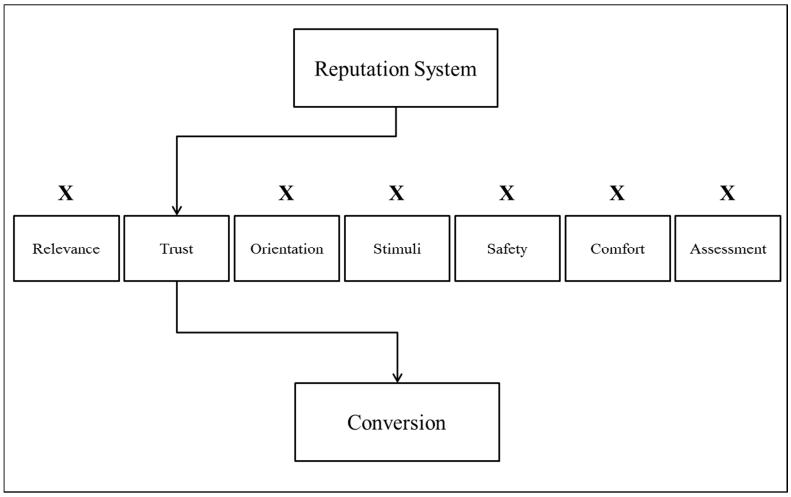



Figure 7: Seven Layers of Conversion (Own illustration based on Morys 2011)

The expectation behind Hypotheses **1** and **2** is to identify a noticeable difference of online users' trust perception towards special dealer offerings. The idea is to compare the activities on the website with an experiment. For a considered period of time the same online offerings are displayed with and without the presence of reputation scores in order to elaborate coherences. Comparing the CTRs of the online audience enables to draw conclusions on the interest of online offerings evoked by trust perception in dependence of reputation scores. CTRs originate from the domain of web-marketing and are defined as the proportion of viewers who click on an online advertisement or banner compared with the overall impressions of the specific item (Baltas, 2003). For this experiment the purpose of the ratio will be slightly modified, since the point of interest is not the effectiveness of an advertisement but the willingness of customers to get further information on an offering based on the perception of trust initiated by dealer ratings. The rate is calculated as follows:

$$CTR = \frac{Clicks}{Impressions} * 100$$

The number of clicks of one special offering divided by the times the offering was displayed in a specific timeframe. For the experiment this means comparing the CTRs for dealers with and without the displaying of reputation scores and to measure the variance of the results. The effects of reputation

scores on trust perception can as well be measured by evaluating if the presence of reputation scores fosters users' willingness to engage in an interaction. Such an analysis results in Hypothesis **2**.

Thereby again a comparison among a representation with and without ratings on the website has to take place. Since CTRs might appear too weak for such an approach, CCRs are applied. The purpose of the ratio is to measure the number of certain specified goals compared with the overall number of visits on a webpage (Perdikaki et al., 2012). The desired action for this study is defined as a user's active attempt to get in touch with a dealer on the marketplace. To do so a user has to search for a so called contact box on the dealer's webpage, enter her or his contact information and add an request for the dealer which than will be send in form of an email. The metric is generated with the following formula:

$$CCR = \frac{Email\ Requests}{Visits} * 100$$

Consequently CCRs are measured as percentage of visits which will result in direct email requests for the dealer. The ratio appears to be particularly suitable since even small variations between the CCRs can have measureable impacts like for instance increased sales revenues (Sismeiro & Bucklin, 2004) and can be further positively associated with an increase in customer loyalty (Conroy & Bearse, 2006).

> **H3a**: A higher reputation score induces higher click/contact conversion.
> **H3b**: A lower reputation score induces lower click/contact conversion.

Hypotheses **3a** and **3b** build upon the assumptions of **H1-2** and address a meaning to the height of a reputation score. It is questioned if the trust perception of online offerings might differ not only by the implementation of a reputation score, but additionally by the height of such a score. This implies an active examination by users of the actual reputation score of a dealer. By this means the influence of the score on a user's decision which offer to choose or with whom to interact is examined. If users actively dispute with a dealer's ratings, it is rational to suggest that dealers should receive higher or lower attention in dependence of their reputation score (Mudambi & Schuff, 2010). In order to measure the possible deviations of trust perception in dependence of the reputation score, again conversion is chosen as a proxy. Therefore the dataset of the implemented reputation system is additionally checked for average clicks and email contacts for each dealer offering and compared in dependence of the numerical rating score. Users' reactions on the ratings are thereby collected and examined. The goal is to identify statistically significant distinctions among the clicks and email contacts in dependence of the rating scores and eventually identify specific behavioral patterns.

> **H4**: A higher rating count has a positive impact on the relation of reputation score and conversion.

A further perceivable part of a reputation system is the rating count upon which a reputation score is based on. Hypothesis **4** therefore does not aim at the effect on trust perception, but is a measure for the

rating credibility of a reputation score. It predicts that users are aware of the existence of dealer ratings by reacting on the total number of ratings a vendor's reputation score consists of. Consequently the rating count moderates the relationship between a reputation system (Reputation score) and the trust perception of online offerings (Conversion). The idea is that amongst those offerings that possess a reputation score, a difference in dependence of the rating count is measureable. This difference might be measurable between reputation scores which are based on just one rating and others which consist for instance of ten and more. As Poston and Speier (2005) argue a large number of ratings would strengthen the validity of a reputation score, while a small amount of ratings would serve as a basis to discount the score and its credibility. A possibility to measure such a credibility enhancing variable is to stay close to the prior Hypotheses. Again, the average clicks and contacts on special offers are evaluated and compared, but in dependence of the rating count.

> **H5**: Reduced rating credibility by *fraud/manipulation/spam* has a negative effect on the strength of the relationship among reputation score and conversion.

The second experiment of this thesis scrutinizes the actual performance and robustness of a given reputation system since all the above described assumptions are dependent on the accuracy of the displayed reputation scores and reviews. As **H4** indicated, the rating count is a measure for rating credibility. However, referring back to the requirements for a reputation system by Dingledine et al. (2000), especially the system's robustness against attacks has thereby particular significance. The approach colludes with the feedback loop of Trust-Action-Learning presented in Figure 1. Users must be sure of the content validity in order to base their trust perception of online offerings upon the reputation score of a vendor. Furthermore the correctness of the scores has a bearing on the trustworthiness and credibility of the whole online marketplace (Kambil & Van Heck, 1998). Fraud and fake filters for reputation systems are essentially more reliable then maybe a few years ago, but there might still be loopholes the current prevention techniques are not able to detect without problems.

The special focus thereby lies upon free text user feedback, since the often highly subjective content is difficult to classify. Users can insert their personal opinions which under circumstances might include useless, manipulative or fraudulent content (Jindal & Liu, 2007). Such feedback should not be considered as objective and possibly banned from the marketplace and vendor's representation. On the other hand extremely positive feedback could imply the attempt of a vendor to represent its reputation better than it actually is. User ratings of a different website feature will therefore serve as testing field. The free text dataset will be evaluated for potential inconsistencies by manual and automated semantic analysis techniques. The results could thereby serve as a way to make reputation systems more robust and consequently affect the trustworthiness and credibility of the whole online marketplace.

## 4.3 Data Collection

*AutoScout24*

The necessary data for the evaluation of the hypotheses was collected in collaboration with AutoScout24, Europe's biggest online car marketplace with access to around two million car offers and over 17 million users per year. The company, with headquarter and biggest market share in Germany, provides an online based marketplace for new and second hand cars, motorcycles, utility vehicles as well as car parts and components. Above that the company implements a number of services on the platform connected to the lifecycle of car purchase, car ownership and finally the disposal of a car. For whole Europe around 40,000 professional car dealers use the marketplace in order to offer their cars or services. AutoScout24 is part of the Scout24 group. A major goal of the company is to exude transparency, trustworthiness and credibility for its private and business users and thereby stand out from their competition.

One from numerous measures deployed by AutoScout24 is the implementation of reputation systems for several of their site features, especially where the marketplace acts as an intermediary between transaction partners. The service providers or product vendors can be rated by their customers to share their experience on the marketplace and as a result foster the reputation of honest vendors and identify less recommendable ones. To evaluate the effects such reputation systems actually have on users' perception of trust, several tests were implemented on the marketplace. The type of collected data has to be classified as quantitative and is used to analyze relationships between specific variables by so called quantifiable phenomena (Patton, 1990). Upon that a quantitative approach for the studies allows to test the validity of the formulated hypotheses (*4.2 Research Framework*) (Blumberg, Cooper, & Schindler, 2008). In general, it is particularly important to consider construct, internal and external validity in all kinds of research (Eisenhardt, 1989). Thereby construct validity stands for the application of correct operational measures and internal validity refers to the ability of establishing casual relationships and thus indicating the legitimacy of the results (Voss, Tsikriktsis, & Frohlich, 2002; Yin, 2003). In order to generalize and eventually apply the findings to different cases of a population in real-life context, external validity is required (Babbie, 2012). With the aid of the collaborative company both external validity and the generalizability of research findings are enabled by large sample sizes of the studies. Moreover, the reliability of the research has to be guaranteed. It refers to the probability of repeating the same study repeatedly and thereby obtain similar results (Yin, 2003). For this study reliability is assured by the customer rating/review forms which serve as basis for data collection in both cases. In the fashion of questionnaires the rating forms are easily standardized (all respondents are exposed to the same form in the same way) and as a consequence easily repeatable. Hereinafter the test setting and applied dataset for both experiments are described in detail.

The focus of the first hypotheses lies on the impact of a reputation system on the trust perception of online offerings. The approach thereby does not concentrate on technical details of reputation systems. For a specified timeframe, information from one of the company's online marketplaces are gathered and the underlying database is used for further evaluations. To investigate the mentioned relationship a specially developed experiment was implemented on one of the company's second-hand-car online market places. Its Dutch online presence emerged thereby as preferable target for the test set-up. Since an own solution of a reputation system was not established yet, third party dealer ratings of one of the company's cooperation partners were used. The partner in question is a well-known Dutch automotive association for all sorts of motorized vehicles. Car dealers, service stations or workshops which are members of the association, can be rated by their customers on the cooperation partner's web presence. Thereby it is possible to share experiences with a dealer by a free-text comment. Furthermore, users can rate the dealer's service quality, expertise and price-quality-relationship of a transaction on a scale from 1 (lowest score) up to 10 (highest score). Above that, an option to recommend the dealer is included. An exemplary rating form is presented in the appendix (A 1). The rating along with the automotive association's logo is implemented into the offerings of dealers on the marketplace. The impact of third party ratings on online users' behavior was already scrutinized by Benedicktus (2011) as well as De Maeyer and Estelami (2011) and found as positively perceived by the online audience. The probability of significant differences in comparison to in-house ratings is consequently rather low. A big advantage with third party ratings is that rating data needs not to be collected first and therefore the time scope fits better in the overall timeframe of the Master project. A further advantage is the rather small probability of fraudulent and manipulative content. The data is filtered by the third party organization and it has to guarantee for content accuracy. On the downside the sample size is going to be reduced in this way, since not the entire listed car dealers are members of the association.

For the further evaluations on conversion in dependence of height of reputation score, the information for each Dutch dealer offering stored in the company's databases is evaluated. The database supplies among others information on *make*, *model*, *price* or *reputation score* for each offering in dependence of a specific dealer. Furthermore, the database calculates average clicks and contacts per offer per day for a specified timeframe. These information are used to draw inferences on the relationship between height of reputation score and conversion as well as reputation count as credibility indicator. The total number of available offerings is displayed by Table 2.

Of the [redacted] have their rating by the association implemented on the web presence (Table 2). This implicates that the data collection timeframe is scheduled for around 10 days in order to generate enough online traffic to gather a significant amount of impressions for the experiment. The target amount of impressions for the data collection period is estimated at approximately one [redacted] per version of the split-test

experiment (Figure 9). The needed metrics which are necessary for the evaluation of results as for instance the number of impressions, number of visits or click-rates on specific items are collected with special web-analytic tools. In this case Adobe® Omniture web analytics software is used. Table 2 gives on overview on the test sample for the first experiment.

| **Test Sample One: Overview** | | | |
|---|---|---|---|
| Data collection timeframe | [redacted] | | |
| | [redacted] | [redacted] | [redacted] |
| Dealers overall | [redacted] | [redacted] | [redacted] |
| Dealers without membership | [redacted] | [redacted] | [redacted] |
| Dealers with membership | [redacted] | [redacted] | [redacted] |
| Dealers with membership and customer rating | [redacted] | [redacted] | [redacted] |
| Average rating score | [redacted] | | |
| Median rating score | [redacted] | | |
| Number of impressions overall | [redacted] | | |
| Number of visits overall | [redacted] | | |

Table 2: Test Sample One: Overview

For the second experiment on fraud, direct rating data from a different website feature is closely examined. The service at hand offers a platform where garages and workshops are able to present and promote themselves and potential customers can book specific services for their cars. The platform thereby accompanies the two parties as an intermediary. In a special field test users booked a service for their car at a workshop via the portal and were asked via an email request to give feedback one day after having the appointment at the workshop. Via a link users get to a rating form (A 2) which consists of five variables to be rated in a scale from one (very bad) to five (very good) stars. The variables incorporate criteria as kindness, competence, cleanness or price value and an overall rating. Further customers can decide, if they would recommend the workshop and finally a textbox allows giving 500 character free text feedback. The collection of user feedback started in April 2012 and still continues. For the experiment rating data from about one year is taken into consideration. In the underlying time [redacted] were collected. With 59% a little more than half of the ratings were submit with a free text comment. Only a marginal number of raters (5%) would not recommend a workshop as worthy to revisit. The average score for the overall rating aggregated [redacted] An overview on the test sample data is presented in Table 3.

The dataset is used to scan the content for potentially fraudulent and manipulative ratings. A manual and automated (in collaboration with the company) semantic analysis will be performed to derive implications especially from the free-text user feedback. The very high number of user recommendations (95%) and high average as well as more robust against statistical outlier median rating score [redacted] indicate a rather improbable influence of fraudulent denigrations. However, there may be potential attempts of workshops to rate themselves, thereby whitewash their reputation and dilute the impact of the platform's reputation system for customers. The purpose of the

experiment is to identify such and corresponding fraudulent behavior. There are standardized prevention tools integrated in the reputation system's architecture. But on top of that only those users may submit a rating which actually booked a service. Consequently fraud and manipulation is connected to a higher effort. Ratings of uninvolved third parties should thereby be excluded.

| **Test Sample Two: Overview** | |
|---|---|
| Data collection timeframe | |
| Number of workshops | |
| Total number of ratings | |
| Number of ratings with free-text feedback | |
| Number of ratings with recommendation | |
| Number of rating with un-recommendations | |
| Average number of ratings per workshop | |
| Average score for overall rating | |
| Median score for overall rating | |

Table 3: Test Sample Two: Overview

## 4.4 Data Analysis

Data for the first experiment is analyzed in a highly standardized manner using statistically approved measures to evaluate empirically the relationship among different variables and thereby receive results according to possible interferences. The aim is to confirm relationships based on results of statistical significance. The second experiment must be seen as a combination of predominantly quantitative examinations of rating data such as frequency analysis, review spam categorizations and detection of duplicate reviews with further qualitative semantic approaches by evaluating the rating content more in-depth. The focus thereby lies predominantly on free-text user feedback and reviews. In what follows the data analysis procedure for both datasets is described in more detail.

Hypothesis **1** and **2** are based on an experimental approach. Therefore a tool, considered as an important statistical aspect of online based applications, is used in order to measure user and customer behavior (Borodovsky & Rosset, 2011). This is done for instance in dependence of variances of design, usability or content on a webpage. The so called A/B-test or split test compares, as implied by the name, two or more versions of an instance which are equal besides one variation. In terms of web pages, a similar amount of online traffic is directed on each of the versions and thereby the impact of the variation on the conversion can be measured. As indicated earlier, the foundations for measurement are the CTRs and CCRs. For a better understanding the effect is exemplified in Figure 8. Normally a currently used version [A] with no changes is compared with a slightly modified version [B] in order to match the discrepancy and if version [B] has a significantly noticeable impact.

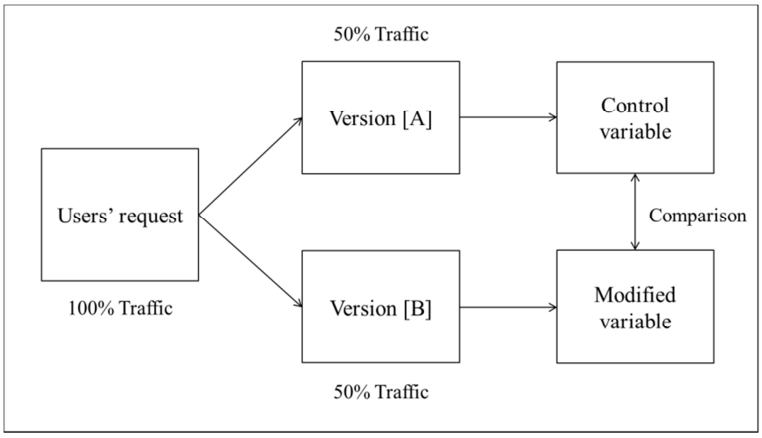



Figure 8: A/B Testing (Own illustration)

In case of this experiment the test setting includes an A/B/C test campaign. Thereby customer ratings are implemented on dealers' market place presence in different settings. In detail, this means that the online offerings of dealers will include a seal of the third party organization, the overall reputation score and a link leading to the detailed customer feedback information. The control variable is ensured by a Version [A] without any integration of the information just mentioned. Version [B] will include the rating information on the *list-view* where potential customers can choose from a list of dealer offerings dependent on their search criteria. The rating information in version [B] also will be displayed on the *details-view* when customers choose one specific offering from the list. And finally the information is shown on the *dealer-information-page* which can be clicked on from the details page to get further dealer information. Version [C] excludes in comparison to version [B] the integration on the list-view. Exemplary representations of how the particular units are integrated in detail, are provided in the appendix (A 3; A 4). Figure 9 gives an additional overview of the testing for the particular versions.

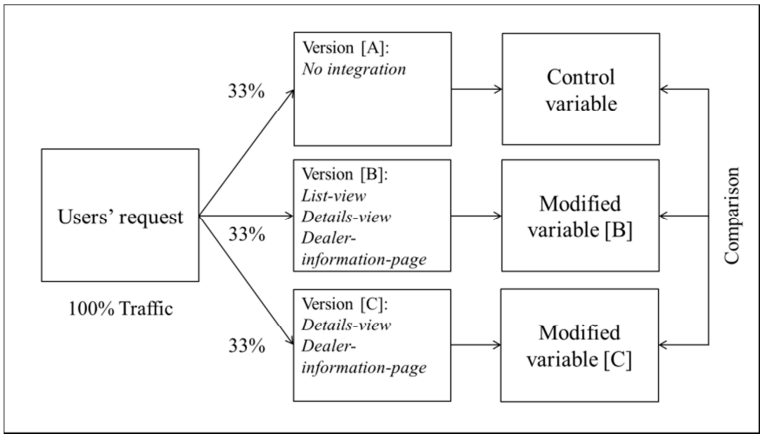



Figure 9: A/B/C-Test Setting (Own illustration)

After collecting the necessary data (CTRs and CCRs) the versions are compared for statistically significant variances of the modified variables in comparison to the control variable. Therefore, the so called *z-Test* is applied with the use of the *z-Score* which can be described as the number of standard deviations between the mean values of the control and modified variations. Using the z-Score is based

on the assumption that a binomial or Poisson distribution can be approximated with a normal distribution for sufficiently large values. So if the sample consists of more than 1000 views events than the following applies:

$$F_{Binomial}(k; n, p) \approx F_{Poisson}(k; \lambda = np)$$

$$F_{Poisson}(x; \lambda) \approx F_{Normal}(x; \mu = \lambda, \sigma^2 = \lambda)$$

In order to determine if the results are significant they are tested with 90% (Significance level α=0.1), 95% (α=0.05) and 99% (α=0.01) confidence intervals around the conversion rate and according ρ-values, respectively.

Hypotheses **3a/b** and **4** are evaluated by the given information on average clicks per day and average contacts per day for each offering provided by the gathered database of test sample one (Table 2). The dataset is arranged in dependence of reputation scores and counts and assessed for significant differences in conversion. In order to value the results for statistical significance, tests for correlation and a series of *t-Tests* are applied.

The data analysis process for hypothesis **5** will involve several steps. The goal is to identify fraud or manipulations within the mostly free-text based customer reviews and comments. In order to structure the search for possible review spam, the approaches by (Jindal & Liu, 2007, 2008) serve as guidance. First review spam has to be categorized in order to know what exactly to search for. Thereby according to Jindal and Liu (2007) three major types of review spam can be identified. Those are *false opinions* including underserving positive or negative opinions (Type 1) either to extensively promote an offer or to deliberately damage reputation. Furthermore, review spam can be categorized into *reviews on brands only* (Type 2) which contain solely information on a brand, manufacturer or transaction partner and not the underlying product or service. And finally the authors classify so called *non-reviews* as type 3 review-spam. Such reviews for instance feature no opinions and therefore do not fulfill their purpose. Type 3 review-spam may include advertisement or other non-reviews such as questions or answers, comments on other reviews or just completely random text that does not have anything to do with the actual item of interest.

The strategy for data analysis consequently follows the principle of getting a first impression of the sentiment for the whole test sample by basic statistics as distribution and frequency analysis. The purpose of the exploration is to develop a feeling for the spectrum of themes, the distribution of positive and negative content and identify key subjects. Type 3 and 2 review spam is approached by a manual semantic search enabled by supervised learning by means of the works of Jindal & Liu as well as related studies. Thereafter follows the assessment of Type 1 review spam. A possible method is to evaluate the data for duplicate and near duplicate reviews. For that the first step contains dismantling the review content from most of punctuation and special characters in order to enable better comparability among the reviews. In a next step the free-text reviews of the whole dataset are

disassembled in their individual n-grams, whereas n-grams are continuous sequences of words (or items) from an underlying text. An automated inter- and intra-dealer n-gram matching, further related to the *shingle* method by Broder (2000) serves the purpose to identify the (near-) duplicate reviews. Results are then manually checked for accuracy and validity.

A further method to identify potentially fraudulent reviews is completed by exploiting a suitable partition of the dataset for group deviations (Mukherjee, Liu, & Glance, 2012). Thereby deviations of a group consisting of potentially fraudulent reviews from genuine reviews are measured. High values may imply manipulations. The overall focus is to determine the likeliness of fraud and manipulation in temporary reputation systems and get a feeling for the proportion of review spam. Further the impact on the relation between a reputations system and the trust perception of online offerings needs to be estimated. To examine the results on significance, related studies on the subject serve as basis for comparisons, such as Jindal and Liu (2008) or Lau et al. (2011).

# Chapter 5: Results

## 5.1 Impact of Reputation Systems on Online Trust Perception

The first step in the data analysis process features the evaluation of the implemented A/B/C-Test experiment on the company's online marketplace website. The intention behind the experiment is to measure differences in conversion in dependence of a visible reputation score (**H1-H2**) in order to imply the effect on online trust perception. To measure the effects, online offerings of the [redacted] with an integrated reputation score were displayed in different ways on the marketplace. As mentioned in section *4.4*, the same dealer offerings are shown once with the reputation score on the list-view as well as on the details page, in a second version solely on the details page and in a third version without any reference to a reputation system as control variable (A 3; A 4). Customers' interest in an online offering and their willingness to engage in an online interaction are measured with CTRs and CCRs. The test was implemented for ten days [redacted] and revealed the following results.

**z-Test for Statistical Relevance of A/B/C-Test Results (CTRs)**

| Version | A: No integration | B: Full integration | C: Partial integration |
|---|---|---|---|
| Impressions | [redacted] | [redacted] | [redacted] |
| Detail hits | [redacted] | [redacted] | [redacted] |
| CTRs | [redacted] | [redacted] | [redacted] |
| Difference to A | | [redacted] | [redacted] |
| Standard error | [redacted] | [redacted] | [redacted] |
| z-score | | [redacted] | [redacted] |
| ρ-value | | [redacted] | [redacted] |
| Confidence level: | | [redacted] | |
| 99% (α=0,01) | | [redacted] | [redacted] |

Table 4: z-Test for Statistical Relevance of A/B/C-Test Results (CTRs)

**z-Test for Statistical Relevance of A/B/C-Test Results (CCRs)**

| Version | A: No integration | B: Full integration | C: Partial integration |
|---|---|---|---|
| Visits | [redacted] | [redacted] | [redacted] |
| Contacts | [redacted] | [redacted] | [redacted] |
| CCRs | [redacted] | [redacted] | [redacted] |
| Difference to A | | [redacted] | [redacted] |
| Standard error | [redacted] | [redacted] | [redacted] |
| z-score | | [redacted] | [redacted] |
| ρ-value | | [redacted] | [redacted] |
| Confidence level: | | [redacted] | |
| 99% (α=0,01) | | [redacted] | [redacted] |
| 95% (α=0,05) | | [redacted] | [redacted] |

Table 5: z-Test for Statistical Relevance of A/B/C-Test Results (CCRs)

From a first glance it becomes apparent that [redacted]

[redacted]

As indicated in *Chapter 4*, CTRs and CCRs are evaluated as reasonable proxy for online trust perception. The experiment now demonstrates that the implementation of a reputation score on dealer offerings in a B2C online marketplace fosters the conversion on the marketplace. In detail this means that both customers show a higher interest in online offerings (demonstrated by higher CTRs) and are more willing to engage in online interactions (demonstrated by higher CCRs). Accordingly, the reputation scores foster the customers trust perception of online offerings. Consequently, Hypotheses **1** and **2** appear to find support in the results of the A/B/C-Test experiment.

However, in order to approve the argumentation above, the results have to be checked for statistical relevance. Since the A/B-Test or split test is a common tool to test the effects of variations on webpages the most frequent analysis technique is by means of a z-Test. Thereby the distribution of the test statistic can be estimated with a normal distribution under the null hypothesis due to very large sample sizes. As a result, if the variance of the population is known and the sample size is large, a lot of statistical tests can be performed as z-Tests.

The name z-Test refers both to the normal distribution which is also known in mathematics as *Z-distribution* and the calculation of the so called z-Score. The score is interpreted as the number of standard deviations of an observation and enables to examine whether the particular score is above, below or equal to the mean of a number of scores. Furthermore the score exemplifies how far a particular score diverges from the mean. For a random variable X= p-$p_c$ the z-Score can be calculated as:

$$z = \frac{p - p_c}{\sqrt{\frac{p(1-p)}{N} + \frac{p_c(1-p_c)}{N_c}}}$$

Whereas N represents the sample size of an experimental variation or treatment and $N_c$ the sample size of the control variable. The results of the z-Test for statistical relevance of the differences in CTRs and CCRs for versions [B] and [C] are demonstrated in the tables before. Thereby, Table 4 goes into the measurable differences in CTRs and the second Table 5 addresses the changes for the CCRs in depth.

[redacted]

Those offers, incorporated with the visible reputation score of a dealer performed significantly better in terms of CTRs and CCRs than the control variable with any integration of a reputation score.

Furthermore the experiment showed that these effects don't just occur when offerings are presented in a list-view and customers directly have the possibility to choose among offers with and without a reputation score. Version [C] with a partial integration of the reputation score solely on the details page received as well significantly higher CTRs and CCRs even though a previous choice among offerings with and without scores was not given (as in version [B]). This circumstance in addition affirms Hypotheses **1** and **2** which assume a reputation system as variable affecting the online trust perception of online offerings. It can be argued that the impact of version [C] on online customers' interest in online offerings and their willingness to engage in online interaction is caused entirely by a fostered online trust perception (provided by the reputation system). There was no direct comparison between offerings with and without a reputation score (like on a list-view) incorporated for version [C]. Customers led to version [C] of the experiment were confronted with the reputation score first when they already chose the vehicle they were interested in. And still those offerings resulted in more conversions and by implication were perceived with higher online trust. As a consequence of the statistical confirmation of the assumption elaborated on behalf of Table 4 and Table 5, both Hypothesis **1** and **2** can be supported. The implementation of reputation scores does have a positive impact on conversion.

**H1: Supported**
**H2: Supported**

In order to evaluate if the height of a reputation score additionally affects the conversion on an online marketplace and therefore trust perception of online offerings, the gathered data was approached once more. The company's database provided for each of the dealer offerings (Table 2) *average clicks per offer per day* as well as the *average contacts per offer per day*. These rates emerged as promising measures to evaluate Hypothesis **3a** and **3b**. Every online offering provided by [redacted] is checked for the just mentioned measures for the given collection time frame of the test sample. Variances in the online trust perception of customers are hereby measured with the differences in mean/median values of clicks and contacts per offering in dependence of the reputation score height.

A first finding however disqualifies the average contacts per offering as valid measure with a high significance, since the frequency of offerings with measurable contacts is remarkably low. As indicated by [redacted] offerings with an implemented reputation score do provide information on registered contacts. In contrast [redacted] offerings show measureable average clicks per day and hence qualify them as an acceptable measure.

**Proportion of Clicks and Contacts per Offering**

| | Only membership | Ratio | Membership and rating | Ratio |
|---|---|---|---|---|
| Offers with measurable clicks | [redacted] | [redacted] | [redacted] | [redacted] |
| Offers with measureable contacts | [redacted] | [redacted] | [redacted] | [redacted] |
| Total | [redacted] | | [redacted] | |

Table 6: Proportion of Clicks and Contacts per Offering

Since the scores for the reputation system at hand are based on the Dutch grading system the range of possible scores reaches from 1 up to 10 which corresponds to a scale from very poor to excellent. Unfortunately, the company's policy towards the representation of reputation scores has to be seen as a minor obstacle. On the marketplace only reputation scores with a range from 8.0/10 to 10/10 are displayed in order to not discriminate dealers with a lower score. Vendors below a score of 8.0 have the same seal of the third party organization on their offerings but without the visible rating score incorporated (A 3). Consequently the first evaluations focus on the implications which can be derived from the scores starting at 8.0.

As already indicated the reputation system is based on the Dutch academic grading system used by the majority of institutions in the Netherlands. Thereby grades from 8.0 on are interpreted already as good or very good and consequently there might be a positive impact on the offerings and how they are perceived by customers. The question arises if any significant differences or patterns among the median clicks for the 8.0 to 10 range are recognizable?

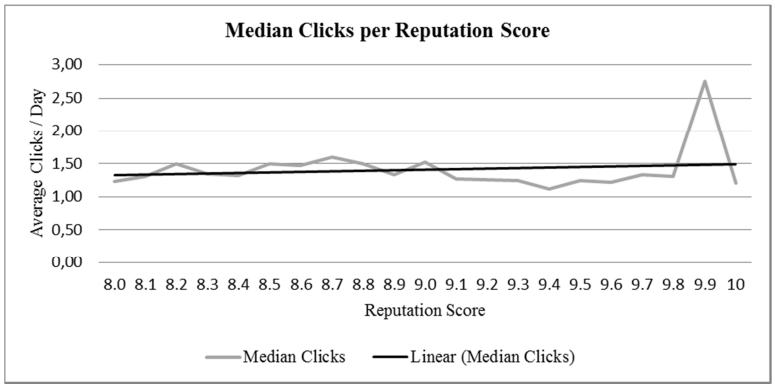



Figure 10: Median Clicks per Reputation Score

Figure 10 shows the median clicks of all offerings in dependence of a certain reputation score. What can be seen is that no obvious trends or patterns between click frequency and rating score are identifiable. In order to elucidate the low distinctions in median values in dependence of the reputation score, Figure 10 makes use of a representation of the results together with a linear trend curve. The rather low fluctuations around the linear trend curve once again illustrate the high homogeneity among the median clicks per offering in dependence of the reputation score. The only major outlier is to be found at a reputation score of 9.9. The reason for this inconsistency must be addressed to the unequal distribution of offerings per reputation score. Taking a closer look on the number of offerings per

score (Figure 11) it becomes obvious that there are very few dealers with a score of 9.9 and consequently a dilution of the median clicks is difficult to rule out. When comparing the actual values which are made available in the appendix (A 5) it becomes apparent that from [redacted]

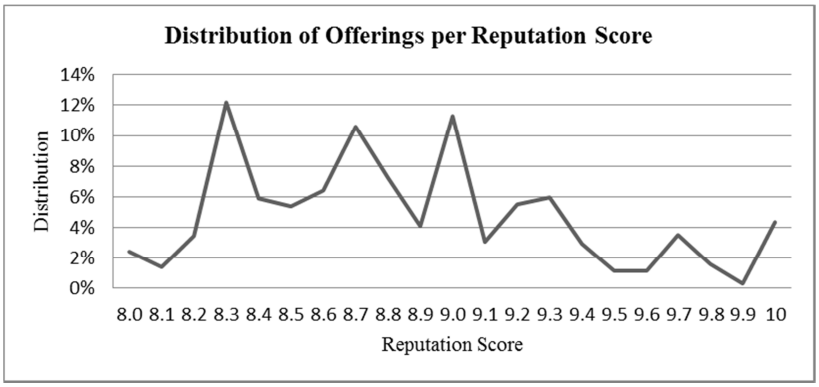



Figure 11: Distribution of Offerings per Reputation Score

To test the above mentioned perceptions on statistical significance, a first step is to check if a correlation among the median clicks and reputation scores height is measurable. Hence, *Pearson's correlation coefficient* can be applied with the values illustrated in A 5. The formula for Pearson's coefficient measures the linear correlation among variables with a value between +1 and -1 whereas values close to 1 indicate a very high or perfect correlation among the variables and corresponding values close to 0 a marginal or no correlation at all. The coefficient is defined as covariance among the variables, divided by the product of the variables' standard deviation. For two variables this means:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \text{ or } r = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \text{ for samples}$$

Applying the formula results in a correlation coefficient between the median clicks and the reputation scores of just 0.1515 and has to be interpreted as a very low correlation among the variables. Since the correlation coefficient reacts rather sensitive on outlier and small changes in the dataset, as the isolation of the median clicks for the 9.9 reputation score, a further test can be made use of.

By separating the total number of offerings into two groups (Group A: Scores 8.0-8.9; Group B: Scores 9.0-10) a two sided *t-Test* with unequal variances should identify if there are statistically significant differences in the median clicks in dependence of the reputation scores. The results can be found in (Table 7). The very high ρ-value of 0.44 indicates that no significant difference among the two groups in terms of their median clicks can be identified. Not even with a low confidence of 80 % (α=0.2) the two groups show any measurable differences from each other.

| t-Test on Click Homogeneity (Scores 8.0-10) | | |
|---|---|---|
| Group | | |
| Mean | | |
| Sample size | | |
| Standard deviation | | |
| ρ-value | | |
| Confidence level: | | |
| 99% (α=0,01) | | |

Table 7: t-Test on Click Homogeneity (Scores 8.0-10)

The results implicate that customers are rather indifferent in terms of their trust perception of online offerings, if an offering shows a reputation score of 8.0 or higher. But how does the group of offerings with a score of 8.0 or higher perform in comparison to the group of offerings with scores ranging from 1.0 to 7.9?

The already mentioned peculiarity of the marketplace does not allow a comparison of every single reputation score and its median clicks. In order to draw inferences from the reputation scores about the customers' trust perception despite this inconvenience, a different approach has to be established. A solution therefor can be found in the separation of the sample again into two groups. Herby it's possible to distinguish between a group C which incorporates all offerings without a visible reputation score and thus lies in the range of 1.0 to 7.9 and a group D which comprises all the remaining offerings with visible reputation scores from 8.0 to 10. The approach enables a comparison of the two groups with statistical measures. Calculating the mean and median clicks for each group, results in the following Figure 12. What can be observed is a distinct disparity among the two groups in terms of customer clicks per offering. It appears that offerings with a higher reputation score than 7.9 tend to perform better in clicks than ones with a lower score. With the aim of eliminating doubts in the performance discrepancy of the two groups, the difference of group *C* and *D* are tested on statistical significance. Applying a further *t-Test,* shows significantly more clicks for the offerings of group D compared to group C.

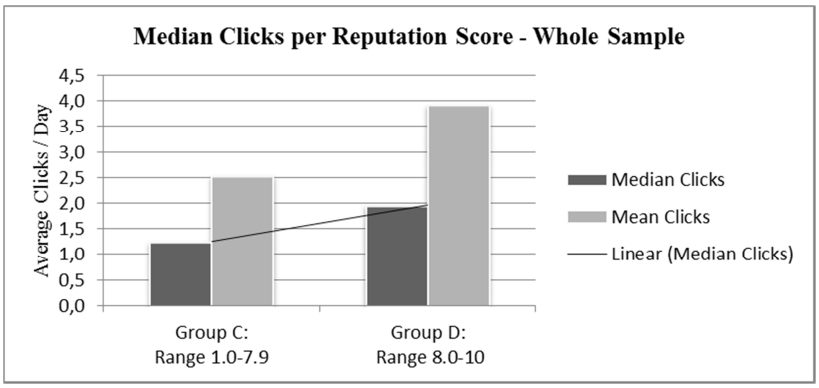



Figure 12: Median Clicks per Reputation Score – Whole Sample

Taking a closer look at the results provided by Table 8, the differences of the two groups are that unambiguous that even with very high standard deviations a difference is measurable with 99% confidence (α=0.01). What does that say about reputation scores and the trust perception of online offerings? Apparently customers do perceive first of all the existence of a higher reputation score and upon that tend to choose offerings with the higher reputation score measured by click conversion of group C and D (Table 8).

| **t-Test on Click Homogeneity (Scores 0,0-10)** | |
|---|---|
| Group | |
| Mean | |
| Sample size | |
| Standard deviation | |
| ρ-value | |
| Confidence level: | |
| 99% (α=0,01) | |

Table 8: t-Test on Click Homogeneity (Scores 1.0-10)

In order to shed more light on the different performance levels as a function of the reputation scores the sample is narrowed down to specific car models out of one car class. In this way, other possible variables affecting the outcome of why offerings with a higher reputation score (8.0-10) perform better than ones with lower scores (0.0-7.9) are dismissed. The four chosen models are the most frequent car types on the marketplace for each brand and act in the car market as direct competitors to each other. Furthermore, the models are matched in terms of production year and price to ensure a high as possible compliance among the offerings and at the same time don't minimize the count of offerings too much. A test for the click conversion in dependence of the reputation score group for each of the four models results in Figure 13 below. The findings implicate that for all four models the pattern in terms of the two reputation score groups is comparable to the results in the whole sample (Figure 12).

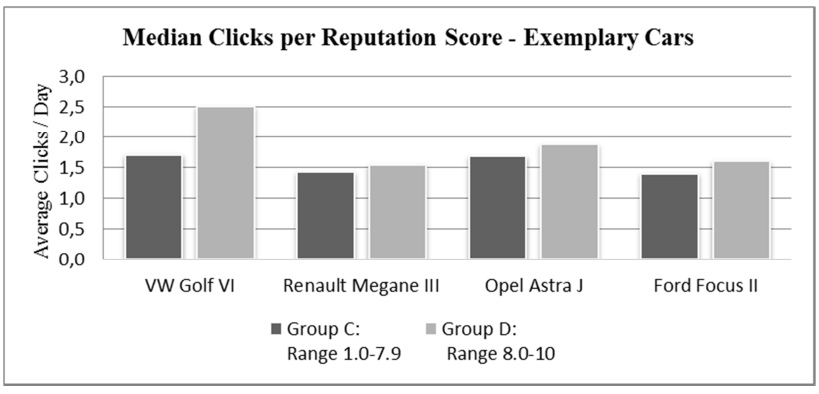



Figure 13: Median Clicks per Reputation Score – Exemplary Cars

Figure 13 demonstrates that all four models received higher clicks when offered by a dealer with a reputation score of 8.9 or higher. Especially for the Golf VI the differences are particularly noticeable with a disparity of 46.53%. In total offerings for the four models performed 20.19% better than those

with lower scores (Table 9). A test for significance however only provided significant results for the models VW Golf VI and Ford Focus II. Thereby the first model shows a difference in median clicks with 99% confidence (α=0.01) and according to that a very low ρ-value of 0.007 (Table 9). The Ford Focus II on the other side provides only 80% confidence (α=0.2) with a t-Test ρ-value of 0.1698. The other two models showed no statistical significance for the difference in median clicks. One possible explanation of such an outcome might be that VW dealers are generally perceived as more trustworthy than others. To interpret the results as indicatory for all four models might be too generic, since the missing significance in the exemplary car test can be addressed to the narrowed down sample from 35,528 offerings overall to only 2,096, but might have as well other reasons.

**t-Test on Click Homogeneity – Exemplary Cars (Scores 1.0-10)**

| Model | VW Golf VI | Renault Megane III | Opel Astra J | Ford Focus II | Total |
|---|---|---|---|---|---|
| Group C: 1.0-7.9 median clicks | | | | | |
| Group D: 8.0-10 median clicks | | | | | |
| Shift | | | | | |
| Sample Size | | | | | |
| t-Test ρ-value | | | | | |
| Confidence level: | | | | | |
| 99% (α=0,01) | | | | | |

Table 9: t-Test on Click Homogeneity – Exemplary Cars (Scores 1.0-10)

However, the evaluation of the dataset demonstrated that customers do perceive reputation systems in terms of the reputation score height. Furthermore, the applied t-Test for significance indicates that a higher score improves the click conversion and therefore the trust perception of online offerings, at least in terms of the whole sample. According to that, the group with lower reputation scores shows a significantly lower click conversion illustrated by the median clicks per offerings. As a result Hypothesis **3a** and **3b** are regarded as supported.

**H3a: Supported**

**H3b: Supported**

In general, both the implemented experiment as well as the following evaluation of the company's database identified a reputation system as an online trust fostering mechanism. In a next step the role of rating credibility as a moderator of the relationship between a reputation system and the trust perception of online offerings, is closely examined.

## 5.2 Rating Count Impact as a Credibility Indicator

As indicated in *Chapter 4*, the rating count a reputation score consists of serves as a credibility indicator for the relationship between reputation score and conversion. Hypothesis **4** is based on the assumption that the impact of a reputation score on trust perception (conversion) varies in dependence of the actual rating count per reputation score. Thereby the question arises if there are measurable differences in the median clicks between offerings with very few ratings/reviews and ones that are based on larger amounts. Test sample one (Table 2) thereby again serves as foundation for the examinations. Tests with the database demonstrated that the median clicks for offerings with a reputation score of 8.0 or higher are very homogeneous (Table 7) and consequently the effects of the rating count can be observed isolated from the reputation score.

In a first step the occurring rating counts for [redacted] offerings are identified. Overall [redacted] rating counts were identified reaching from reputation scores based on only 1 up to 999 counts. The frequency of offerings per each rating count class needs to be calculated in order to determine the median clicks for each class. In the following Figure 14 the median clicks per rating count are displayed.

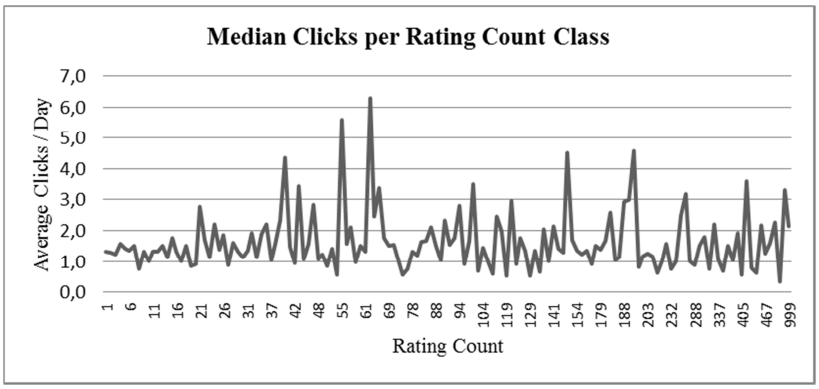



Figure 14: Median Clicks per Rating Count Class

As can be seen there are no obvious patterns identifiable in terms of median clicks in dependence of the rating count. A test for correlation between rating count class and median clicks therefore showed a very low Pearson's correlation coefficient of just 0.0337. One possible reason for this rather random click frequency could be addressed to the distribution of the number of offerings per rating count class. In the next two figures the actual number of offerings per class is visualized. The first graph (Figure 15) shows all rating counts and demonstrates how unevenly the amount of offerings is distributed upon those classes. It appears that the majority of the offerings' reputation scores are based solely on one up to nine ratings/reviews. Taking a closer look on the following distribution (beginning with a rating count of ten) of offerings per reputation count class the very high fluctuations in terms of number of offerings per classes become visible (Figure 16). While several classes provide a few hundred offerings there are others that have a few to almost none as basis for evaluations. Taking a closer look for instance on the rating count classes from [redacted]

[redacted] Consequently the validity of Figure 14 has to be assessed as rather low. An overview of all classes with the corresponding number of offerings and median clicks can be found in the appendix (A 6).

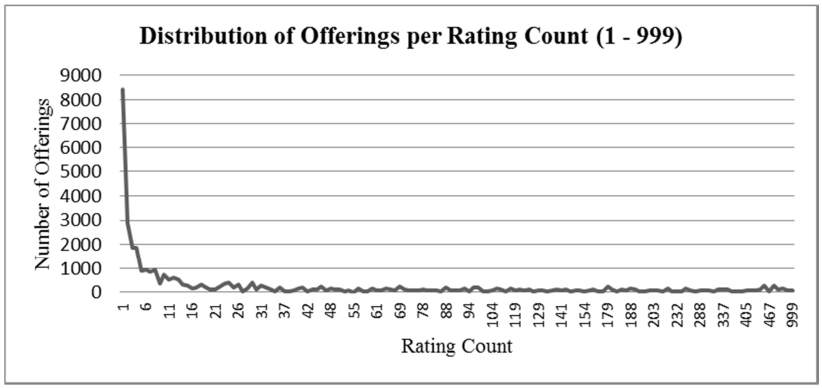



Figure 15: Distribution of Offering per Rating Count (1-999)

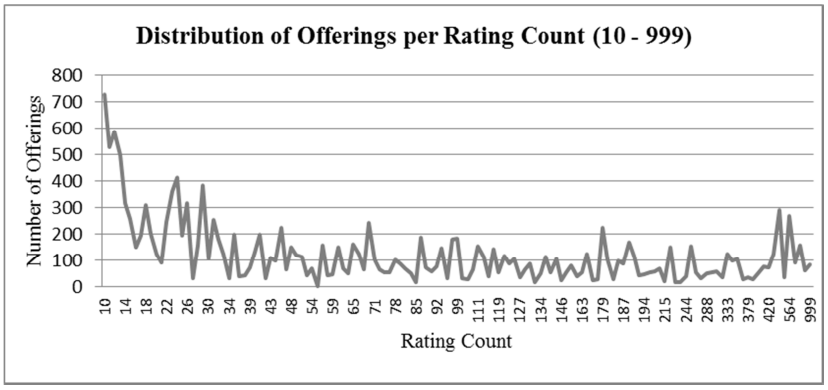



Figure 16: Distribution of Offerings per Rating Count (10-999)

In order to circumvent this peculiarity and guarantee certain validity for the impact of the rating count on customers' online trust perception, the dataset has to be approached in another way. There are known rating counts reaching from 1 up to 999. For the purpose of a valid interpretation of results a proper classification of the rating counts needs to be elaborated. The foundation of the classification is the distribution of offerings around specific rating count classes (Figure 15). Therefore offerings which present reputation scores based on only one rating/review serve thereby as starting point. [redacted] To enable comparability the remaining classes are adjusted on the one hand in terms of number of offerings and on the other in terms of reasonable intervals. The resulting classifications are shown in Table 10. Furthermore Figure 17 provides a visualization of the distribution of offerings per elaborated rating count classes.

**Distribution of Offerings per Rating Count Class**

| Rating count class | [redacted] | | | | | | |
|---|---|---|---|---|---|---|---|
| Median clicks | [redacted] | | | | | | |
| Number of offerings | [redacted] | | | | | | |
| Distribution | [redacted] | | | | | | |

Table 10: Distribution of Offerings per Rating Count Class

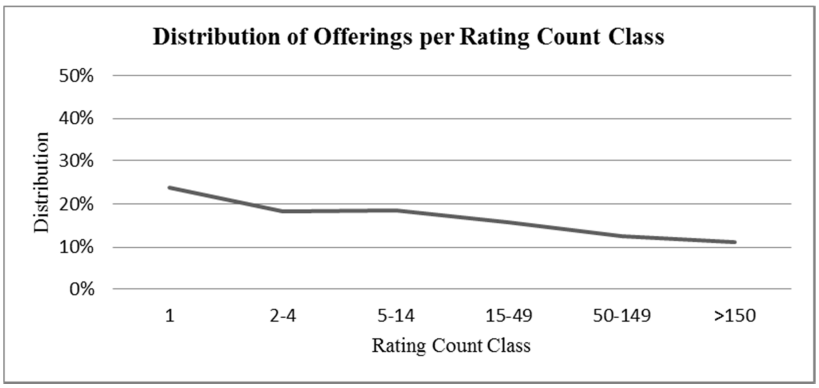



Figure 17: Distribution of Offerings per Rating Count Class

The elimination of the high fluctuations as well as the adjustments in terms of number of offerings per rating count class enables a realistic comparison of the median clicks per class. In this way inferences of the impact of rating count on the relation of reputation score and trust perception of online offerings can be drawn. The comparison of the median clicks for each of the rating count classes shows the following pattern (Figure 18).

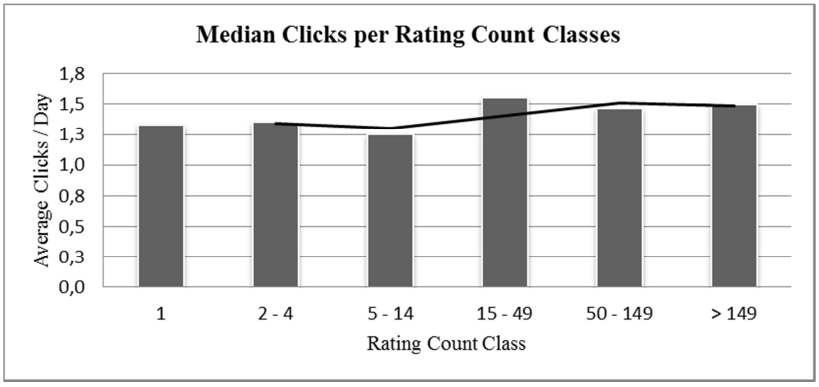



Figure 18: Median Clicks per Rating Count Classes

As can be seen the amount of median clicks per rating count class clearly rises till a certain level of just about 50 counts per reputation score. From there on it stays on a rather homogeneous level of around 1.5 median clicks (Table 10).

This pattern implies that customers do perceive a reputation score differently in dependence of the rating count. Consequently, the rating count influences the strength of the relation between reputation system and their trust perception of online offerings. As it shows, reputation scores that are based on more counts perform better in terms of median clicks until a certain level. From there on the median clicks stay on a consistent level. It can be argued that customers are rather indifferent in their trust

perception from the level of 15 counts and more. This means reputation scores based on 50 or 150 counts are perceived rather evenly. In order to evaluate these patterns for statistical significance again a t-Test is applied. Thereby two groups can be elaborated from Figure 18. The first one incorporates rating counts from one until 14 and the second one accordingly includes all the counts from there on, beginning with 15 counts. The t-Test should exemplify if the two groups differ distinctly from each other in order to address significance.

| **t-Test on Click Homogeneity (Rating Count Groups)** | | |
|---|---|---|
| Group | | |
| Mean | | |
| Sample size | | |
| Standard deviation | | |
| ρ-value | | |
| Confidence level: | | |
| 99% (α=0,01) | | |

Table 11: t-Test on Click Homogeneity (Rating Count Groups)

The results indicate a difference between the two groups with 99% (α=0.01) confidence according to the t-Test ρ-value of 0.008. So even though the differences may appear not that substantive, they still can be measured with statistical significance.

Now what does that say about Hypothesis **4**? The evaluation showed findings which do support the assumption that the relation between reputation score and conversion is positively affected by a higher amount of ratings/reviews a reputation score consists of. Consequently Hypothesis **4** as well finds support in the elaboration above.

**H4: Supported**

## 5.3 Impact of Fraud /Manipulation/Spam

In the following analysis the possible dilution of a reputation system's credibility and thereby its impact on the trust perception of online offerings is closely examined. Test sample two (Table 3) provides the basis for evaluations. The focus is directed on customer ratings/reviews with free-text feedback. As indicated by Table 3 the dataset delivers [redacted] of ratings of the whole dataset. Furthermore the gathered data provides among others additional information which is similarly valuable for evaluations, such as:

**Provided Information by Test Sample Two**

| Information on: | Explanation: |
|---|---|
| [redacted] | [redacted] |

Table 12: Provided Information by Test Sample Two

To get a first impression on customers' rating behavior the dataset is evaluated for the ratings/reviews distribution in terms of their underlying reputation score. Figure 19 indicates a clear distribution towards high and therefore positive rating scores. As can be seen, the majority of ratings/reviews lies within the range of 4.0 up to 5.0 star ratings. A high amount of purposely negative ratings/reviews therefore already can be ruled out.

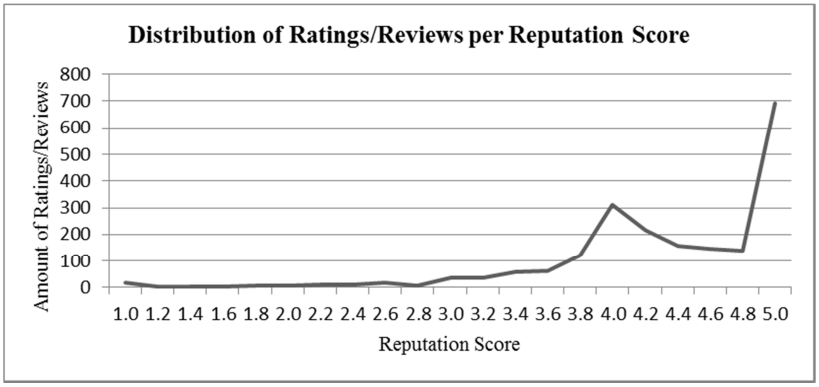



Figure 19: Distribution of Ratings/Reviews per Reputation Score

The intention of the analysis is to identify potentially fraudulent, manipulative or spam reviews to draw inferences on the robustness of the given reputation system. In more detail this means to assess if an impact on the relationship between a reputation system and the trust perception of online offerings might occur.

The task is thereby approached by a manual and automated semantic analysis of the free-text reviews with a following manual evaluation of the outcomes. Upon that, experiences of related studies serve as guidance in terms of classifications and comparisons of the results. As indicated in *section 4.4*, the thesis follows the review-spam classification of Jindal and Liu (2007) which identified three major types of review spam: Type 1- Fraudulent and manipulated content; Type 2- Reviews without information on the underlying service or product; Type 3- Non-reviews without any information fulfilling the purpose of a review.

Additional testing for possibly manipulated content is completed by exploiting a suitable partition of the dataset for group deviations. Following the study of Mukherjee et al. (2012) high deviations among a group of potentially fraudulent and genuine ratings/reviews may cause dilutions of a reputation score in order to change the sentiment on a product or service. On a 5-star rating scale the maximum possible deviation consequently is 4. The deviation of the groups for a workshop service can be calculated as follows:

$$D(g,s) = \frac{|AR_{s,g} - \overline{AR}_{s,g}|}{4}$$

The formula displays the deviation from a genuine group $g$ on a workshop service $s$, whereas $AR_{s,g}$ equals to the average reputation score for a workshop by the genuine group and $\overline{AR}_{s,g}$ accordingly to the average score of the potentially fraudulent group. |…| denotes the absolute value to circumvent negative (below zero) values. The larger the deviation the higher the damaging impact. As measure for separating the groups the rating information on "workshop recommendable: YES or NO" is used. Workshops with at least 15 ratings/reviews were taken into consideration.

**Test for Deviation D(g,s) - Results**

| workshop | # Ratings | $AR_{s,g}$ | $\overline{AR}_{s,g}$ | $D(g,s)$ | workshop | # Ratings | $AR_{s,g}$ | $\overline{AR}_{s,g}$ | $D(g,s)$ |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 48 | 4.54 | 2.00 | 0.63 | **7** | 20 | 4.42 | 3.00 | 0.35 |
| **2** | 36 | 4.44 | 1.00 | 0.86 | **8** | 19 | 4.38 | 3.00 | 0.34 |
| **3** | 24 | 4.34 | 3.00 | 0.33 | **9** | 19 | 4.36 | - | - |
| **4** | 23 | 4.50 | 2.00 | 0.62 | **10** | 17 | 4.47 | - | - |
| **5** | 20 | 4.45 | - | - | **11** | 16 | 4.31 | - | - |
| **6** | 20 | 4.55 | - | - | **12** | 15 | 4.53 | - | - |

Table 13: Test for Deviation D(g,s) - Results

Only 12 workshops fulfilled the requirements of at least 15 ratings/reviews. Table 13 furthermore indicates that for only half of the analyzed workshops a $\overline{AR}_{s,g}$-group was identifiable. Three workshops show a deviation of more than 0.50. A manual review analysis of the concerned workshops showed in terms of workshop [1] that the un-recommendations appear to be justified and well explained in the reviews. In case of workshop [2] the highest D(g,s) can be recorded. And here it is where the reviews show potentially fraudulent behavior. In both cases for workshop [2] the overall

rating and topic of complaint is nearly the same “booked service was canceled”. The negative review for workshop [4] showed no intention of fraudulent behavior.

Since the evaluation for manipulations by group deviations showed no comprehensive results the above mentioned classification method by Jindal and Liu (2007) is applied. Supervised learning by related studies enables to evaluate the dataset for type 2 and 3 review spam. Thereby [redacted] (Table 3) free-text reviews are labeled manually with regard to the just mentioned categories. Since type 2 and 3 has to be classified more as spam than willingly manipulative or fraudulent behavior, the ratings are evaluated as a whole instead of splitting up the dataset dependent on dealer id. Both types (2 and 3) are consolidated as non-reviews (Lau et al., 2011). The analysis process is further described in the appendix (A 7).

**Evaluation for Type 2 and 3 Review-spam - Results**

| Spam Type | # identifications | Proportion | Real world distribution |
|---|---|---|---|
| Type 2 | [redacted] | [redacted] | Jindal & Liu (2008) Lau et al. (2011) |
| Type 3 | [redacted] | [redacted] | |
| Total (non-reviews) | [redacted] | [redacted] | **2.0%** |

Table 14: Evaluation for Type 2 and 3 Review-spam - Results

Scanning the dataset manually for non-reviews (Type 2 and 3) revealed a total amount [redacted] free-text reviews that fit into the classification. A comparison to the total number of reviews results [redacted] of the dataset. According to related studies a regular distribution between genuine and non-reviews lies around 2.0% (Table 14). Now how is that huge gap interpretable? Since the classification is more suitable for product reviews, especially labeling Type 2 reviews was connected to some inconsistencies. For instance, a review text as *Workshop XY is great, easily reachable and I really liked the owner* has to be classified as Type 2 spam because no information on the underlying service was given. The same text with the integration of the merest information on the quality of the service however passed the filter: *Workshop XY is great, **very competent work**, easily reachable and I really liked the owner.*

Labeling Type 3 review spam showed no further ambiguities and includes reviews without any context to the workshop or underlying service which consequently do not fulfill the basic requirements of a review. Incomplete, grammatically unreadable reviews and questions also were addressed to this group. Examples for type 3 spam are for instance “I could not find the workshop” or random letter strings as “ssfhhhgfff”. Furthermore reviews with an obvious discrepancy among numerical reputation score and free-text review were classified as Type 3 review spam. Very short reviews as “great job”, “everything ok” or “Bad service” were not identified as Type 3 since they provide a measurable sentiment towards the workshop and service.

The results so far revealed a relatively high potential for spam reviews in the sample. Although they may not fulfill the primary requirements of a review, they are not potentially fraudulent, manipulative or intentionally misleading. The analysis of the much more harmful group of untruthful type 1 review spam is approached by an automated semantic analysis further related to the *shingle* method (Broder, 2000). Thereby the search focus lies upon duplicate and near duplicate content. In a first attempt free-text reviews of the whole dataset are disassembled in their individual n-grams, whereas n-grams are continuous sequences of words (or items) from an underlying text. Test sample two (Table 3) provides 1,219 free-text reviews. An evaluation of the n-gram frequency showed a range lasting from 1-grams up to reviews consisting of 84 terms. The average length of the reviews amounts to 11.84 terms and the median to 4. Using this knowledge, n-gram classifications for the test are 4-, 5-, 6-, 10-, 12-, 16-grams. The diagram below exemplifies the distribution of reviews in dependence of their n-gram length. As can be seen the majority of reviews is based on rather short n-gram sequences.

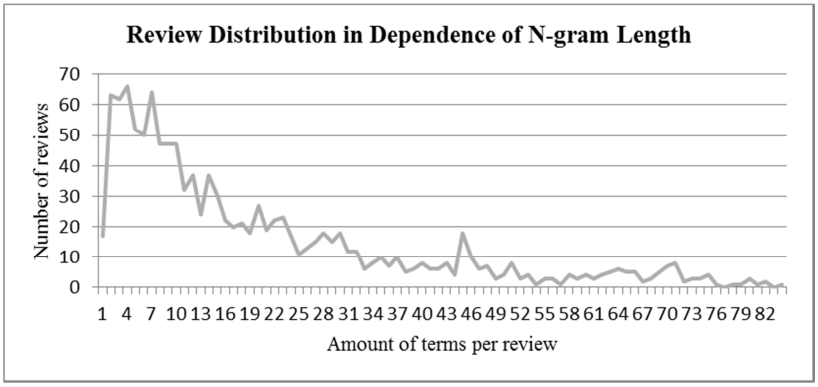



Figure 20: Review Distribution in Dependence of N-gram Length

In a next step every single 4-, 5-, 6-, 10-, 12-, and 16-gram is matched with the dataset of free-text reviews for possible duplicates or near-duplicates. All measures were applied with *Perl* programming language. Insights into the used scripts are given in the appendix (A 8; A 9; A 10). For all 6-gram classes the method identified duplicate matches. Table 15 gives an outline on the hits. The method turned out to be very effective as long as the length of the n-grams is reasonably high. From 6 items on the dataset revealed a number of (near-) duplicates which might incorporate fraudulent content.

**N-gram Matching on Test Sample Two - Results**

| #-gram | 4-gram | 5-gram | 6-gram | 10-gram | 12-gram | 16-gram |
|---|---|---|---|---|---|---|
| Identified (near-) duplicates | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] |

Table 15: N-gram Evaluation on Test Sample Two - Results

However, the lower the n-gram the more matches are found but without particular validity. As for the 4-grams, matching identifies many duplicates in text fragments which are based on common phrases like “I can recommend the” or “I was very satisfied”. Furthermore many inter-dealer matches were found. Yet, a manual evaluation showed no manipulative correlation what so ever within the inter-dealer matches.

As a consequence of the inconclusive results so far the analysis is expanded to an intra-dealer n-gram matching. Thereby the method is applied for every single dealer which showed conspicuities in the first matching approach and then checked for 1- up to 8-grams. The results indicate that n-gram matching in dependence of single dealers appears to be more suitable at least for the given test sample.

**N-gram Matching on Suspicious Dealer Ratings – Results**

| Workshop | # Reviews | # (near-) duplicates | % of total | Workshop | # Reviews | # (near-) duplicates | % of total |
|---|---|---|---|---|---|---|---|
| [redacted] | | | | | | | |

Table 16: N-gram Matching on Suspicious Dealer Ratings - Results

The second evaluation using the n-gram matching method revealed several interesting findings. Among the 20 dealers identified in the first matching attempt, all of them show (near-) duplicates in their customer reviews. By use of the Jaccard Coefficient similarity scores among the intra dealer (near-) duplicates are determined. The coefficient is calculated as follows:

$$J(A,B) = \frac{|S(A) \cap S(B)|}{|S(A) \cup S(B)|}$$

Whereas |S(A) ∩ S(B)| represents the total number of similar (n-gram) attributes of two reviews and |S(A) ∪ S(B)| the union of all (n-gram) attributes of both reviews. A perfect similarity consequently reflects in a score of 1 and no similarity in a score of 0 respectively.

[redacted] Most of the matches were identified as potentially fraudulent content. In those cases of lower similarity scores still longer fragments (up to 8-grams) showed a perfect matching what definitely can rule out a random coincidence. Interestingly, many dealers that showed (near-) duplicate reviews had very few reviews, in some cases even just the two duplicates what goes in-line with the observations of earlier works (Jindal & Liu, 2008). If the reviews were submit on purpose or occurred as a consequence of technical inconsistencies of the implemented reputation system cannot be assessed with absolute certainty. The additional information provided by the dataset (Table 12) revealed that some duplicates were submitted on the same date, for the same car and for the same service what rather implies inconsistencies of the reputation system and a double rating of one person. Others again differ in the just mentioned additional information but still provide the same or near the same reviews.

When it comes to the content of the duplicate reviews, all except one instance show a positive or very positive sentiment towards the workshop and underlying service. The near-duplicate found with a negative sentiment responds to a not accomplished service work for a customer's car. As inferred from the two reviews the workshop was not able to change the cambelt of a car. Even though the rating appears legitimate, the fact that there is a near duplicate with almost the same content submitted on the same day appears dubious. The similarity score of the mentioned reviews ads up to 0.28.

In general, manipulations and intentionally fraudulent behavior cannot be verified perfectly with the chosen inquiry. However, the found (near-) duplicates demonstrate a very high potential of doubtful assertions. In total 20 dealers from the dataset showed potentially fraudulent (near-) duplicate reviews which represent [redacted] of the whole test sample's reviews. Together with the before examined Type 2 and 3 non-reviews the number of concerned [redacted] of the test sample. The regular distribution of Type 1 to 3 review spam or untruthful and non-reviews is approximated with 6% (Jindal & Liu, 2008; Lau et al., 2011). Consequently, the examined dataset showed an above average frequency of doubtful reviews or ones which don't fulfill the purpose of a review. Now what does that say about hypothesis **5**? A statistically significant impact of the rating credibility upon the relation between a reputation system and the trust perception of online offerings cannot be calculated with the given preconditions. However, what is out of question is the fact that the identified reviews both by manual and automated semantic analysis approaches, have no valuable message for customers. Since customers do perceive the existence of reputation systems and thereby get affected in their trust perception of online offerings (Hypotheses **1** and **2**), it is reasonable to argue that a dilution of the reviews by untruthful and non-reviews may have a negative effect on the relation between a reputation system and the trust perception of online offerings. Even though the intention of manipulation or fraudulent behavior on the examined online marketplace may differ from more typical marketplaces with consumer goods such as *amazon.com* or alike, still a potential for fraud cannot be excluded[redacted] of the test sample reviews showed manipulative or spammed content.

Concluding, the findings on reduced rating credibility by the likeliness of fraud showed a potential for untruthful and non-reviews and therefore at least possible dilutions of the rating credibility. In a few cases willingly manipulative behavior cannot be ruled out. Undoubtedly, untruthful as well as non-reviews may have an impact on customers' perception of the reputation score and consequently dilute the trust perception of online offerings. Therefore hypotheses **5** finds as well support.

**H5: Supported**

# Chapter 6: Discussion and Conclusion

The intention of the Master thesis project was to clarify several questions concerning the contemporary role of trust in an online environment, the role of reputation systems as means to foster online trust perception and the robustness of such system against fraudulent and manipulative behavior. The need for research in the mentioned field of interest arose from the fact of trust as a major success factor, especially in the constantly growing field of online based activities. At the same time, the variable of trust is difficult to define both in the digital and in the real-world environment in terms of interpersonal relations, business cooperation or alike. The discussion of the results follows the analogical order of the research questions presented in C*hapter 1*. Thereafter, the limitations of the research approach, recommendations for practice and finally suggestions for future research will be presented.

## 6.1 Determinants of Online Trust

An extensive literature review enabled to answer the first two research questions: (a) *What are the main determinants of online trust?* and (b) *How are reputation systems and fraud assessed in terms of online trust?* The analysis of literature revealed trust in online settings as the mediating variable between the online environment and any online activity. The variable itself is determined by confidence, competence and benevolence criteria which as a precondition derive from offline settings. Furthermore, other variables directly affect online trust. The trust model presented in C*hapter 2* clarifies that first of all features of the online environment shape the perception of online trust. This leads directly to the second research question. Reputation systems embedded as trust mechanisms in the online environment were recognized by literature as means for trust communication and perception. In fact, their impact was identified as most influential among established online trust mechanisms. However, the danger of fraud and manipulations must not be neglected. Even very recent studies argue that reputations systems are among the most influential but at the same time most liable trust mechanisms. Without proper prevention mechanisms these systems might easily become victims of fraud.

In addition, literature revealed that gathering online trust is not a onetime procedure. The presented model implicates a process with a recurring cycle of trust, action and learning. Thereby the variable *learning* directly determines the formation of online trust by user experience, familiarity and satisfaction, gained with former online activities.

## 6.2 Impact of Reputation Systems on Online Trust Perception

The main research focus of the thesis lay on the question: (c) *Does a reputation system measurably affect the perception of online trust?* As elucidated by literature, reputation systems are trust mechanisms aiming at the active promotion of online trust. The intention was to scrutinize this

assumption with an experiment. On trial implemented reputation scores on a well-known online marketplace served as testing field. Conversion was chosen as a proxy for online trust perception. The results of the A/B/C-Test experiment showed statistically significant results (95% and 99% confidence) for click as well as contact conversion (CTRs and CCRs). Thereby the same offerings were presented to the online audience with and without implemented reputation scores. Referring back to (c), the implementation of a reputation score lead in all test settings to higher conversion and gives support for a positive impact on online trust perception. Further evaluations of the gathered data identified the height of a numerical reputation score as an additional factor affecting conversion and therefore trust perception. In the test environment higher scores resulted in significantly higher conversion than lower scores. Significant differences in the click performance (99% confidence) were found between the rating score group of 1 to 7 points and the one of 8 to 10 points. Interestingly the results further revealed that from a score of 8 till 10 customers show a rather indifferent behavior.

Additional testing focused on the number of ratings a reputation score is based on. The rating count advanced as a credibility indicator for the relationship among reputation score and conversion. Because of an uneven distribution, six rating count classes with an approximate number of offerings were arranged. The evaluated test sample recorded rising conversion until [redacted]. A test for significance between the click performance of the [redacted] a difference with 99% confidence. The findings therefore identify the rating count as a valid credibility indicator until a certain level. From there on, added counts showed no more significant impact.

### 6.3 Likeliness of Manipulations and Review-spam

A further goal was to examine the effectiveness of reputation systems in terms of robustness against attacks. As the research question (d) *How likely is fraud and manipulation of a current reputation system?* implicates the possibility of fraud, manipulations and spam in free-text comments was observed. The given dataset of free-text reviews was approached with manual and automated sematic analysis methods, following to some extend related studies to achieve comparability of results. The manual analysis identified overall [redacted] reviews as so called non-reviews. Compared with a real-world distribution of approximately 2% non-reviews, the amount was rather high. To identify the more harmful group of untruthful reviews an automated n-gram matching approach was applied. A measure for potentially manipulated content was determined by duplicate and near duplicate reviews as done by related studies. Matching occurred on inter- and intra-dealer level. While the inter-dealer assessment showed no manipulative behavior, intra-dealer matching provided [redacted] with similarity scores reaching from [redacted] Overall the test sample revealed a ratio of untruthful and non-[redacted] what is higher than the approximated real-word distribution. Regarding the research question, the likeliness of fraud and manipulations still exists in contemporary reputation systems.

This leads directly to the last research question (d) *Is there a need for a higher layer of trust which evaluates the rating's trustworthiness?*. A manipulative intention cannot be verified perfectly by the evaluation methods at hand. However, the found matches present inconsistencies and enough reason to doubt about their accurateness. Same goes for the manually labeled non-reviews. Such reviews dilute to some extend the validity of a reputation system. Even though they may inhabit a positive or negative sentiment, they provide no information or connection to the underlying product or service. Reputation systems are based in the broadest sense on the principle of word of mouth and trust transitivity. The correctness of the rating and review information has to be assessed as decisive for the success of these trust mechanisms. The need for a higher layer of trust cannot be entirely rejected. Advanced automated semantic analysis and opinion mining tools may support the established fraud filters in the search for misleading content and help to make the systems more robust against attacks.

## 6.4 Limitations of the Research

As in any qualitative or quantitative research approach, the Master thesis project as well is not without its limitations. Often in research a given sample is assessed for major generalizability in order to achieve applicability of the results for similar or related topics. In case of the underlying study the test sample for evaluating the impact of reputation systems on trust perception of online offerings as well as the likeliness for fraud and manipulations, was considered as reasonably extensive. Yet, for some in depth analyses the size of the dataset had to be minimized and assertions of statistical significance were not always possible. Still, the results showed implications which mostly correlate with findings from related studies.

In terms of the review evaluations for fraud and manipulations, perfect evidence for willingly manipulative behavior cannot be achieved with the utilized methods. However, the methods showed high potential for the measurement of untruthful and non-reviews and may find their way through further research into practice.

## 6.5 Recommendations for Practice

The research showed that trust is undoubtedly the success factor in online environments and enables sustainable competitive advantages. Thereby, online trust is a foundation for users and customers to approach in online activities. Online entities have to bear this in mind and find ways to generate, foster and sustain trust with and within their users and customers. For the cooperating company of this study, reputation systems were identified as a promising measure to achieve this goal. The company at hand should focus on the development of an own reputation system and implement it on all the company's main website features where the web presence acts as an intermediary or direct service provider. An own approach gives the ability to control the robustness of such a system. The findings showed that trust mechanisms like reputation systems are after all not without failure. Filtering for fraudulent and manipulative content has to be challenged, but at the same time the content should as well be

approached for misleading non-reviews. The used automated n-gram matching method exemplified a possible way for adjusting a reputation system's fraud filter. However, the company should exploit this prototype to develop advanced analysis tools which for instance incorporate semantic lexicons to detect doubtful content.

In general the influence of such systems on trust perception seems to have arrived in practical use. Entities deploying such mechanisms must therefore focus on the robustness and thereby not only on fraud prevention but as well on semantic content analyses in terms of fraudulent and non-reviews.

## 6.6 Suggestions for Future Research

The findings of the thesis offer insights into the relationship of reputation systems, online trust perception and the credibility of reputation systems. Additional research should strengthen the ties in these relationships by elaborating direct measures for online trust perception. In this study conversion was used as a proxy for trust perception. A further possibility is to enhance or refute the cross-sectional study findings with longitudinal approaches. Furthermore, the research shows that former findings on reputation systems are often based upon the effects on online marketplaces for consumer goods. The applicability for intermediary and service providing web sites are not without problems. Upon that, the scope on B2C platforms may be obsolete. Advancing fields as C2C social commerce are perfect examples for this shift. Research should examine the established relations in these new environments. Web based experiments appear very promising to get insights into these new domains.

The issue of untruthful and non-review feedback is closely connected to these new fields. The elaborated filtering mechanisms showed potential, but need to be classified as groundwork. Research should build upon the underlying findings and enhance semantic analysis methods. Advanced algorithms and the application of semantic lexicons could serve as an approach for future research. The filtering for (near-) duplicate reviews seems to be not enough to successfully prevent the dilution of reputation scores. Thereby, social commerce and C2C platforms appear as a reasonable testing field for the deployment of such new fraud prevention and content filtering measures.

# Bibliography


Aiken, K. D., & Boush, D. M. (2006). Trustmarks, objective-source ratings, and implied investments in advertising: Investigating online trust and the context-specific nature of Internet signals. *Journal of the Academy of Marketing Science, 34*(3), 308-323. doi: 10.1177/0092070304271004

Andersen, P. H. (2001). Relationship development and marketing communication: Towards an integrative model. *Journal of Business & Industrial Marketing, 16*(3), 167-183.

Andersen, P. H., & Kumar, R. (2006). Emotions, trust and relationship development in business relationships: A conceptual model for buyer-seller dyads. *Industrial Marketing Management, 35*(4), 522-535. doi: 10.1016/j.indmarman.2004.10.010

Anderson, E., & Weitz, B. (1989). DETERMINANTS OF CONTINUITY IN CONVENTIONAL INDUSTRIAL CHANNEL DYADS. *Marketing Science, 8*(4), 310-323. doi: 10.1287/mksc.8.4.310

Anderson, J. C., & Narus, J. A. (1990). A MODEL OF DISTRIBUTOR FIRM AND MANUFACTURER FIRM WORKING PARTNERSHIPS. *Journal of Marketing, 54*(1), 42-58. doi: 10.2307/1252172

Angriawan, A., & Thakur, R. (2008). A parsimonious model of the antecedents and consequence of online trust: An uncertainty perspective. *Journal of Internet Commerce, 7*(1), 74-94.

Arcand, M., Nantel, J., Arles-Dufour, M., & Vincent, A. (2007). The impact of reading a web site's privacy statement on perceived control over privacy and perceived trust. *Online Information Review, 31*(5), 661-681. doi: 10.1108/14684520710832342

Arino, A., de la Torre, J., & Ring, P. S. (2001). Relational quality: Managing trust in corporate alliances. *California Management Review, 44*(1), 109-131.

Atif, Y. (2002). Building trust in e-commerce. *IEEE Internet Computing, 6*(1), 18-24. doi: 10.1109/4236.978365

Ba, S. L., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quarterly, 26*(3), 243-268. doi: 10.2307/4132332

Babbie, E. R. (2012). *The practice of social research*: Wadsworth Publishing Company.

Baltas, G. (2003). Determinants of internet advertising effectiveness: An empirical study. *International Journal of Market Research, 45*(4), 505-513+520.

Barber, B. (1983). *The logic and limits of trust*: Rutgers University Press New Brunswick, NJ.

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.

Belanger, F., Hiller, J. S., & Smith, W. J. (2002). Trustworthiness in electronic commerce: the role of privacy, security, and site attributes. *Journal of Strategic Information Systems, 11*(3-4), 245-270. doi: 10.1016/s0963-8687(02)00018-5

Beldad, A., de Jong, M., & Steehouder, M. (2010). How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust. *Computers in Human Behavior, 26*(5), 857-869. doi: 10.1016/j.chb.2010.03.013

Benedicktus, R. L. (2011). The effects of 3rd party consensus information on service expectations and online trust. *Journal of Business Research, 64*(8), 846-853.

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), 310-323.

Berscheid, E., & Walster, E. (1974). PHYSICAL ATTRACTIVENESS. *Advances in experimental social psychology, 7*(1), 157-215.

Blois, K. J. (1999). Trust in business to business relationships: An evaluation of its status. *Journal of Management Studies, 36*(2), 197-215. doi: 10.1111/1467-6486.00133

Blumberg, B., Cooper, D. R., & Schindler, P. S. (2008). *Business Research Methods* (Vol. 2): McGraw-Hill Higher Education New York.

Borodovsky, S., & Rosset, S. (2011). *A/B Testing at SweetIM: The Importance of Proper Statistical Analysis.* Paper presented at the Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on.

Boyd, J. (2003). The rhetorical construction of trust online. *Communication Theory, 13*(4), 392-410. doi: 10.1111/j.1468-2885.2003.tb00298.x

Broder, A. Z. (2000). Identifying and filtering near-duplicate documents. In R. Giancarlo & D. Sankoff (Eds.), *Combinatorial Pattern Matching* (Vol. 1848, pp. 1-10).

Chang, H. H., & Wong, K. H. (2010). Adoption of e-procurement and participation of e-marketplace on firm performance: Trust as a moderator. *Information and Management, 47*(5-6), 262-270.

Chen, M., & Singh, J. P. (2001). *Computing and using reputations for internet ratings.* Paper presented at the Proceedings of the 3rd ACM Conference on Electronic Commerce.

Chen, Y., & Xie, J. (2005). Third-party product review and firm marketing strategy. *Marketing Science, 24*(2), 218-240.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research, 43*(3), 345-354.

Child, J. (2001). Trust - The fundamental bond in global collaboration. *Organizational Dynamics, 29*(4), 274-288. doi: 10.1016/s0090-2616(01)00033-x

Clemons, E. K., Gao, G. G., & Hitt, L. M. (2006). When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems, 23*(2), 149-171. doi: 10.2753/mis0742-1222230207

Conroy, P., & Bearse, S. (2006). Customer conversion. *The Changing*.

Corbitt, B. J., Thanasankit, T., & Yi, H. (2003). Trust and e-commerce: a study of consumer perceptions. *Electronic commerce research and applications, 2*(3), 203-215.

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: concepts, evolving themes, a model. *International Journal of Human-Computer Studies, 58*(6), 737-758.

Cova, B., & Salle, R. (2000). Rituals in managing extrabusiness relationships in international project marketing: A conceptual framework. *International Business Review, 9*(6), 669-685.

Das, T. K., & Teng, B. S. (1998). Between trust and control: Developing confidence in partner cooperation in alliances. *Academy of Management Review, 23*(3), 491-512.

Davis, F. D. (1989). PERCEIVED USEFULNESS, PERCEIVED EASE OF USE, AND USER ACCEPTANCE OF INFORMATION TECHNOLOGY. *MIS Quarterly, 13*(3), 319-340. doi: 10.2307/249008

De Maeyer, P., & Estelami, H. (2011). Consumer perceptions of third party product quality ratings. *Journal of Business Research, 64*(10), 1067-1073.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management science, 49*(10), 1407-1424.

Dingledine, R., Freedman, M. J., & Molnar, D. (2000). Accountability measures for peer-to-peer systems *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*: O'Reilly Publishers.

Dini, F., & Spagnolo, G. (2009). Buying reputation on eBay: Do recent changes help? *International Journal of Electronic Business, 7*(6), 581-598.

Doney, P. M., & Cannon, J. P. (1997). An examination of the nature of trust in buyer-seller relationships. *Journal of Marketing, 61*(2), 35-51. doi: 10.2307/1251829

Duan, H., & Liu, F. (2012). *Building robust reputation systems in the E-commerce environment.* Paper presented at the 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review, 14*(4), 532-550.

Fielding, N. G. (2012). Triangulation and Mixed Methods Designs Data Integration With New Research Technologies. *Journal of Mixed Methods Research, 6*(2), 124-136.

Flavian, C., Guinaliu, M., & Gurrea, R. (2006). The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & Management, 43*(1), 1-14. doi: 10.1016/j.im.2005.01.002

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research, 19*(3), 291-313. doi: 10.1287/isre.1080.0193

Garg, S. K., Versteeg, S., & Buyya, R. (2013). A framework for ranking of cloud computing services. *Future Generation Computer Systems-the International Journal of Grid Computing and Escience, 29*(4), 1012-1023. doi: 10.1016/j.future.2012.06.006

Gefen, D. (2000). E-commerce: the role of familiarity and trust. *Omega, 28*(6), 725-737. doi: 10.1016/s0305-0483(00)00021-9

Gefen, D. (2002). Reflections on the dimensions of trust and trustworthiness among online consumers. *Data Base for Advances in Information Systems, 33*(3), 38-53.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Inexperience and experience with online stores: The importance of TAM and trust. *IEEE Transactions on Engineering Management, 50*(3), 307-321.

Gefen, D., & Straub, D. W. (2004). Consumer trust in B2C e-Commerce and the importance of social presence: Experiments in e-Products and e-Services. *Omega, 32*(6), 407-424.

Geyskens, I., Steenkamp, J. B. E. M., & Kumar, N. (1998). Generalizations about trust in marketing channel relationships using meta-analysis. *International Journal of Research in Marketing, 15*(3), 223-248.

Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science, 23*(4), 545-560. doi: 10.1287/mksc.1040.0071

Grabner-Kraeuter, S. (2002). The role of consumers' trust in online-shopping. *Journal of Business Ethics, 39*(1-2), 43-50. doi: 10.1023/a:1016323815802

Gretzel, U., & Fesenmaier, D. R. (2006). Persuasion in recommender systems. *International Journal of Electronic Commerce, 11*(2), 81-100. doi: 10.2753/jec1086-4415110204

Gutowska, A., Sloane, A., & Buckley, K. A. (2009). On Desideratum for B2C E-Commerce Reputation Systems. *Journal of Computer Science and Technology, 24*(5), 820-832.

Haas, D. F., & Deseran, F. A. (1981). TRUST AND SYMBOLIC EXCHANGE. *Social Psychology Quarterly, 44*(1), 3-13. doi: 10.2307/3033857

Haefliger, S., Monteiro, E., Foray, D., & von Krogh, G. (2011). Social Software and Strategy. *Long Range Planning, 44*(5), 297-316.

Hoffman, D. L., Novak, T. P., & Peralta, M. (1999). Building consumer trust online. *Communications of the ACM, 42*(4), 80-85.

Houser, D., & Wooders, J. (2006). Reputation in auctions: Theory, and evidence from eBay. *Journal of Economics & Management Strategy, 15*(2), 353-369. doi: 10.1111/j.1530-9134.2006.00103.x

Hu, N., Bose, I., Koh, N. S., & Liu, L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems, 52*(3), 674-684. doi: 10.1016/j.dss.2011.11.002

Hussey, J., & Hussey, R. (1997). *Business research A practical guide for undergraduate and postgraduate students MacMillan Business*: Macmillan Publishers Limited.

Jarvenpaa, S. L., & Toad, P. A. (1996). Consumer reactions to electronic shopping on the World Wide Web. *International Journal of Electronic Commerce, 1*(2), 59-88.

Jarvenpaa, S. L., Tractinsky, N., & Vitale, M. (2000). Consumer Trust in an Internet Store. *Information Technology and Management, 1*(1-2), 45-71.

Jensen, C., Potts, C., & Jensen, C. (2005). Privacy practices of Internet users: Self-reports versus observed behavior. *International Journal of Human Computer Studies, 63*(1-2), 203-227.

Ji, S. X., & Wang, X. M. (2011). *Research of online trust in C2C electronic commerce based on website institutions.* Paper presented at the ICEE 2011 International Conference.

Jindal, N., & Liu, B. (2007). *Analyzing and detecting review spam.* Paper presented at the Data Mining, 2007. ICDM 2007.Seventh IEEE International Conference.

Jindal, N., & Liu, B. (2008). *Opinion spam and analysis.* Paper presented at the WSDM 2008 Proceedings of the 2008 International Conference.

Joinson, A. N., Reips, U. D., Buchanan, T., & Schofield, C. B. P. (2010). Privacy, trust, and self-disclosure online. *Human-Computer Interaction, 25*(1), 1-24.

Jones, K., & Leonard, L. N. K. (2008). Trust in consumer-to-consumer electronic commerce. *Information and Management, 45*(2), 88-95.

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.

Jøsang, A., & Pope, S. (2005). *Semantic constraints for trust transitivity*. Paper presented at the Conferences in Research and in Information Technology Series.

Kambil, A., & Van Heck, E. (1998). Reengineering the Dutch Flower Auctions: A Framework for Analyzing Exchange Organizations. *Information Systems Research, 9*(1), 1-19.

Kee, H. W., & Knox, R. E. (1970). Conceptual and methodological considerations in the study of trust and suspicion. *Journal of Conflict Resolution, 14*(3), 357-366.

Kim, D. J., Ferrin, D. L., & Rao, H. R. (2008). A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision Support Systems, 44*(2), 544-564.

Komiak, S. Y. X., & Benbasat, I. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS Quarterly, 30*(4), 941-960.

Koufaris, M., & Hampton-Sosa, W. (2004). The development of initial trust in an online company by new customers. *Information & Management, 41*(3), 379-397. doi: 10.1016/j.im.2003.08.004

Krasnova, H., Spiekermann, S., Koroleva, K., & Hildebrand, T. (2010). Online social networks: why we disclose. *Journal of Information Technology, 25*(2), 109-125. doi: 10.1057/jit.2010.6

Kumar, R., & Nti, K. O. (1998). Differential learning and interaction in alliance dynamics: A process and outcome discrepancy model. *Organization Science, 9*(3), 356-367. doi: 10.1287/orsc.9.3.356

Lau, R. Y. K., Liao, S. Y., Chi-Wai Kwok, R., Xu, K., Xia, Y., & Li, Y. (2011). Text mining and probabilistic language modeling for online review spam detection. *ACM Transactions on Management Information Systems, 2*(4), 1-30.

Lauer, T. W., & Deng, X. (2007). Building online trust through privacy practices. *International Journal of Information Security, 6*(5), 323-331.

Lee, M. K., & Turban, E. (2001). A trust model for consumer internet shopping. *International Journal of Electronic Commerce, 6*(1), 75-92.

Liao, C., Palvia, P., & Lin, H. N. (2006). The roles of habit and web site quality in e-commerce. *International Journal of Information Management, 26*(6), 469-483. doi: 10.1016/j.ijinfomgt.2006.09.001

Lindgaard, G., & Dudek, C. (2003). What is this evasive beast we call user satisfaction? *Interacting with Computers, 15*(3), 429-452.

Liu, L., & Munro, M. (2012). Systematic analysis of centralized online reputation systems. *Decision Support Systems, 52*(2), 438-449.

Liu, L., Munro, M., & Song, W. (2010). *Evaluation of collecting reviews in centralized online reputation systems.* Paper presented at the WEBIST 2010 - Proceedings of the 6th International Conference on Web Information Systems and Technology.

Lua, E. K., Chen, R., & Cai, Z. (2011). *Social Trust and Reputation in Online Social Networks.* Paper presented at the 2011 IEEE 17th International Conference on Parallel and Distributed Systems.

Luhmann, N., Davis, H., Raffan, J., & Rooney, K. (1979). *Trust and Power: two works by Niklas Luhmann*: Wiley Chichester.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709-734.

Mayzlin, D. (2006). Promotional chat on the internet. *Marketing Science, 25*(2), 155-163.

McKnight, D. H., & Chervany, N. L. (2002). What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology. *International Journal of Electronic Commerce, 6*(2), 35-59.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2000). *Trust in e-commerce vendors: a two-stage model.* Paper presented at the Proceedings of the twenty first international conference on Information systems.

McKnight, D. H., Kacmar, C. J., & Choudhury, V. (2004). Shifting Factors and the Ineffectiveness of Third Party Assurance Seals: A Two-Stage Model of Initial Trust in a Web Business. *Electronic Markets, 14*(3), 252-266.

Mesch, G. S. (2012). Is online trust and trust in social institutions associated with online disclosure of identifiable information online? *Computers in Human Behavior, 28*(4), 1471-1477. doi: 10.1016/j.chb.2012.03.010

Metzger, M. J. (2006). Effects of site, vendor, and consumer characteristics on web site trust and disclosure. *Communication Research, 33*(3), 155-179. doi: 10.1177/0093650206287076

Mollering, G. (2006). Trust, institutions, agency: towards a neoinstitutional theory of trust *Handbook of trust research* (pp. 355 - 376).

Moorman, C., Zaltman, G., & Deshpande, R. (1992). Relationships Between Providers and Users of Market Research: The Dynamics of Trust. *Journal of marketing research, 29*, 314-328.

Morys, A. (2011). *Conversion-Optimierung: Praxismethoden für mehr Markterfolg im Web*: Entwickler. Press.

Mudambi, S. M., & Schuff, D. (2010). WHAT MAKES A HELPFUL ONLINE REVIEW? A STUDY OF CUSTOMER REVIEWS ON AMAZON.COM. *MIS Quarterly, 34*(1), 185-200.

Mukherjee, A., Liu, B., & Glance, N. (2012). *Spotting fake reviewer groups in consumer reviews.* Paper presented at the WWW '12 Proceedings of the 21st international conference.

Murphy, G. B., & Tocher, N. (2011). Gender differences in the effectiveness of online trust building information cues: An empirical examination. *Journal of High Technology Management Research, 22*(1), 26-35.

Nosko, A., Wood, E., & Molema, S. (2010). All about me: Disclosure in online social networking profiles: The case of FACEBOOK. *Computers in Human Behavior, 26*(3), 406-418.

Pan, Y., & Zinkhan, G. M. (2006). Exploring the impact of online privacy disclosures on consumer trust. *Journal of Retailing, 82*(4), 331-338.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*: SAGE Publications, inc.

Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce, 7*(3), 101-134.

Pavlou, P. A., & Gefen, D. (2004). Building effective online marketplaces with institution-based trust. *Information Systems Research, 15*(1), 37-59. doi: 10.1287/isre.1040.0015

Perdikaki, O., Kesavan, S., & Swaminathan, J. M. (2012). Effect of Traffic on Sales and Conversion Rates of Retail Stores. *Manufacturing & Service Operations Management, 14*(1), 145-162. doi: 10.1287/msom.1110.0356

Poston, R. S., & Speier, C. (2005). Effective use of knowledge management systems: a process model of content ratings and credibility indicators. *MIS Quarterly, 29*(2), 221-244.

Rasmusson, L., & Jansson, S. (1996). *Simulated social control for secure Internet commerce.* Paper presented at the Proceedings of the 1996 workshop on New security paradigms.

Ratnasingam, P., & Phan, D. D. (2003). Trading partner trust in B2B E-commerce: A case study. *Information Systems Management, 20*(3), 39-50.

Reichheld, F. F., & Schefter, P. (2000). E-Loyalty: Your secret weapon on the web. *Harvard Business Review, 78*(4), 105-113.

Remenyi, D., Williams, B., Money, A., & Swartz, E. (1998). *Doing research in business and management: an introduction to process and method*: Sage Publications Limited.

Resnick, P., Kuwabara, K., Zeckhauser, R., & Friedman, E. (2000). Reputation systems. *Communications of the ACM, 43*(12), 45-48.

Ridings, C. M., Gefen, D., & Arinze, B. (2002). Some antecedents and effects of trust in virtual communities. *Journal of Strategic Information Systems, 11*(3-4), 271-295.

Rosseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: Across-discipline view of trust. *Academy of Management Review, 23*(3), 393-404.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of personality, 35*(4), 651-665.

Salam, A. F., Iyer, L., Palvia, P., & Singh, R. (2005). Trust in e-commerce. *Communications of the ACM, 48*(2), 73-77.

Saunders, M. N. K., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*: Pearson Education.

Schlosser, A. E., White, T. B., & Lloyd, S. M. (2006). Converting web site visitors into buyers: How web site investment increases consumer trusting beliefs and online purchase intentions. *Journal of Marketing, 70*(2), 133-148.

Schurr, P. H., & Ozanne, J. L. (1985). INFLUENCES ON EXCHANGE PROCESSES - BUYERS PRECONCEPTIONS OF A SELLERS TRUSTWORTHINESS AND BARGAINING TOUGHNESS. *Journal of Consumer Research, 11*(4), 939-953. doi: 10.1086/209028

Shankar, V., Urban, G. L., & Sultan, F. (2002). Online trust: a stakeholder perspective, concepts, implications, and future directions. *Journal of Strategic Information Systems, 11*(3-4), 325-344. doi: 10.1016/s0963-8687(02)00022-7

Shen, X., & Zhong, Y. (2008). *Empirical research in the antecedents of online trust in C2C market*. Paper presented at the 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing, Vols 1-31. <Go to ISI>://WOS:000263466104199

Sillence, E., Briggs, P., Harris, P., & Fishwick, L. (2007). Health Websites that people can trust - the case of hypertension. *Interacting with Computers, 19*(1), 32-42. doi: 10.1016/j.intcom.2006.07.009

Sismeiro, C., & Bucklin, R. E. (2004). Modeling purchase behavior at an E-commerce web site: A task-completion approach. *Journal of marketing research, 41*(3), 306-323. doi: 10.1509/jmkr.41.3.306.35985

Smith, D., Menon, S., & Sivakumar, K. (2005). Online peer and editorial recommendations, trust, and choice in virtual markets. *Journal of Interactive Marketing, 19*(3), 15-37.

Sultan, F., Urban, G., Shankar, V., & Bart, Y. (2002). Determinants and role of trust in e-business: A large scale empirical study. *Working Paper, Sloan School of Management, MIT, Cambridge, MA 02142*.

Sun, Y., & Liu, Y. (2012). Security of online reputation systems: The evolution of attacks and defenses. *IEEE Signal Processing Magazine, 29*(2), 87-97.

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers, 13*(2), 127-145. doi: 10.1016/s0953-5438(00)00031-x

Tullberg, J. (2008). Trust-The importance of trustfulness versus trustworthiness. *Journal of Socio-Economics, 37*(5), 2059-2071.

Tzafrir, S. S. (2005). The relationship between trust, HRM practices and firm performance. *The International Journal of Human Resource Management, 16*(9), 1600-1622.

Urban, G. L., Amyx, C., & Lorenzon, A. (2009). Online Trust: State of the Art, New Frontiers, and Research Potential. *Journal of Interactive Marketing, 23*(2), 179-190.

Voss, C., Tsikriktsis, N., & Frohlich, M. (2002). Case research in operations management. *International Journal of Operations & Production Management, 22*(2), 195-219. doi: 10.1108/01443570210414329

Wang, H. Q., Lee, M. K. O., & Wang, C. (1998). Consumer privacy concerns about Internet marketing. *Communications of the ACM, 41*(3), 63-70. doi: 10.1145/272287.272299

Whitby, A., Jøsang, A., & Indulska, J. (2004). *Filtering out unfair ratings in bayesian reputation systems.* Paper presented at the The third international joint conference on autonomous agent systems.

Wijnhoven, A. B. J. M. (2012). *Information services design: A design science approach for sustainable knowledge*: Routledge New York.

Wijnhoven, A. B. J. M., Ehrenhard, M. L., & Alink, T. (2012). *High Reliable Internet Hosting: A Case Study of Handling Unanticipated Threats to Reliability*. Paper presented at the International Conference on Information Systems, Orlanda, USA.

Witkowski, T. H., & Thibodeau, E. J. (1999). Personal bonding processes in international marketing relationships. *Journal of Business Research, 46*(3), 315-325. doi: 10.1016/s0148-2963(98)00031-9

Wu, G., Hu, X., & Wu, Y. (2010). Effects of Perceived Interactivity, Perceived Web Assurance and Disposition to Trust on Initial Online Trust. *Journal of Computer-Mediated Communication, 16*(1), 1-26. doi: 10.1111/j.1083-6101.2010.01528.x

Yang, Y., Hu, Y., & Chen, J. (2005). *A web trust-inducing model for e-commerce and empirical research.* Paper presented at the Seventh International Conference on Electronic Commerce, Vols 1 and 2, Selected Proceedings.

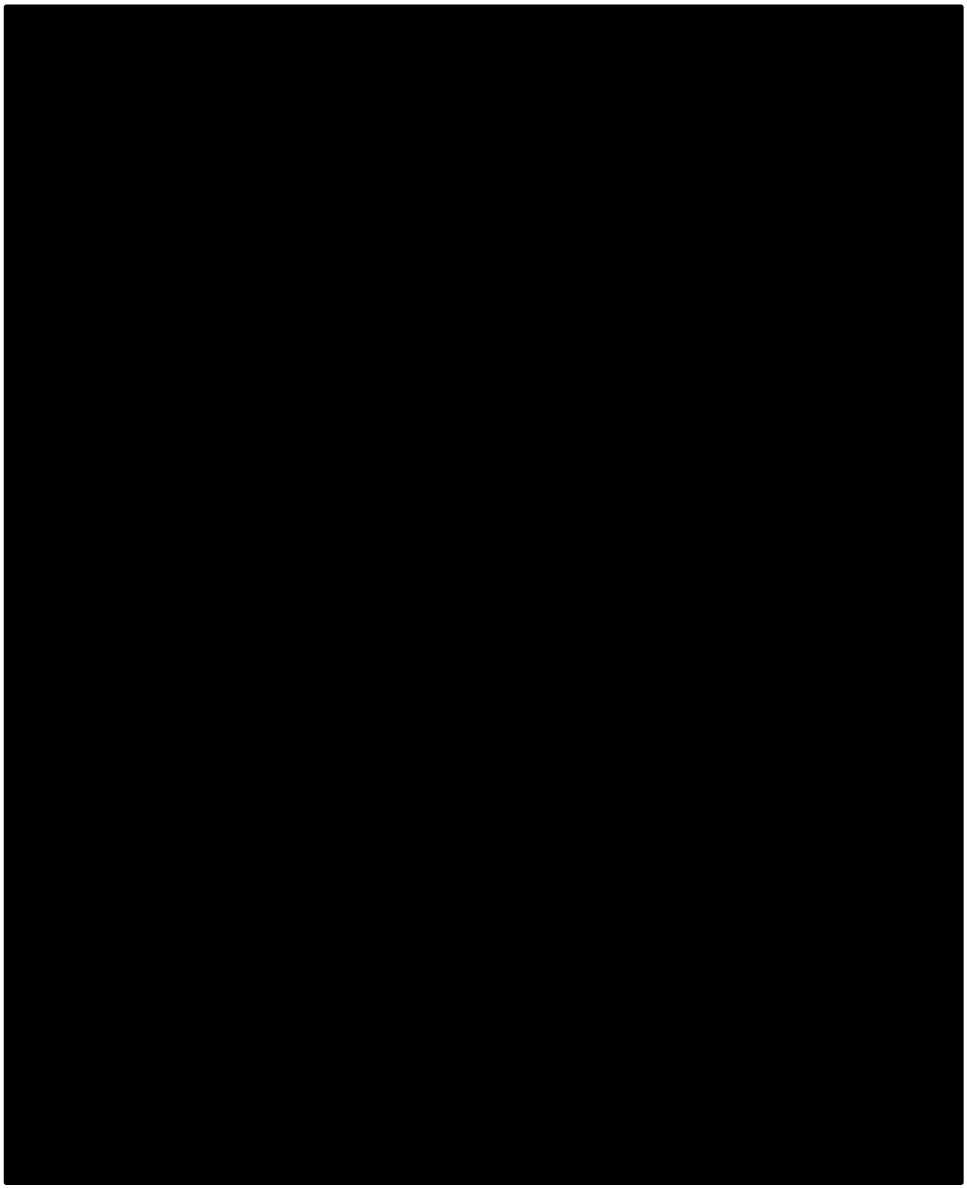
Yin, R. K. (2003). Case study research: Design and methods (applied social research methods): Thousand Oaks. California: Sage Publications.

Yoon, S. J. (2002). The antecedents and consequences of trust in online-purchase decisions. *Journal of Interactive Marketing, 16*(2), 47-63.

Zhang, Y., Bian, J., & Zhu, W. (2013). Trust fraud: A crucial challenge for China's e-commerce market. *Electronic commerce research and applications, (in press)*.

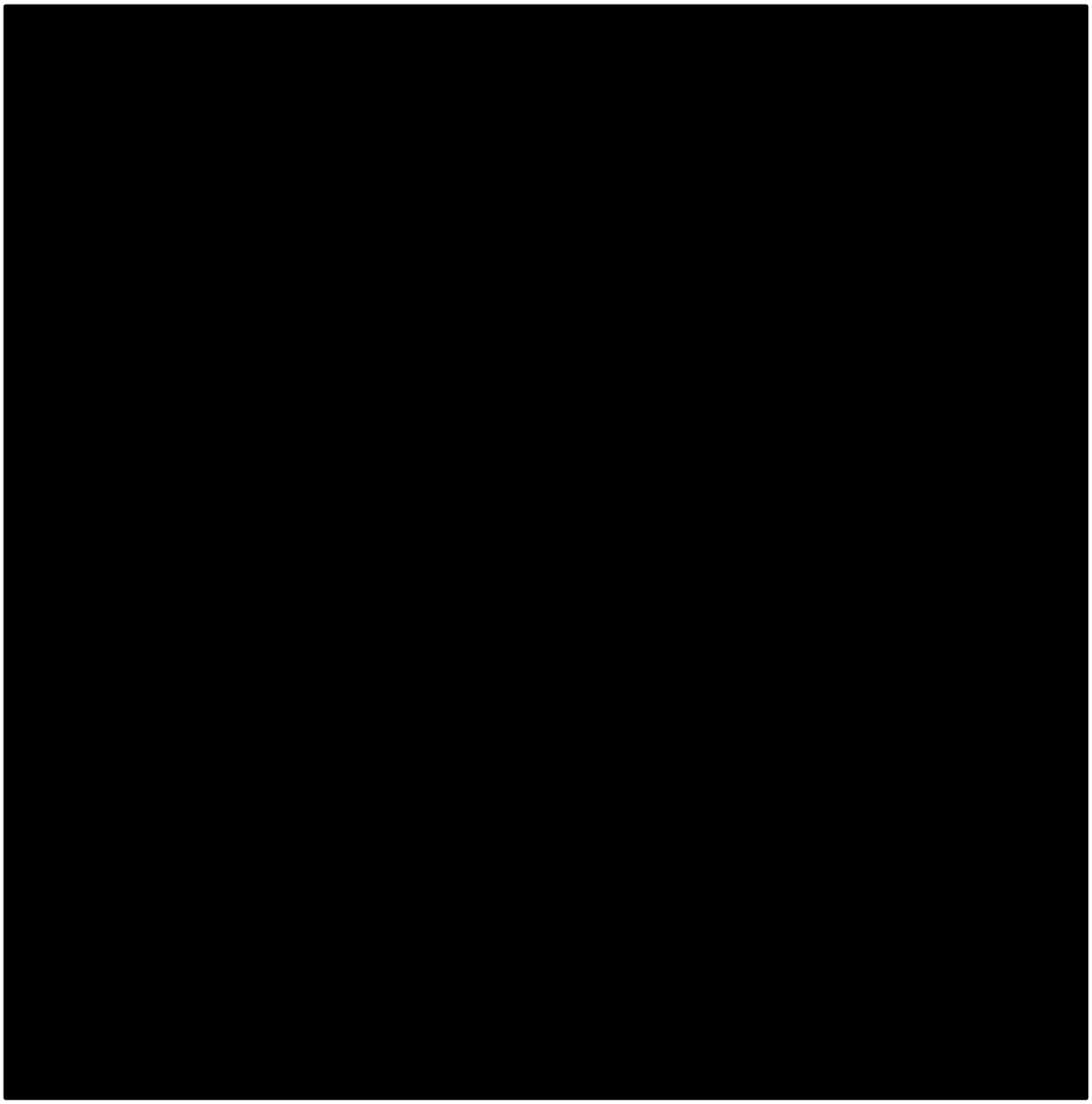
# APPENDIX

## A 1: Rating Form for Test Sample One

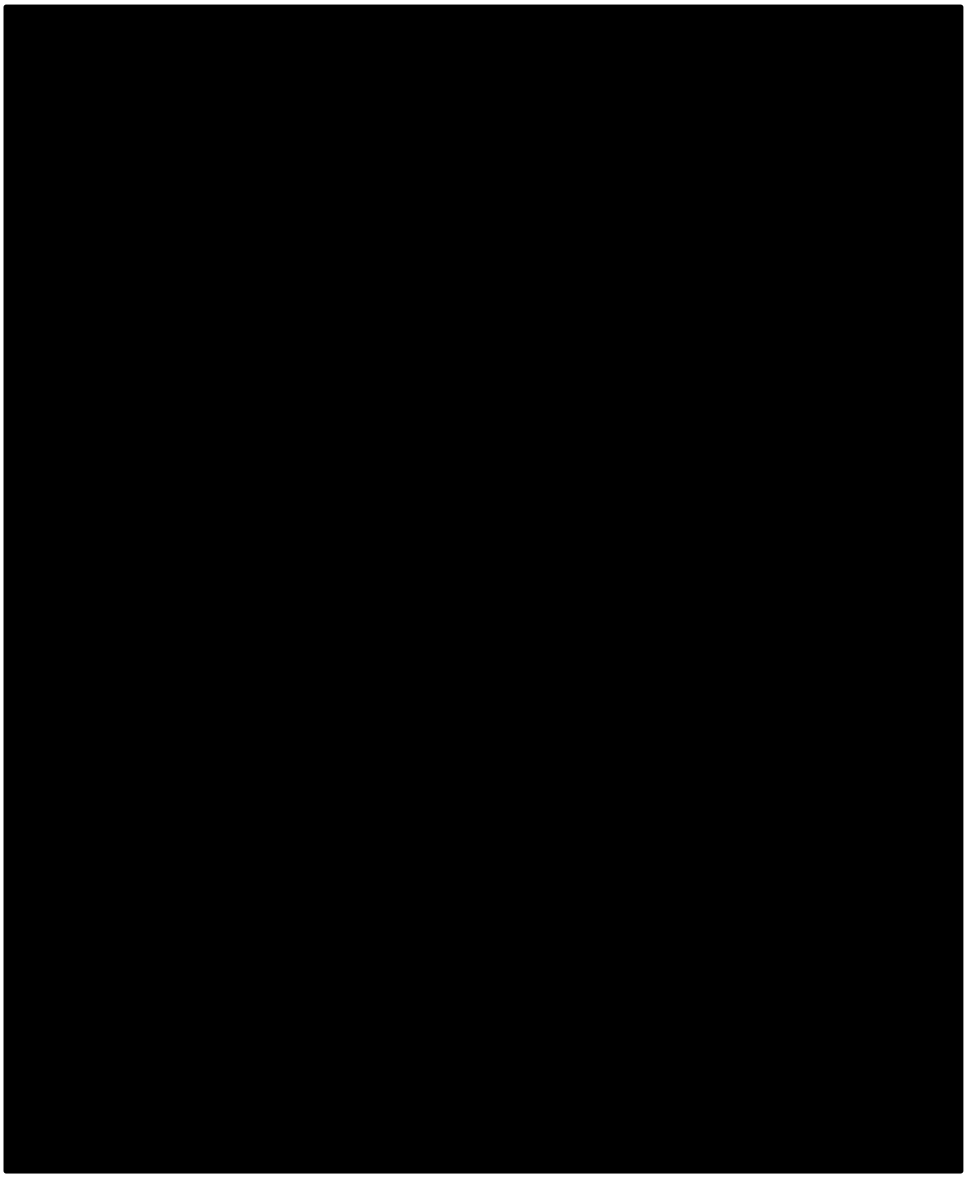


**A 2: Rating Form for Test Sample Two**

**A 3: A/B/C-Test Integration on List-View**

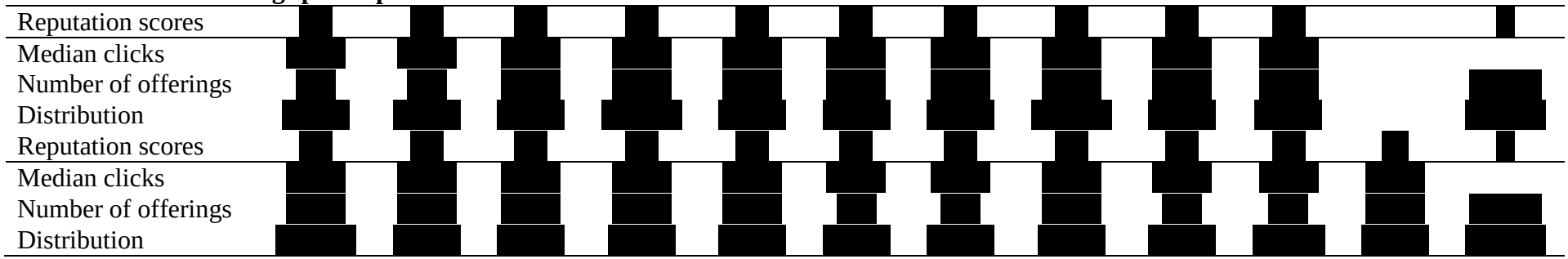


**A 4: A/B/C-Test Integration on Details-View**



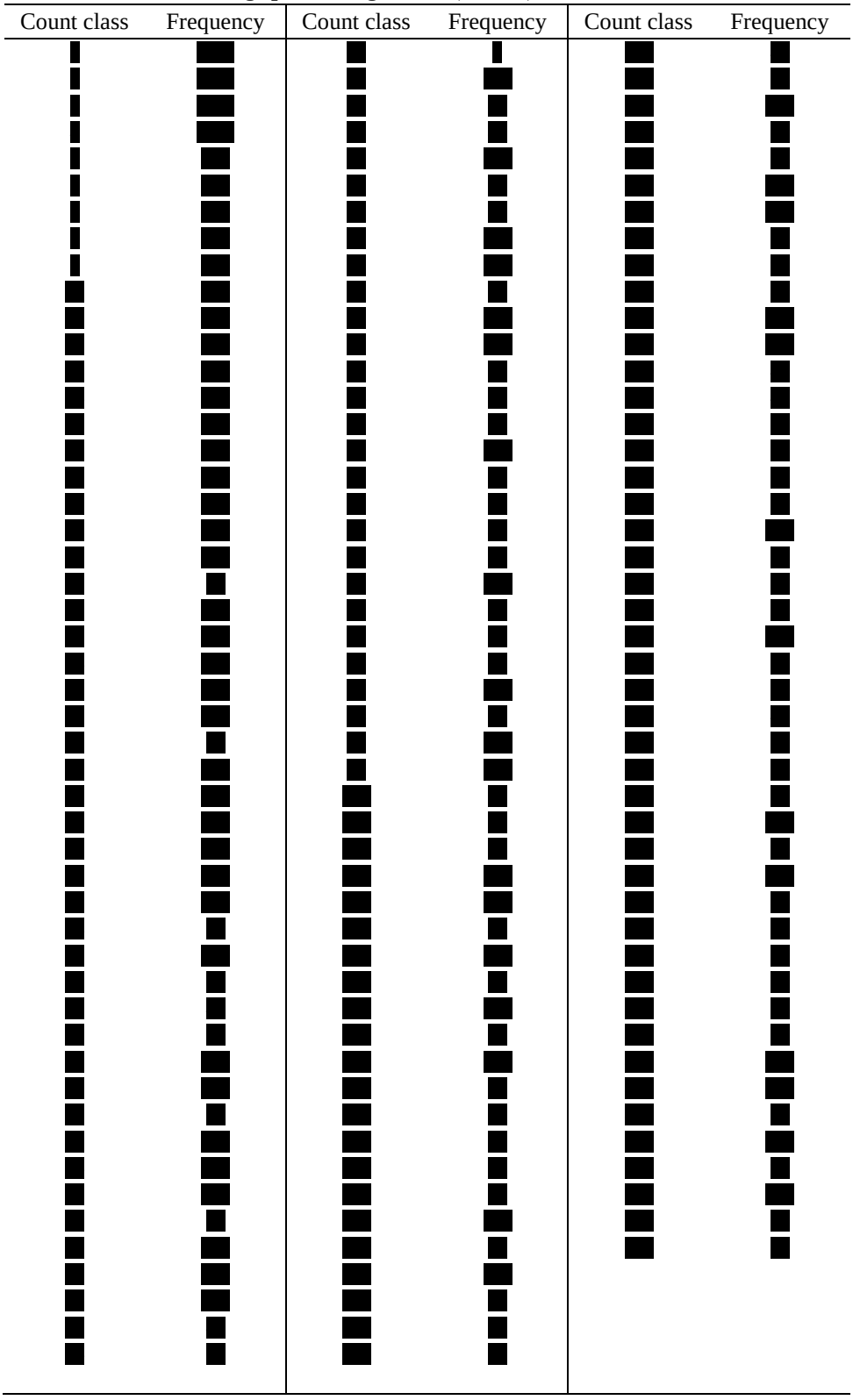
**A 5: Distribution of Offerings per Reputation Score**

**Distribution of Offerings per Reputation Score**



## A 6: Distribution of Offerings per Rating Count Class (1 – 999)

**Distribution of Offerings per Rating Count (1 - 999)**



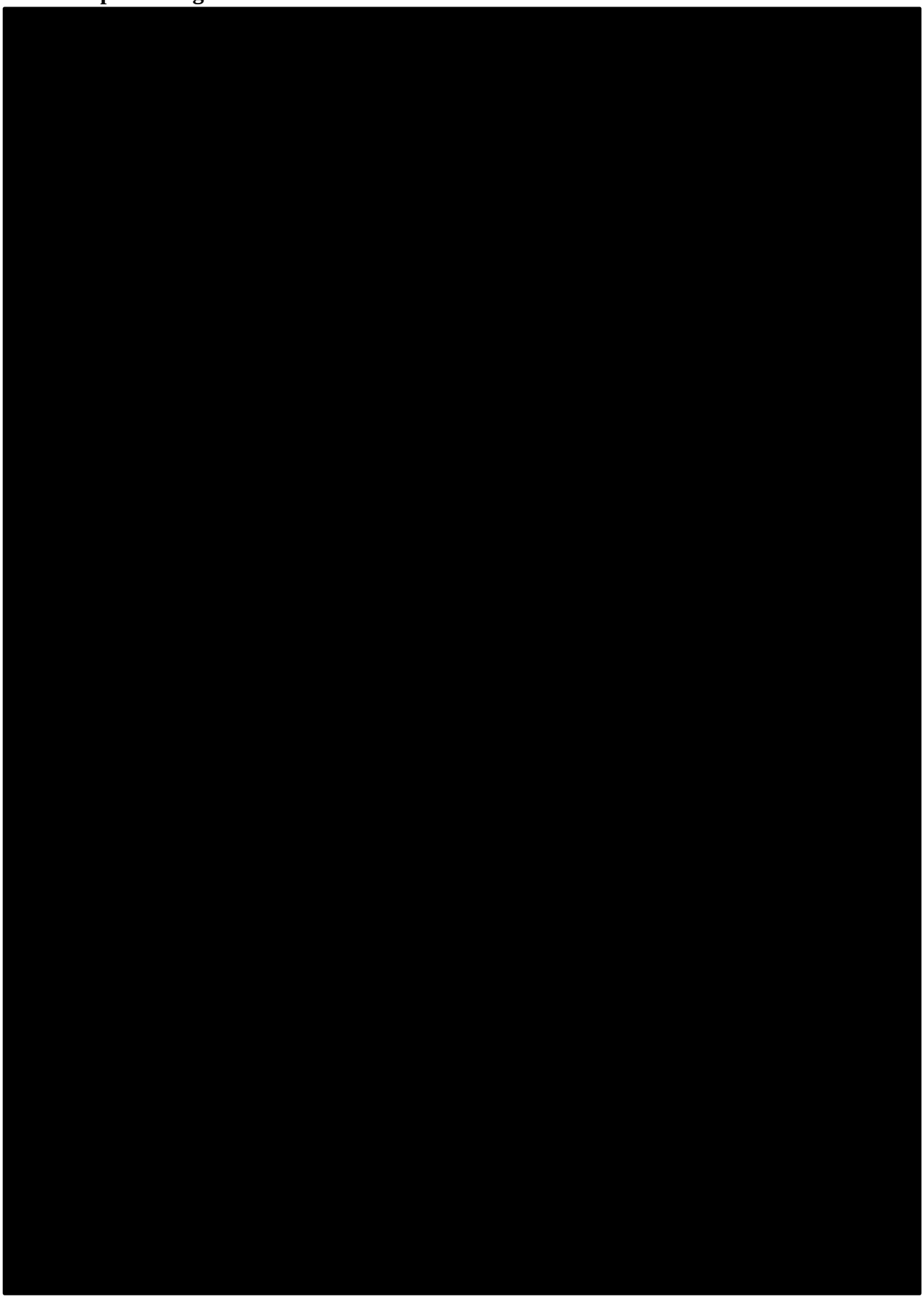
## A 7: Scanning for Type 2 and 3 Review-spam

Procedure:

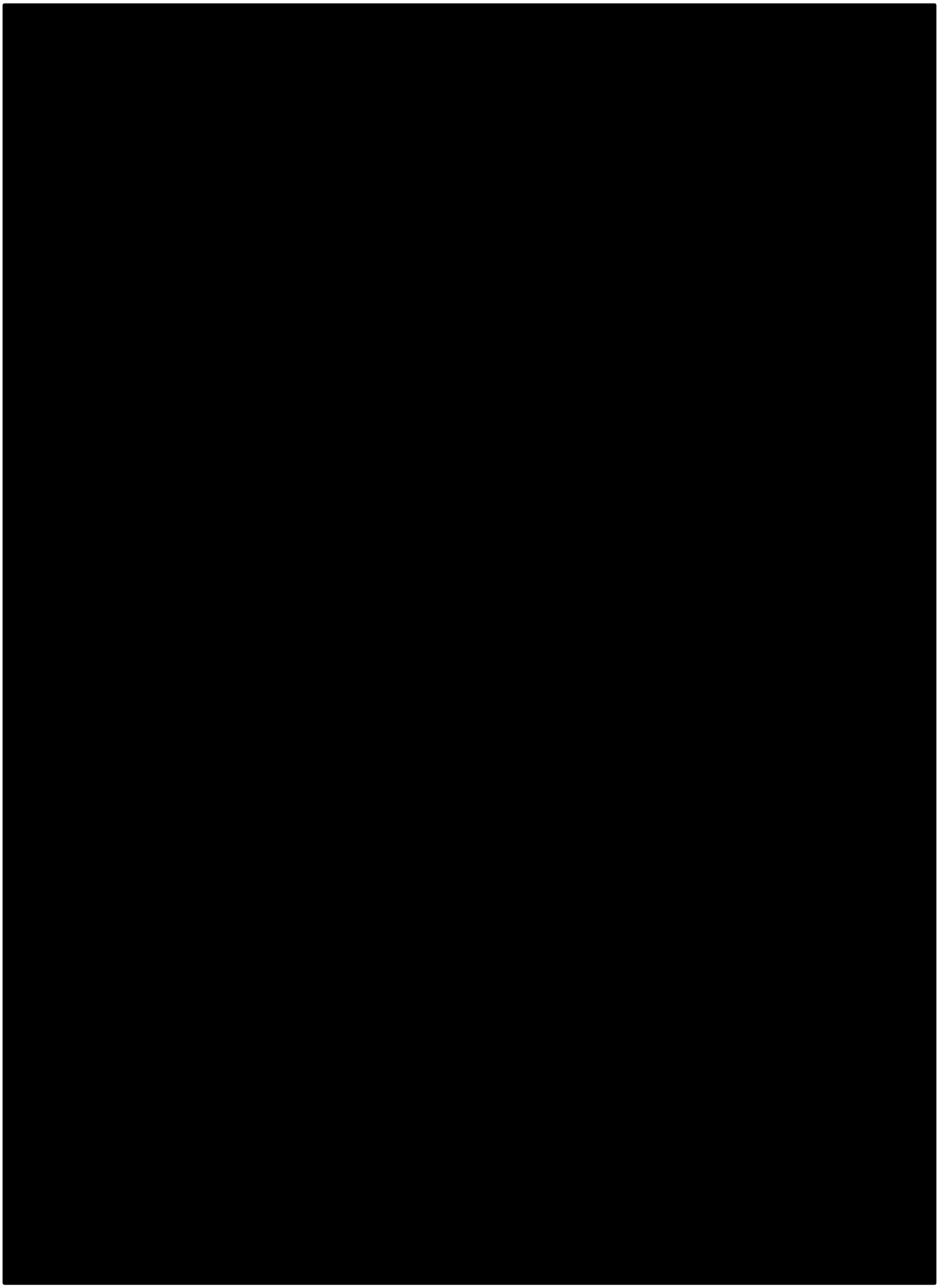
The company's review database was converted into MS Excel in order to enable editing of the data. All ratings without a free-text review as well as unnecessary information were dismissed. The preprocessing enabled to work with the above visualized view. All of [redacted]-text reviews were now scanned manually each after another for type 2 and 3 review-spam. Additional information as the numerical reputation scores, rating-date or service type supported the manual evaluation process. Identified non-reviews were color-coded. After the examination of the whole dataset, filtering for the color-coded reviews allowed to make an overview of all targets.

The criteria by which the non-reviews were labeled correspond to the approach presented by Jindal and Liu (2007).

## A 8: Preprocessing of the Free-Text User Feedback



**A 9: N-Gram Generation**



**A 10: N-Gramm Matching**