BI Norwegian Business School - campus Oslo

GRA 19703

Master Thesis

Thesis Master of Science

Chatbots in service recovery: Crackpot or Jackpot?

Navn:	Peter Daniel Lund Magnusson, Arne Aleksander Rånnerud

Start:	15.01.2019 09.00
Finish:	01.07.2019 12.00

Chatbots in service recovery: Crackpot or Jackpot?

A comparative investigation of the impact of chatbots and human chat agents in service recovery on firm and recovery satisfaction.

Date of submission:
26.06.2019

Deadline:
01.07.2019

Campus:
BI Oslo

Exam code and name:
GRA 19703 Master Thesis

Program:
Master of Science in Business with major in Marketing

Supervisor:
Hannah Snyder

Acknowledgements

This master thesis completes our Master of Science degree in Business with a major in marketing at BI Norwegian Business School.

First of all, we would like to offer our deepest gratitude to our supervisor Hannah Snyder for all her support, guidance and feedback on the thesis. Her advice and supervision was extremely valuable in the completion of our thesis, and we would not have managed to come up with the same results without her help. Your time and effort have been highly appreciated.

Secondly, a big thanks to Adam Tyrcha is in order. We are forever in your debt for the time you took out of you busy schedule, to help us. Your comments, thoughts and eye for details has truly been inspirational and utmost appreciated.

Lastly, we would like to thank all our family and friends for support and motivations during our studies at BI and especially in the most difficult and struggling periods. Finally, we want to give huge thanks to all the participants who completed our questionnaire and a special thanks to those who completed our pre-test and provided us with feedback on improvements. Now, we are both really excited to bring our valuable knowledge from BI further into work and practise.

Best regards,

Arne Rånnerud & Peter Magnusson

Abstract

Problem Service recovery has been a heavily researched and studied area for several years. As technology has continued to develop and been adapted and incorporated into this field, new research avenues have also opened up, many of which have not been studied in much depth as of yet. With very limited research on firms using chatbots as a service encounter in service recovery, we wanted to investigate this further. Therefore, we wanted to check if there was any difference between a customer interacting with a human or a chatbot in a service recovery situation and measure this with regard to satisfaction with the firm and recovery.

Purpose The purpose of this study is to analyse the effect of an unsuccessful and a successful service recovery, provided by either a human, chatbot or a combination of the two, on satisfaction with the firm and service recovery.

Research Design A scenario-based survey experiment was chosen to answer the research problem and questions for this thesis. The design used was a 3 (Chatbot, Human, Chatbot+Human) x 2 (Unsuccessful, Successful) between subject design.

Findings Customers are proven to be more satisfied when interacting with a chatbot in a successful service recovery, than interacting with a human or a combination of the two. However, when the customers interact with a chatbot in a unsuccessful recovery, they are even more unsatisfied than when interacting with a human or the combination. We also found a significant relationship between the outcome variable and satisfaction with recovery and firm, and a significant interaction effect between the outcome and encounter variables and satisfaction with firm and recovery.

Keywords Service recovery, customer satisfaction, chatbot, human, service failure, E-commerce, anthropomorphism,

Table of Contents

1.0 Introduction

Global economies continue to grow, with the service industry being one of the primary drivers of this growth. Indeed, 70% of global GDP is accounted for by the service industry, and this is expected to continue to grow even further (Wan & Chan, 2019). There has been a rapid development of service technologies which has spurred this immense growth (Wan & Chan, 2019). On the background of this development, the traditional service encounter has been fundamentally changed (Larivière et al., 2017). The improvement of communication and information technologies is changing how customers are interacting with each other, and with service providers, which again may influence how customers perceive the entire service experience (Froehle & Roth, 2004). Similarly, consumers are purchasing increasing amounts of goods and services on the internet, meaning there is less physical interaction between an employee and the customer. Concurrently, chatbots have increasingly begun to be used in the service industry. A chatbot is defined as “a computer program designed to simulate conversation with human users, especially over the internet” (MyClever et al., 2018). Many other terms are also used instead of chatbots such as: virtual agent, dialogue system, machine conversation system, conversational agent etc. (Shawar & Atwell, 2007). The development of chatbots can be seen as a part of reducing the extent of human interaction, which has been considered a crucial part of the traditional service experience (Holloway & Beatty, 2003). Using chatbots in the service encounter is one of many examples in which companies turn customer service into self-service (Rust and Huang, 2018: Fluss 2017).

Chatbots rely on Artificial Intelligence (AI) or machine learning (ML) in order to simulate human communication. The design of natural language processing (NLP) is the form of AI that allows chatbots to understand human speech and interact with human beings. Meanwhile, machine learning helps chatbots to improve themselves and learn through communication with customers (Campbell-Miller, 2017; Ciechanowski et al., 2018). The relative intelligence of the chatbot influences how human the conversation feels. Moreover, there is a consensus among customers that chatbots should be able to solve both complicated and less complicated cases. Due to the technological advances, customers have higher expectations when interacting in online environments and require a higher level of service quality (Gronroos &

Voima, 2013). The report made by MyClever, Drift, Salesforce, & SurveyMonkey Audience (2018), shows that one of the potential problems people expect a chatbot to solve, is where a complaint needs to be resolved (35% of people surveyed). In accordance with previous studies, the report displays what people expect a chatbot or virtual assistant to be like. To our knowledge, there is relatively limited evidence of companies using chatbots in complaint handling situations, but as mentioned, people have set expectations regarding what a chatbot should be able to deliver. Some respondents propose a chatbot should be smart, high performing, seamless and personable (Samora, 2018), whereas others want to interact with someone who can show empathy, address complex needs and offer assurance (Arcand, 2017).

There is a long list of potential benefits chatbots possibly could offer to both service providers and their customers. Chatbots can provide service in all hours of the day, and answer customer inquiries instantly. However, the main motivations behind implementing chatbots can be said to be cost driven. Valuable and expensive human work power can be placed elsewhere to do more pressing and important tasks, whereas chatbots can answer the more basic and simplistic inquiries. Nonetheless, chatbots do not always meet customer needs and expectations (Mimoun, Poncin, & Garnier, 2012). Unsatisfactory online service encounters have been found to negatively affect word of mouth, loyalty and intention to repurchase a product (Oliver, 1997). In order to turn the unsatisfactory online service encounters to satisfactory encounters, the service provider should be able to offer good service recovery procedures in the form of compensation, discounts or apologies (Kelley, Hoffman, & Davis, 1993).

The present study makes several important contributions. Firstly, by bridging service recovery with chatbots, we add a new perspective to previous research and help explain how technology influences service recovery encounters. We add knowledge to previous theories regarding the interplay of humans and technology and challenge them by introducing service recovery attempts handled by chatbots. The reason for this is that customers, companies and technology have new roles in the service encounter due to a technological shift (Laviere et. al., 2017). Previous research has identified user intentions and potential benefits of using chatbots in service encounters, but research regarding service recovery in the lens of chatbots is to our knowledge very limited. For instance, an organized service recovery policy is crucial in order to maintain satisfied and loyal customers. With this in mind, it is of considerable importance to examine how consumers actually respond when encountering a chatbot in a complaint situation. If customers are not satisfied, loyalty is reduced, and an increase in switching behavior and negative word of mouth may happen, all of which have costs attached to them. Thus, in our research we will investigate whether the potential benefits and motivations for implementing chatbots in fact serve their true purpose.

We will also offer several managerial and theoretical contributions, with concrete advice on how service firms should address the increasing use of technology. Our study shows that there is a significant difference in level of satisfaction with the firm and the recovery, in a successful or unsuccessful encounter, depending on if the customer is faced with a chatbot or a human chat agent. For future managers this means that one needs to be observant when implementing and replacing traditional customer service jobs with technology. We also demonstrate the importance of chatbots working together with humans. Since chatbots is a relatively new field of research, our thesis will outline in great extent future topics that need to be evaluated in the lens of chatbots.

Our thesis is organized as following. The second section will give an overview of the fields of service recovery, satisfaction constructs, the changing environment of technology and the interaction between chatbots and humans. The research presented in this section is evaluated, in order to present four hypotheses. Our method and data collection is described in section 3, while our results are presented in section 4. We discuss our results with regard to our hypotheses in section 5, followed by managerial implications, limitations, and a list of suggestions for further research.

1.1 Research Objective and Research Question

1.1.1 Research objective and contribution

There has been extensive research conducted into service recovery in the past, whereas for chatbots the body of literature is increasing. However, the two combined have yet to be researched. We therefore find it necessary to extend current literature regarding service recovery and investigate it in regard to chatbots. Our overall research objective is to gain greater knowledge about how customer satisfaction is affected during service recovery situations which are facilitated by technology in various degrees. To do so, we will investigate how customers respond differently in encounters performed by a chatbot or a human chat agent, and when the encounters are either successful or unsuccessful. Hence, our aim is to test whether the complete loss of the human element affects the customers satisfaction towards the firm, as well as the recovery itself. Also, chatbots and human chat agents in symbiosis are included in order to see if the human element in some form is still needed to maintain satisfied customers in service encounters.

With our thesis, we aim to contribute new theoretical insight to a new field of research, involving chatbots and complaints in customer service encounters. In conjunction to our theoretical contribution, our thesis will provide useful knowledge for managers and developers of chatbots. An understanding of service recovery through chatbots is required for managers when implementing new technologies in their businesses, as well as regarding hiring and reallocation of staff.

1.1.2 Research question

To allow for a better understanding of this area of research, we have come up with these research questions:

> *RQ1: How will satisfaction with firm and recovery be affected by a service recovery attempt performed by a chatbot, as opposed to human chat agents?*
>
> *RQ2: How will satisfaction with firm and recovery be affected by a service recovery attempt performed by a chatbot and a human chat agent working together?*

2.0 Research Background

2.1 Service encounter 2.0

The service encounter 2.0 is defined as: *"any customer-company interaction that results from a service system that is comprised of interrelated technologies, human actors, physical/digital environments and company/customer processes"* (Laviere et al., 2017). What is more, is that technology is fundamentally changing the nature of the service encounter (Laviere et al., 2017). Through rapid technological advances, such as the development of AI and more novel technologies such as smartphones, the classical roles of employees, customers and technology in the service encounter are changing (Laviere et al., 2017).

The findings of Laviere et. al. (2017) indicate that technology can strengthen or substitute the classical service employee, which in turn leads to customers and employees taking on roles as an enabler, innovator, coordinator or differentiator. Hence, the interaction between customers and companies has changed, which further adds on to the ideas of how customers and companies' interplay (Laviere et. al., 2017: Shostack, 1985). In the light of the technological "evolution" customers are now engaging in a "quasi social relationship" with new forms of artificial intelligent beings, such as computers (Biocca & Harms, 2002). Companies are interacting with customers through technology, customers are interacting with companies through technology, and people are increasingly communicating with one another through technology-based medias (Shankar et al., 2016).

2.2 Service Recovery

For a company to succeed today, customer satisfaction and loyalty is crucial. One aspect of retaining and improving the satisfaction and loyalty of the customers, is to adopt a well-functioning recovery process after a service failure. Service recovery refers to the action taken by the company providing the service, in order to solve the customer complaint originating from a service failure (Grönroos, 1988). Service failures are often impossible to avoid due to non-human and human errors, and therefore these types of failures will often lead to dissatisfaction with the company (Kau & Loh, 2006). Bitner et al. (1990) found that most of the service failures in a interpersonal service encounter are due to employee behavior, while the recoveries of these failures have also been shown to be a result of employee behavior (Forbes, Kelley & Hoffman, 2005). In a technology-based service encounter, such as e-commerce, the human element of the encounter is erased. With a non-existing human element present, the recovery process has become far more difficult (Kelley and Davis, 1994).

A service failure occurs when the service provider is not able to deliver the service in the way the customer would expect (Bhandari & Polonsky, 2011). In order to retain the customers that experience this type of failure, an effective complaint handling has shown that you will be able to convert these oppressed customers to satisfied and loyal ones (Gilly & Hansen, 1992; Ndubisi, Malhotra & Miller, 2013). Spreng, Harrell & Mackoy (1995) also found that the overall service recovery is even more important than the original service failure that led to the recovery, and this illustrates how important the recovery process is for the company.

Moreover, service failures in e-commerce are inevitable (Wang, Wu, Lin & Wang, 2011). The costs of leaving a customer unsatisfied is greater in e-commerce compared to the original brick and mortar stores as it is on a virtual platform, with little direct interaction between the company and customers. Additionally, there are relatively small switching costs for a customer in e-commerce, compared to other industries in which chatbots are present. Indeed, a study by Forbes et al. (2005) confirms that online consumers are likely to consider other places to purchase, regardless of the online retailers' effort of retaining the business after the service failure.

Research supports the notion of a recovery paradox; saying that if a failure takes place, and the recovery measure is highly effective, there is an opportunity to acquire higher customer satisfaction than if the failure did not happen (Magnini, Ford, Honeycutt & Markowski, 2007). This highlights incentives for why a firm should emphasize an effective service recovery strategy and maintaining customer relations. Following this, there is a correlation between the strength of the service recovery initiative and spreading positive word of mouth (Magnini et al., 2007; Berry, 1995). Additionally, Hart et al. (1990, 149) states that any problem an employee has a chance to resolve is an opportunity to go on beyond work requirements to win a customer for life. Furthermore, in the service recovery process, the responsiveness and courtesy shown by the employee will have a positive impact on how the consumer evaluates your company (Komunda & Osarenkhoe, 2012). Research done by Sousa & Voss (2009) shows that in e-service, the intentions of customer loyalty are negatively affected by the service failure, but that the resolution has a positive effect on the loyalty.

2.3 Satisfaction constructs

Customer satisfaction is crucial for every company to survive, creating a long-term relationship with its customers and sustaining a profitable future, especially for companies working within e-commerce (Cyr, 2008). Here, the switching cost of changing to a competing brand or company is very limited for a consumer, and therefore it is crucial to keep them satisfied enough to stay on as a consumer (Tax et al., 1998). One important aspect to sustain these consumers and keep a high level of satisfaction is to deliver a seamless, efficient and justified recovery when a consumer has experienced a service failure. When a service failure happens, it has been proven that a recovery from this failure would lead to a higher level of satisfaction, especially if the outcome is positive (Andreassen, 2000; Sousa & Voss, 2009). Being satisfied with the complaint response would also lead to a higher repurchase intention from the already dissatisfied customer (Halstead & Page, 1992). In addition, since several studies have measured the outcome of the service recovery in achieving customer satisfaction, word of mouth and future repurchase intentions (Bitner et al., 1990; Smith et al, 1999; Dong et al., 2008), our study will mainly focus on the satisfaction with the recovery and the impact on satisfaction with the firm.

Extensive efforts have been made into researching the relationship between customer satisfaction and loyalty in the past. We already know that customer satisfaction affects loyalty in a non-linear manner (Bowen & Chen, 2001), and when satisfaction reaches a certain level, the loyalty will increase dramatically. Similarly, when satisfaction has a decline, the loyalty will decrease equally (Oliva et al., 1992; Bowen & Chen, 2001). Further, Maxham & Netemeyer (2002) proved that two important and distinct aspects of satisfaction (satisfaction with recovery and overall firm satisfaction) affected word of mouth (WOM) intent and purchase intent. This is an assumption we take with us in this research, which will not be tested specifically. The constructs of satisfaction, both with firm and with the recovery, in the present study have been adopted from Maxham & Netemeyer's (2002) study.

2.4 E-commerce as a marketplace

E-commerce, also known as electronic commerce, is the part of the internet where one can sell and buy goods or services. In this type of industry, chatbots are predicted to stand for almost 85% of customer service interactions by 2020 (Julia, 2018). The use of chatbots is required to meet customer expectations and provide excellent customer service to the customers whenever they need it. By meeting customer expectations, the companies manage to attract new customers and retain the existing ones, resulting in repeated business by the customers (Anderson & Srinivasan, 2003). The rapid development of e-commerce has resulted in the growth of excessive information which can be overwhelming for a customer (Vegesna, Jain & Porwal, 2018).

2.5 The great debate: Chatbots or Human chat agents?

2.5.1 Chatbots

AI based service agents interact with customers in a similar manner as humans do in a human-to-human chat encounter, but instead of having a human chat agent answering customer inquiries, there is a computer program that steers the communication (Wünderlich & Paluch, 2017). Froehle & Roth (2004) created a theoretical framework that defines the role of technology in the service encounter. Chatbots replace the human element of the service encounter completely, which is termed as *technology-generated customer contact* (TGCC) (Froehle & Roth, 2004). There are many potential benefits present with a TGGC and use of chatbots in the service encounter. For instance, customers can contact companies without having to verbally connect with an employee (Fuss, 2017). Also, potential benefits with the use of chatbots that were prominent in a 2018 survey made by Drift is that one can get 24-hour service (64 %), get instant answers (55%) and receive answers to simple questions (55%) (MyClever et al., 2018). Brandtzaeg & Følstad (2017) also document that the most important motivation for customers when using chatbots were productivity reasons. The speed, ease of use and convenience were the main reasons (Brandtzaeg & Følstad, 2017).

Coupled with the argument of instant gratification, 35% of the respondents to Drifts study anticipated chatbots to be used for resolving complaints (MyClever et al., 2018). Familiarity with the chat platform can also explain consumer motivations for using chatbots. Consumers are increasingly using social media, such as Facebook messenger and WhatsApp, to chat and stay in contact with their friends and family (MyClever et al., 2018). The extensive use of social media by consumers to seek help in the U.S, where millions of requests are sent on Twitter each month, is a clear evidence of the potential benefit of chatbots (Xu, Liu, Guo, Sinha &

Akkiraju, 2017). Customers expect their request to be solved within an hour, but in reality, it will take human operators on average 6,5 hours to respond (Xu et al, 2017). Accordingly, the motivations for businesses to incorporate chatbots include artificial intelligence replacing human service jobs (Huang & Rust, 2018). Further, the technology of chatbots is easily accessible for firms to implement (Wünderlich & Paluch, 2017). Previously, chatbots in e-commerce have been successfully used in roles of a shopping assistant (Shawar & Atwell, 2007). Chatbots as shopping assistants performed tasks such as giving users information regarding price and products. Hence, user expectations were met, customers felt it was easy to use and that the computer made their life easier (Shawar & Atwell, 2007).

Furthermore, implementing chatbots can be seen a cost reduction measure, as expensive human capital can be moved to other more important areas of the business (Xu et al, 2017). Accordingly, chatbots have close to zero incremental costs attached to usage (Wirtz et al., 2018). Thus, chatbots can be seen as a promising candidate to be an alternative to traditional customer service (Brandtzaeg & Følstad, 2017) and can have positive impacts on both the service provider and the customer. The arguments made for chatbots indicates that there is potential for a firm to become more customer centric and cost effective with implementing chatbots in service recovery. However, there is limited research on this topic, and potential benefits and consumers motivation for use needs to be investigated further.

2.5.2 Human chat agent

To begin with, we define human involvement in a service encounter as a customer directly interacting with human personnel, called a chat agent in this case (Bitner, 1990). As opposed to interacting with a chatbot, Froehle & Roth (2017) defines this interaction as *Technology mediated customer contact* (TMCC). TMCC illustrates when the employee and customer is not physically co-located during the encounter, but there is a human element present (Froehle & Roth, 2004). The chat agent interacts with the customer on an online platform, learns and understands the customer's inquiry and delivers the requested service to the customer (Bitner, 1990).

Literature regarding human interaction on a written platform indicates that employees which show positive emotions in a service encounter, correlate positively with a customer's evaluation of service quality (Pugh, 2001). Following this train of thought, service with a smile, or an employee going on and beyond for the customer can lead to a higher customer satisfaction. However, since it is on a written platform, service with a smile is perhaps not as relevant, but going on and beyond for the customer can be a critical aspect in generating customer satisfaction. Literature suggests that customized treatments such as friendliness is important to create long term loyalty, which displays the importance of the correlation between customer retention and profits (Reichheld, 1993). A human chat agent can pick up subtle linguistic cues and personalize a conversation for the individual customer (Fuss, 2017). Additionally, a loyal customer can create a ripple effect in the sense of positive word of mouth (Gremler & Brown, 1999), which stresses the importance of maintaining customer satisfaction.

2.5.3 Chatbots vs. Humans

Customer needs are not always satisfied with chatbots and frustrated customers can publicly concern their discontent with TGGC and chatbots (Wünderlich & Paluch, 2017). There are several challenges present when implementing chatbots in real-life interactions with consumers on an online platform, such as an overall scepticism and resistance to take it in use (Araujo, 2018). Potential blockers to the use of chatbots were identified to be consumers answering that they would prefer a real-life assistant (43%), worrying that the chatbot would make a mistake (30%) and if the chatbot wasn't able to respond in a friendly manner (24%) (MyClever et al., 2018).

Moreover, empathy can be considered to be a human skill, which we do not associate with machines. It can involve picking up subtle linguistic cues, moods and patterns. This has previously been identified as one of the reasons why customers prefer to speak to a human chat agent instead of a chatbot (MyClever et al., 2018). Chatbots will have to solve problems with some degree of intelligence, which includes cognitive abilities, social capabilities and affective sensitivity such as showing appropriate emotional responses (De Angeli, Johnson & Coventry, 2001). Chatbots carry the risk of being perceived as cold, socially restricted, untrustworthy and incompetent, which can result in great customer frustrations (Brave, S., & Nass, C., 2002; Feine, Morana, & Gnewuch, 2019). Furthermore, Wünderlich & Paluch (2017) propose that communication-related cues have an impact on customers perceptions of authenticity. This is exemplified in a response from their study: “I want to be treated well. Sometimes, you only get these standard answers.” (Wünderlich & Paluch, 2017). Additionally, the complexity of the human languages creates difficulties for chatbots, and it has been shown that people communicate with the chatbot for a longer time, but with shorter messages compared to a human chat agent (Hill, Ford & Farreras, 2015).

Concurrently, chatbots represent a new form of customer interaction which is designed to increase the quality of information given, with intention to increase user satisfaction (Wünderlich & Paluch, 2017). Furthermore, Araujos (2018) research suggest that when chatbots are given anthropomorphic cues, or human like abilities, a positive effect upon relationship building is expected. By imitating human behaviour, especially in a text-based platform, chatbots can almost be indistinguishable from humans (Wirtz et al., 2018). For instance, in Wünderlich & Paluch (2017) study, they found that 38% of their participants were uncertain whether they encountered a human or a chatbot, while 18% guessed wrong. Van Doorn et al. (2017) suggest that the more human like the automated service agent is, the more forgiving the customer might be when the agent causes a service failure. Duffy (2003) states that too strong anthropomorphic cues given to a service robot can lead to overly optimistic expectations among users which can lead to disappointment. Furthermore, chatbots can strengthen the relationship with customers and can revolutionize how companies stay in touch with its customers (Hyken, 2017). Chatbots will not get frustrated by challenging customers and they do not have bad days, which human chat agents might experience (Hyken, 2017). Hyken (2017) continue to explain that chatbots can for instance send out messages on customers birthdays to check up on them, and in some ways can deliver more human experiences than an actual human.

Following the previous section, chatbots with more human-like attributes, may help combat distrust some users have towards computer-based systems (Zamora, 2017). Further, customer service chatbots can be taught to detect subtleties and complexities of the human language (Wilson et al., 2017). For instance, Yahoo is developing algorithms for chatbots to be able to read between the lines, to detect and understand when customers are using sarcasm (Wilson et al., 2017). Furthermore, there are machine-learning systems which helps digital assistant, such as chatbots, to answer inquiries with sympathy (Wilson et al., 2017). It is worth mentioning that these systems are under development, and at present moment, human trainers are necessary to train the chatbots. Supervised learning of this kind, can result in chatbots that are more equipped to solve on-the-fly problems and reduce time spent on inquiries (Wilson et al., 2017).

It is apparent from previous research that both chatbots and human chat agents can result in satisfied customers in a recovery situation. However, we know of the uncanny valley effect, which involves customers tending to feel discomfort towards technology in a human-machine interaction (Ciechanowski et al., 2018). Simpler text chatbots were proven to induce less uncanny valley effects and less negative effects than a machine displaying an avatar (Ciechanowski et al., 2018). Similarly, Moon and Conlon (2002) established empirical support for a general decision-making bias in which they coined person-sensitivity bias. The authors made a direct comparison between humans and objects (industrial robots) in a good and a bad performance when performing the same task. The result from this study suggests that in a good experience, humans evaluate humans more positively than objects. Subsequently, when the experience is bad, objects are evaluated more positively than humans (Moon & Conlon, 2002). Thus, individuals get too much credit when things go well, and too much blame when things do not go so well (Moon & Conlon, 2002). With this in mind, we propose that a successful service recovery situation involving a chatbot will result in less satisfied customers, as opposed to an encounter with a human chat agent. However, a unsuccessful service recovery encounter with a chatbot will then result in less dissatisfaction towards the firm and the recovery, compared to the same situation performed by a human chat agent.

H1 *A successful service recovery attempt made by a chatbot will generate less customer satisfaction towards the firm and the recovery itself, as opposed to a successful recovery attempt made by a human chat agent.*

H2 *A unsuccessful service recovery attempt made by a chatbot will generate greater customer satisfaction towards the firm and the recovery itself, as opposed to a unsuccessful recovery attempt made by a human chat agent.*

2.5.4 Chatbots and human chat agents in symbiosis

Until now, most of our discussion has been to identify which of the two agents, chatbots or human chat agent, is the option that provides the greatest customer satisfaction. However, technology has affected the service encounter in a manner which makes it possible for humans and technology to interplay, in order to create a better service encounter (Froehle & Roth, 2004; Laviere et. al., 2017). As discussed, there are areas where humans outperform technology and vice versa, however when working together there is great potential to solve the most complex customer inquiries (Nadella, 2016). What is meant by this, is that many chatbots have the option to transfer the customer to a human chat agent. Chatbots can initiate the service encounter, and then handover the conversation to a human chat agent instead, which can reduce the number of routine inquiries managed by human service employees (Feine et al., 2019).

Jarrahi´s (2018) research investigates the complementarity of humans and AI, to see how the two can bring in their own strengths in decision making processes in regard to uncertainty, complexity and equivocality. For instance, AI can be an extension to human’s cognition when considering complexity, while humans provide a more holistic and intuitive way of dealing with uncertainty and equivocality in decision making in the organization (Jarrahi, 2018). An example is Garry Kasparov, who was the first world champion in chess to be beaten in a game of chess by a machine in 1997 by IBM Deep Blue (Collins, 2018). Kasparov then said: “If you can't beat them, join them” (Collins, 2018). The result was that chess players assisted by machines beat the singular machine, showing that a collaboration of the two was highly beneficial. Similarly, chatbots have the potential to replace humans completely in some sectors; however, these findings suggest that these systems should be designed to augment and not replace human contribution (Jarrahi, 2018). In other words, the chatbot will improve human efficiency, while humans will improve chatbot efficacy (Tripathy, 2018). Subsequently, chatbots are superior at collecting customer data from support interactions, which humans chat agents in turn can use to perform a more personalized response to a service failure (Hyken, 2017). This is evident in handoff situations, in which the chatbot is not able to answer the customer inquiry and is handed off to a human chat agent. Here, the human chat agent will start with an understanding of what the problem is and who the customer is.

However, a symbiotic relationship does not suggest that the result is beneficial for one or both. There are drawbacks with being transferred from a chatbot to a human chat agent, as potential benefits are removed when considering them individually. Most importantly, the aspect of instant gratification with answers from a chatbot right away disappears. Similarly, one does not receive service 24 hours of the day as one would with just operating with a chatbot. On the contrary, it is argued that it is with the employees the customers build a bond and develops trust with, which an important factor of maintaining customer loyalty (Reichfeld 1993). Aligned with this is the argument of empathy, which is to be identified as a human skill which chatbots do not possess in similar degree. We therefore expect that chatbots and humans working together will result in positive satisfaction in a successful encounter, but lower levels as opposed to evaluating them as separate entities. However, in an unsuccessful encounter we believe the literature indicates that the level of satisfaction would be greater in encounters where chatbots and humans work together compared to just encountering one of them.

H3 *A successful service recovery done by chatbot+human chat agent, will generate less satisfaction towards the firm and the recovery itself, as opposed to an encounter with just one of the agents.*

H4 *An unsuccessful service recovery done by chatbot+human chat agent, will generate more satisfaction towards the firm and the recovery itself, as opposed to an encounter with just one of the agents.*

2.6 Conceptual Framework

The framework for our study portrays the elements used to investigate the relationship between humans and chatbots compared to human-to-human relationship. As the relationship between constructs of this thesis framework have a limited establishment from previous literature and research, we have tried to come up with the most expected construct of the framework we believe will fit best for our research. More specifically, the framework illustrates how satisfaction with firm and recovery are affected depending on which encounter (chatbot, human chat agent or a chabot+human) the respondent is faced with and the outcome of service recovery (successful vs. unsuccessful). Our main interest is to study the constructs of service recovery and customer satisfaction in E-commerce. We will study each construct and analyse the encounters to see if similarities and differences between respondents' reactions are found.

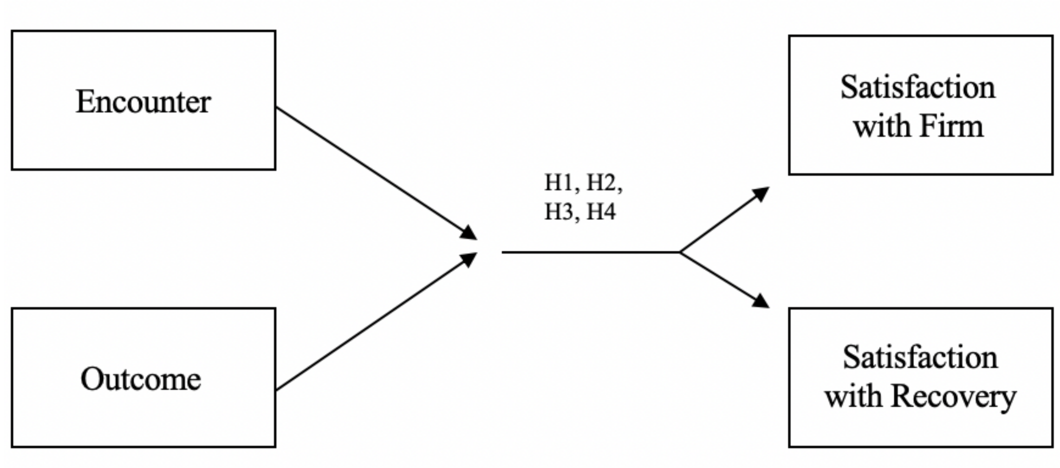


Figure 1: *Conceptual framework*

3.0 Methodology

Based on previous literature and our discussion earlier in this paper, we wanted to measure how customer satisfaction towards a company is affected due to a successful/unsuccessful service recovery attempt made by either a human chat agent, a chatbot or a chatbot and human working together. To examine the four aforementioned hypotheses, we developed a quantitative 3 (Chatbot, Human & Chatbot+Human) x 2 (Successful & Unsuccessful) between subject design. The experimental testing was carried out through a scenario-based survey experiment, where we developed six different scenarios (appendix 1), followed by survey questions (appendix 2) related to each scenario.

3.1 Overall study design

A scenario-based survey experiment was chosen due to the extensive expenses and ethical issues that follows a real-life setting and field studies (Kim & Jang, 2014).

Subsequently, a service recovery study from real-life or lab is hard to replicate due to all the activities that follow a service recovery attempt after a service failure (Smith & Bolton, 1998). The advantages of using scenarios are that the participants eliminate difficulties associated with observation of service failure/recovery incidents in the field of e-commerce (Smith et al., 1999). Likewise, experimental based scenarios avert inconvenient response biases due to memory lapse, consistency factors and rationalization tendencies (Smith & Bolton, 1998).

3.2 Pre-test

Before conducting the survey, a pre-test of the main questionnaire was organized in order to establish that the different scenarios worked as expected and were perceived as realistic. This was necessary as a measure to increase the ecological validity of our study. A small sample of 41 persons was recruited to conduct the pre-test. At the end of the questionnaire, the respondents had the opportunity to comment on improvements, if something was unclear or misunderstood in the survey or just provide general feedback. The feedback received was used to improve some of our questions, because a few respondents commented that the questions were too similar or unclear. Further, we received some comments about our scenarios, some strange wording and a few spelling mistakes we also had to rewrite. We also tested the realism of our scenarios using two items from Dabholkar's' (1996) study, measured on a seven-point Likert scale. The items were rewritten to match our survey and sounded like: “The scenario described in the beginning was realistic” and “I had no difficulty imagining myself in the situation”. Means of the realism checks rated 5.39 on a scale from 1 (Strongly disagree) to 7 (Strongly agree), showing that our six different scenarios seemed very realistic, and confirming the ecological validity of our study. Overall, with the pre-test we could conclude that our scenarios seemed realistic enough and we could go on conducting our main survey.

We also tested the internal consistency and reliability of our scales using Cronbach’s Alpha. Hair et al. (2016, 90) says that an acceptable level of Cronbach's Alpha should not be lower than limit of 0.6 to 0.7 In the table below, we can see both of our scales indicated a very high level of reliability, and therefore we could conclude that the scales worked as expected.

Table 1: Cronbach's Alpha Pre-test

Scale	Cronbach´s Alpha	N of items
Satisfaction with recovery	0.960	3
Satisfaction with firm	0.955	3

3.3 Sample

The data collection of our main survey entailed 289 respondents, but after removing incomplete answers and extreme outliers, we were left with 203 respondents. The majority of our respondents were Norwegian citizens, but as we posted this on social media (Facebook) we had limited of control of everyone conducting the survey. Due to the "evenly present" function in Qualtrics, we managed to get an approximately equal amount of respondents for each of our six scenarios. As shown in table 2, there were at least 30 respondents for each of the scenarios. In addition, as all questions in our survey had compulsory responses, there were no missing values from the 203 respondents.

Table 2: *Distribution of respondents*

	Successful	Unsuccessful
Chatbot	30	39
Human	33	37
Chatbot + Human	32	32

98 (48.3%) of the respondents were female and 104 (51.2%) were male, with one respondent choosing the option "Other". 79.3 % of the respondents had interacted with a chatbot before, 12.8 % had never interacted with a chatbot and the last 7.9% did not know if they had interacted with a chatbot before. This shows that many have actually used chatbots today, but still, 58.1% of our respondents prefer to speak with a human or employee instead of an automated service (16.3%) and as much as 25.6% do not have any preferences. The age of the respondents ranged from below 20, to 69 years of age. The age group of 20-29 counts for the biggest part of the respondents, with 174 people responding to the questionnaire selecting this group (85.7%). This was followed by 9 (4.4%) people in the range 30-39, 5

(2.5%) people in the range 40-49, 3 (1.4%) people in the range 50-59 and 11 (5,3 %) people between 60-69, and one person under the age of 20 (0.5%). Furthermore, the majority of the respondents have a bachelor degree as their highest level of education completed (50.2 %), 32% have a master degree, 14.8% have a high school degree, 1% have finished elementary school, 0.5% have a doctorate as highest level and 1.5% chose the option other. Most of the respondents were either full-time students (39.9%) or full-time employed (53.2%). We further investigated the income of the respondents, and 31% had an income of 200.000 Norwegian krone (NOK) or less, 34% had an income between 200.000-499.999 NOK, 25.1% of the respondents earned between 500.000-799.999 NOK and the last 9.8% earned more than 800.000 NOK. Lastly, we also asked the respondents about their marital status, where 41.4% of the respondents are single, 42.8% are in a relationship, 14.3% are married or in a domestic relationship, and the last 1.5% are either divorced or widowed.

3.4 Design

Table 2 below represents the 3x2 between subject design of our study. Respondents were randomly assigned into six scenarios depending on if they were faced with a human chat agent or a chatbot and if the encounter was successful or not. The six different scenarios are found in appendix 1. Additionally, the last two groups include the ones who had an unsuccessful encounter with a chatbot, but is transferred to a human where one will have a successful or unsuccessful encounter. The design presented in table 2 requires respondents for six different treatment groups, each subjected to different scenarios. All the scenarios and questions were presented in English in order to not confuse or create any confounds regarding translation into Norwegian.

Table 3 - *3x2 between subject design*

		Outcome of encounter	
Encounter	Human	Successful	Unsuccessful
	Chatbot		
	Chatbot + Human		

The contextual setting of our scenarios is a complaint situation on the e-commerce platform of a fictional brand called Beds & Pillows Inc. in Norway. A fictional brand was used due to customer´s tendency to be more forgiving of mistakes in positive pre-existing relationships (Wan & Chan, 2019; Jones, Mothersbaugh & Beatty, 2000). Furthermore, with a fictional brand, there would be no biases regarding the company and its offerings, and the name of the company is revealing to what they actually sell. The clothes and textile industry were chosen due to lower barriers of substituting brands compared to the banking industry where chatbots are also used extensively. In addition, the general population of Norway is sizable and disperse, thus the scenarios need to provide a context that is relatable for the majority of the population. Furthermore, the respondents would be introduced to one of the most common failures that occur in an e-service, which is process failures (Holloway & Beatty, 2003). To recover from this failure, we will use the recovery strategy of "addressing the problem and correcting it", proposed by Forbes et al. (2005) which is considered to be the most desirable and results in the most favorable levels of satisfaction.

When designing the scenarios in which the respondent is faced with a chatbot, there are several things to consider. Firstly, one need to consider how intelligent the chatbot should be. For instance, the Turing test determines if a computer can exhibit intelligent behavior indistinguishable of that of a human (Radziwill & Benton, 2017). Even though the test has received a lot of criticism, the test is yet to be passed by any robot (Todorovic, 2015). Hence, it is important that the respondents will be able to distinguish and recognize that they are talking to a chatbot, because it is not believable that chatbot responses are indistinguishable to that of a human, yet. Thus, it is clearly stated who the respondent is talking with, indicated with a name tag above the chat, as well as an introduction in the form of a greeting.

Furthermore, Saarem´s (2016) article discusses how humans perceive computers that talk, and has used this knowledge to create a guideline for chatbot design. These guidelines were considered when designing our scenarios. Firstly, to help minimize the chance that there is a discrepancy between user expectation and the chatbot´s capabilities, the chatbot should be upfront about what it can and cannot do. Secondly, the chatbot used for our research will not have a gender specific name, as people respond differently depending on what gender they are faced with

(Saarsem, 2016). This logic is also behind the reasoning of not having a picture of a human or a chatbot next to the conversation. Also, the scenarios are worded as similar as possible in order to maintain consistency.

3.5 Procedure

To collect the data, we used an internet-based survey, created in Qualtrics. The participants were recruited through convenience sampling on Facebook, and participation was voluntary.

Firstly, each participant was randomly assigned to one of the six scenarios, that either consisted of a successful or unsuccessful recovery attempt made by a chatbot or a human. The service recovery strategy offered by both the human and the chatbot in the successful scenarios was "Correction" of the complaint/problem, as this is the most commonly used strategy in a service recovery situation (Forbes et al., 2005). After reading the scenario the participants were asked to answer a survey, in order to see how they responded either to a chatbot or a human in the recovery situation and questions related to their satisfaction with the recovery. To this end they also answered questions regarding satisfaction with the company, in order to measure differences between the chatbot and the human agent.

The questions we used to gather information regarding the service failure, recovery and satisfaction or dissatisfaction with the recovery and the organization/firm are included in appendix 2. Each of the participants answered/were told: (1) to read scenario they received, (2) provide his/her satisfaction with the recovery, (3) to rate the satisfaction of the company regarding the recovery, and (4) answer the demographic questions related to our study.

3.6 Instruments

The survey questions were all adapted from earlier research and validated scales. The seven-point Likert scale, ranging from 1 = very dissatisfied to 7 = very satisfied, was used to measure the satisfaction with recovery and satisfaction with firm, adapted from Johnson et al. (2001) and Maxham & Netemeyer (2002), as shown in table 3. Some of the questions were also modified to fit our survey questions and purpose of the study, which can be seen in appendix 2, showing the whole survey. Satisfaction with the recovery and satisfaction with firm were also the two dependent variables of our study. As already mentioned in the pre-test, the reliability of the scales was checked with the Cronbach's Alpha coefficients and this was also done in regard to the main survey, presented in the result section.

Table 4: *Measurement Items*

Measurement Items	
Construct: Satisfaction with recovery	
Johnson et al., 2001; Maxham & Netemeyer, 2002	1. How satisfied are you with the resolution to your problem? 2. To what extent does the resolution of your problem meet your expectations? 3. If you imagine the ideal resolution to this problem. What is your opinion about the resolution?
Construct: Satisfaction with firm	
Johnson et al., 2001; Maxham & Netemeyer, 2002	1. How satisfied are you with Beds & Pillows Inc? 2. To what extent does Beds & Pillows meet your expectations? 3. If you imagine the ideal customer service for a home essential shop. What is your opinion on Beds & Pillows Inc services?

3.7 Reliability and validity

3.7.1 Reliability

The assessment of the degree of consistency between multiple measurements of a variable is defined as reliability (Hair et al., 2014, 123). Similar to the pre-test, the Cronbach's Alpha levels are examined to confirm the internal consistency. When the reliability of our constructs is confirmed, the proceeding validity checks are performed in order to have obtained consistency of our study in the first place.

The Cronbach's Alpha levels allows us to conclude that both scales of our study achieve a high level of reliability., Satisfaction with recovery (α = .907) and Satisfaction with firm (α = .925) coefficient both scores above 0.7, as shown in table 5 (Hair et al. 2016, 90). Thus, our scales measure what they are intended to measure: satisfaction with recovery and satisfaction with firm. Therefore, our two dependent constructs were created using three items, taking the mean scores of all the items in the construct. The reliability of our constructs is established, and the different types of validity are further assessed in the next section.

Table 5: *Cronbach´s Alpha*

Scale	Cronbach´s Alpha	N of items
Satisfaction with recovery	0.907	3
Satisfaction with firm	0.925	3

3.7.2 Validity

Validity is defined as the degree to which a measure accurately represents what is supposed to or correctly represent the concept of study (Hair, Black, Babin, & Anderson, 2014, 7). The construct validity represents in what extent the items or constructs of the study measures what they are supposed to measure. Constructs validity consist of three different subtypes of validity which are convergent, discriminant and nomological validity. We find it sufficient to assess the convergent and discriminant validity in this study, because the relationship of our model/constructs have never been measured before to our knowledge and we have no theoretical background to check for nomological validity of our study. In that case, convergent validity assesses the degree to which measures of the same constructs are correlated (Hair et al., 2014, 124). High correlation indicates that items are measuring their intended constructs, and this is important to strengthen the relationship of the items of the construct in this study. Further, we have discriminant validity which assesses that the conceptually similar concepts are distinct from each other (Hair, et al., 2014, 124). In this case the correlation should be low between the different constructs, in order to be different from each other. There are three items related to each of our dependent variables, which are a part of measuring the construct of overall satisfaction (Janssens et al., 2008).

A factor analysis was performed in order to control if the items loaded correctly on each construct, as well as checking the different types of validity of the survey. All of the six items had factor scores loading on one component, with only one factor with eigenvalue > 1 and explaining almost 76.8% of the variance in this component, as shown in appendix 3. Hence, the construct validity of our study is threatened, which makes it more difficult to measure the intended convergent and discriminant validity. In this case we have provided a more in-depth analysis of the construct validity, where we have checked two of its subtypes.

First of all, reliability can also be an indicator of convergent validity as we confirmed in the previous section (4.1.1) (Hair et al., 2014, 619). All the items of the two constructs indicate a high convergent reliability from table 10, with relatively high Cronbach Alpha scores. To determine the discriminant validity of our constructs, we do not want the constructs to be correlated. This was already confirmed when we checked for multicollinearity of our dataset using Pearson Correlation, and therefore we can confirm a good discriminant validity of our constructs as we did not have any problems with a high correlation.

Content validity is the assessment of the correspondence of the items included in a construct or summated scale (Hair et al, 2014, 123), which can be done through pre-tests with multiple subpopulations or expert judges. In this case it is important that the items of our study represent the characteristics of the constructs and are not to be considered as irrelevant to the constructs, in order to prevent any possible biases to arise. From the previous section (3.6), we showed that the items representing our dependent constructs are gathered from preceding theories and studies.

By randomizing the different scenarios in our study, we could sustain the internal validity of the study. Likewise, by addressing potential confounds such as controlling for age, gender and incorporating appropriate anthropomorphic cues internal validity of the study is maintained. Further, the experimental setting of our study does increase the internal validity of our research, because we have the opportunity to control the setting and make it "clean". The external validity of our study is set to generalize beyond specific settings of our study (Hair et al., 2014, 268). In other words, the results of our study are representative of the larger population, and not just the smaller sample group used in this study.

3.8 Assumption Checking

3.8.1 Observations and dependent measurement

Prior to running a MANOVA, there are several assumptions to consider. Firstly, the observations need to be randomly and independently chosen from the population.

This assumption is met as our questionnaire was distributed on social media, the respondents could voluntarily participate in the experiment, at the time most convenient to them.

Further, there is an assumption stating that all dependent variables need to be either continuous or scales variables, and all independent variables need to be categorical. This assumption is met as all items representing the dependent variables; Satisfaction with Recovery and Satisfaction with Firm, were measured on a 7-point Likert scale. Both the independent variables of Outcome and Encounter, represented one category in each level of the variables.

3.8.2 Normality

Multivariate normality is also one of the assumptions that needs to be checked before running the MANOVA. Firstly, skewness is a measure of asymmetry and normality of the dataset (NIST, 2013), and items with values ranging from -1 to +1 are to be identified as a substantially skewed distribution (Hair et al, 2010, 36). We can see from table 6 that there are no items that lay outside the interval of -1 to +1, as well as all items being of the positive kind. This means our data is moderately skewed to the right. Similarly, kurtosis helps explain how the distribution of our dataset is compared to the normal distribution. All of our items have a negative kurtosis, which indicates a light-tailed distribution or, meaning a flatter distribution in comparison to the normal distribution (NIST, 2012). As SPSS does not provide a way to test the multivariate normal distribution, a sufficiently large sample size is enough, (20 elements for each of the independent variables measuring the dependent ones) for the Multivariate Central Limit Theorem to hold (Statistics Solutions, 2013), and we can assume that the multivariate normality assumption holds as well. This can also be checked statistically by determining the univariate normality (Janssens et al., 2008, 113). This is done by observing the Kolmogorov-Smirnov and Shapiro-Wilk statistics, shown in table 7. Shapiro-Wilk statistics is more commonly used when you have smaller dataset (less than 2000 elements), and therefore this is interpreted in the table below.

Table 6: Descriptive statistics of dependent variables

			Kurtosis		Skewness	
	Mean	**Std.Deviation**	**Statistic**	**Std.Error**	**Statistic**	**Std.Error**
Satisfaction_Recovery	3.447	1.748	-.943	.340	.221	.171
Satisfaction_ Firm	3.090	1.405	-.471	.340	.363	.171

From the results in table 7, it is apparent that the null hypothesis of normality can not be rejected in several combinations of our two dependent variables. In 3 out of the 10 different groups the null hypothesis can not be rejected ($p > .05$). This means that the data of these groups is normally distributed. For the other 7 groups the null hypothesis can be rejected, indicating a non-normal distribution of our sample. When testing for larger sample sizes ($n > 200$), both of the test statistics are extremely sensitive to minor deviations from normality. Thus, a rejection from the null hypothesis is no implication that the deviation is big enough to motivate a distortion of the statistical analysis (Janssens et al., 2008, 113). That said, a graphical inspection of the normality is performed in addition to the formal testing, since our sample size consisted of 203 respondents, and should apply as an assumption for a normally distributed sample group.

Table 7: *Test of Normality*

	Outcome	Encounter	Kolmogorov-Smirnov Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
Satisfaction_ Recovery	Unsuccessful		.154	108	.000	.910	108	.000
	Successful		.078	95	.184	.970	95	.028
		Human	.109	70	.039	.950	70	.007
		Chatbot	.166	69	.000	.880	69	.000
		Chatbot+Human	.098	64	.200	.971	64	.133
Satisfaction_ Firm	Unsuccessful		.119	108	.001	.921	108	.000
	Successful		.105	95	.011	.983	95	.240
		Human	.101	70	.075	.956	70	.015
		Chatbot	.138	69	.002	.946	69	.005
		Chatbot+Human	.113	64	.042	.963	64	.052

Outliers were identified by examining the box plot and the whole data set. As outliers were removed in the beginning of the analysis, these outliers were not identified as extreme outliers anymore. By investigating the histograms and the Q-Q plots of the different groups related to our dependent variables (Appendix 3), we can conclude with a distribution which is close to a normal distribution. Some of the histograms did not show a perfect bell curve, but from the Q-Q plots the points were close to the diagonal line, which suggests a normal distribution. Therefore, we can with some violations to the normal distribution in the formal testing, conclude with not having a big impact on the MANOVA analysis of this study.

3.8.3 Equality of covariance matrices

Table 8: *Box´s Test of Equality of Covariances Matrices*

Box´s M	F	Df1	Df2	Sig.
17.923	1.164	15	194243.120	.292

The Box´s test verifies the assumption of homogeneity of covariance, which also assumes multivariate normality. From table 8 we confirmed that the null hypothesis from the test can not be rejected ($p > .001$). Thus, the covariances matrices between our variables are equivalent. What is more, we have almost equal group sizes between the six different groups from table 2. In addition, Levene's test show that for each of the dependent variables, the null hypothesis can not be rejected, with a not significant result ($p > .05$), indicating that there exist equal group variances. This further supports the assumption of equal covariance matrices across the groups. Interpretation and analysis can therefore continue further, as also this assumption for MANOVA is met.

Table 9: *Levene's Test of Equality of Error Variances*

	Levene Statistic	Df1	Df2	Sig.
Satisfaction_Recovery	1.476	5	197	.199
Satisfaction_Firm	1.392	5	197	.229

3.8.4 Multicollinearity

The last assumption to check before running the factorial MANOVA, is that the two dependent variables cannot be too correlated. We used Pearson Correlation to check whether or not the two dependent variables correlated to each other too much. From table 9 we can see that the correlation value of $r = .790$ indicates that we do not need to be worried about multicollinearity, as long as the value is less than $r < .90$, according to Hair et al. (2014, 196). Significant correlation among our dependent variables can also be checked using *Bartlett's test of sphericity* as done in the factor analysis (Appendix 3). Here, we can also confirm that a significant degree of intercorrelation exist ($p < .001$) (Hair et al., 2014).

Table 10: *Pearson Correlation*

		Satisfaction_Recovery	Satisfaction_Firm
Satisfaction_Recovery	Pearson Correlation	1	.790
	Sig. (2-tailed)		.000
	N	203	203
Satisfaction_Firm	Pearson Correlation	.790	1
	Sig. (2-tailed)	.000	
	N	203	203

3.9 Analysis

To interpret and analyse the data we have gathered with the subscription software Qualtrics, we have used SPSS statistics. This is a software package used to provide statistical analysis. With this software we have among other things been able to present our sample data, descriptive statistics, test the validity and reliability of our data by doing a factor analysis and checking for Cronbach's Alpha coefficients. Further, as we have two dependent variables in our conceptual framework, a MANOVA (multivariate analysis of variance) analysis of our data will be used in order to test our hypothesis and define the relationship between the different groups in our independent variables (Encounter and Outcome), and how they affect the dependent variables of satisfaction with firm and recovery.

4.0 Results

4.1 Descriptive statistics

Table 11 depicts the descriptive statistics of the six different scenarios regarding each of the two dependent variables. The encounter that has the greatest Satisfaction with recovery is a successful encounter with a chatbot (M = 5.333, SD = 1.138), followed by a successful human encounter (M = 4.778, SD = 1.611) and successful human+chatbot encounter (M = 4.104, SD = 1.611). A successful encounter with a chatbot (M = 4.400, SD = 1.236) also results in the greatest score for Satisfaction with firm compared to a successful encounter with a human chat agent (M = 3.919, SD = 1.367). The least satisfied group of Satisfaction with recovery in an unsuccessful encounter is those interacting with a chatbot (M = 1.786, SD = 1.025), followed by Human (M = 2.496, SD = 1.113) and Chatbot+Human (M = 2.771, SD = 1.162). When looking at satisfaction with firm, the order of the least satisfied group is the same as for with satisfaction with recovery. We will dig deeper into the different effects of the variables and how the relationship between these are related or not related to each other, in the analysis below.

Table 11: *Descriptive statistics*

	Outcome	Encounter	Mean	Std. Deviation	N
Satisfaction_Recovery	Successful	Chatbot	5.333	1.138	30
		Human	4.778	1.611	33
		Chatbot + Human	4.104	1.066	32
	Unsuccessful	Chatbot	1.786	1.025	39
		Human	2.496	1.113	37
		Chatbot + Human	2.771	1.162	32
		Total	3.447	1.748	203
Satisfaction_Firm	Successful	Chatbot	4.400	1.236	30
		Human	3.919	1.367	33
		Chatbot + Human	3.458	1.057	32
	Unsuccessful	Chatbot	2.120	.843	39
		Human	2.387	1.070	37
		Chatbot + Human	2.635	1.293	32
		Total	3.090	1.405	203

4.2 Multivariate analysis of variance (MANOVA)

As our study involves two independent and two dependent variables, a factorial multivariate analysis of variance (MANOVA) is an appropriate approach to take. Our independent variables are labelled *Outcome* and *Encounter*, which describes if it was successful or unsuccessful and which of the Chatbot, Human or Chatbot+Human encounters the respondents were faced with. The mean differences are based on our dependent variables which measure satisfaction, named *Satisfaction with firm* and *Satisfaction with recovery.*

The assumptions of performing a MANOVA were met above in the methodology section and to examine if there are any group differences between the variable combinations, this is done by the multivariate test presented in table 12. According to Jansen's (2008, p. 118) the Wilk's Lambda is the one normally operated with, which is then adapted to check for any significant effect of our independent variables on the dependent ones from the multivariate test. From table 12, we can firstly see that there was a statistically significant difference of satisfaction with recovery and firm based on outcome of the scenarios, ($F(2,196) = .552$, $p < .001$; Wilks $\Lambda = .496$, partial $\eta^2 = .504$). Further, the interaction between Outcome and Encounter is statistically significant ($F(4,392) = 6.878$, $p < .001$; Wilks $\Lambda = .873$, partial $\eta^2 = .066$), and therefore the null hypothesis that there do not exist any differences between the groups is rejected. For the Encounter variable, there is no statistically significant difference between the different encounters ($F(4,392) = .552$, $p > .05$; Wilks $\Lambda = .989$, partial $\eta^2 = .006$). Since the interaction term is significant, the effect of the Outcome variable on the dependent variable is not the same between the different Encounters. We have some significant effects, thus an univariate ANOVA should be performed on each dependent variable. In SPSS this is automatically done when running a MANOVA, and the output is presented in appendix 5.

Table 12: *Multivariate Tests*

Multivariate Tests

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Wilks' Lambda	.090	995.535	2.000	196.000	.000	.910
Outcome	Wilks' Lambda	.496	99.753	2.000	196.000	.000	.504
Encounter	Wilks' Lambda	.989	.552	4.000	392.000	.698	.006
Outcome * Encounter	Wilks' Lambda	.873	6.878	4.000	392.000	.000	.066

The test of between subjects' effects in appendix 5 shows that there are significant group differences for both dependent variables satisfaction with recovery and satisfaction with firm, with regards to the Outcome variable and interaction effect between the two independent variables. To account for the Type 1 error to occur when performing multiple ANOVAs, an alpha correction of the critical p-value (.05) is done (Janssens et al., 2008). In this case we can accept a statistically significance at $p < .025$ (=.05/2). Further, we find a significant main effect of our Outcome variable on both dependent variables; Satisfaction with Recovery ($F(1, 197) = 200.337$, $p < .025$; partial $\eta^2 = .504$) and Satisfaction with Firm ($F(1, 197) = 91.465$, $p < .025$; partial $\eta^2 = .317$). The main effect of the different encounters individually, was not statistically significant for Satisfaction with Recovery ($F (2, 197) = .467, p > .025$; partial $\eta^2 = .005$) or Satisfaction with firm ($F (2, 197) = .569, p > .025$; partial $\eta^2 = .006$). Therefore, we have no main effect of the Encounter variable on the dependent variables. From the multivariate test we have in addition got an interaction effect between our two independent variables, with a statistically significant effect on Satisfaction with recovery ($F (2, 197) = 14.227, p < .025$; partial $\eta^2 = .126$) and Satisfaction with Firm ($F (2, 197) = 6.663, p < .025$; partial $\eta^2 = .063$). In this case we know that there exist group differences when the two independent variables are dependent upon the two satisfaction variables. By applying these group differences on the interaction effect, we are able to answer our hypothesis in the next section. Since the independent variables only consist of two groups, we are not able to do a Post Hoc test and test the significant effect between the various groups. In this case, we will use the differences in means to answer the hypothesis.

Our hypothesis are summarized in table 13 below, and further in this paragraph, the results will be presented in greater detail. We confirmed from appendix 5, a significant interaction effect between the Outcome and Encounter variables on the two dependent variables, Satisfaction with Recovery ($F (2, 197) = 14.227, p < .025$; partial $\eta^2 = .126$) and Satisfaction with Firm ($F (2, 197) = 6.663, p < .025$; partial $\eta^2 = .063$). In this case, when a successful recovery is done individually by a chatbot ($M = 5.333$, $SD = 1.138$) and a human ($M= 4.778$, $SD = 1.611$), the chatbot has generated a higher satisfaction with the recovery than a human encounter. Similar results are achieved with the satisfaction of the firm, indicating a higher satisfaction for a chatbot ($M= 4.400$, $SD = 1.236$) than a human ($M= 3.919$, $SD = 1.367$). We do in this case not find support for **H1.** We do not find support for **H2** either**,** when we find that a chatbot ($M = 1.786$, $SD = 1.025$) does not generate a greater satisfaction with recovery than a human ($M = 2.496$, $SD = 1.113$) in an unsuccessful attempt done by both encounters. This means that people are more unsatisfied with the chatbot when it fails, than a human. With the satisfaction of a firm, almost the same results are carried out with a chatbot ($M = 2.120$, $SD = .843$) having less satisfied respondents than a human ($M = 2.387$, $SD = 1.070$) Next, the attempt of a chatbot and a human ($M = 4.104$, $SD = 1.066$) working together generate less satisfaction with the recovery than a recovery done individually by a human ($M = 4.778$, $SD = 1.611$) and a chatbot ($M = 5.333$, $SD = 1.138$) in an successful recovery. Less satisfaction for the firm was also generated for a chatbot and human (M= 3.458, SD = 1.057) working together, than for a chatbot (M= 4.400, SD = 1.236) and human (M= 3.919, SD = 1.367) individually. Thus, we find support for **H3.** Lastly, we find that human and chatbot together (M = 2.771, SD = 1.162) generate a higher satisfaction with the recovery than the Human (M = 2.496, SD = 1.113) and the chabot (M = 1.786, SD = 1.025) in an unsuccessful recovery. Firm satisfaction is again higher for a chatbot and a human (M = 2.635, SD = 1.293) working together, than a chatbot (M = 2.120, SD = .843) and human (M = 2.387, SD = 1.070) working individually, and we do find support for **H4**.

Table 13: *Summary of hypothesis*

Hypothesis	**Encounter**		**Outcome**	**Mean Difference (I-J)**		**Supported/Not Supported**
				Satisfaction with Recovery	Satisfaction with firm	
H1	Chatbot	Human	Successful	.556	.481	Not Supported
H2	Chatbot	Human	Unsuccessful	-.709	-.268	Not Supported
H3	Chatbot+Human	Chatbot	Successful	-1.229	-.942	Supported
		Human	Successful	-.674	-.461	
H4	Chatbot+Human	Chatbot	Unsuccessful	.985	.516	Supported
		Human	Unsuccessful	.275	.248	

5.0 Discussion

The technological advances we are facing are truly making their imprint on the service encounter at present moment. Chatbots are a part of this shift, and contribute to the change in which customers and companies are interacting with each other. Building on previous research regarding service recovery, effective complaint handling is able to convert dissatisfied customers to satisfied ones (Gilly & Hansen, 1992; Ndubisi, Malhotra & Miller, 2013). What is more, to our knowledge there is limited research made on chatbots in service recovery situations and its effect upon customer satisfaction. However, what is obvious from our research background section, is that previous research indicates that both chatbots and human chat agents have qualities which could lead to satisfied customers in complaint situations. Hence, it was of utmost importance to research this in greater detail. The main purpose of our master thesis was to investigate how customers respond to a successful or unsuccessful complaint situation when faced with chatbots and human chat agents. We measured satisfaction towards the firm and the recovery itself in order to see how customers reacted in situations which were in various degrees facilitated by technology. A service encounter with a fictional e-commerce company was chosen, as compared to other industries in which chatbots are deployed, switching intention is higher. Four hypotheses were constructed, in which two were supported and two were not supported.

Our first two hypotheses, H1 and H2, look at the constructs of satisfaction towards the firm and satisfaction with the recovery, and let us make a direct comparison in the differences in customers evaluation of a chatbot and a human chat agent in a service recovery situation. H1 and H2 were not supported, which demonstrates that similar responses made by a chatbot or a human chat agent in a complaint situation resulted in different evaluation by the responders. In other words, in a successful service recovery situation, customers evaluate chatbots more favourably than human chat agents in relation to both satisfaction towards the firm and the recovery itself. Meanwhile, in unfavourable situations, chatbots will lead to less satisfaction amongst customers compared to human chat agents. Hence, these results contradict those of Moon & Conlon (2002), in which it was humans that were evaluated more extreme compared to machines. By extreme, Moon & Conlon mean that humans were evaluated more positively in favorable situations, and more negatively in unfavorable situations.

The literature that laid the groundwork for our study, presented in section 2, provided arguments that were in favor of the use of chatbots in complaint situations. These can help explain why our results challenge those of Moon & Conlon (2002). Firstly, technology has gotten a more prominent role in today's society, which is highlighted by the terminology coined by Larivière et al. (2017): service encounter 2.0. Technology is changing the way customers interact with organisations, which again is altering the roles of the involved actors (Larivière et al. 2017). Accompanied with this is the fact that customers are increasingly using technologies, such as social media, to stay in touch with their friends and families

(MyClever et al., 2018). In addition, chatbots are accessible on mobile devices round the clock (Zamora, 2017). Consequently, we can assume from the literature that customers thoughts and expectations towards technology have changed dramatically since 2002 as well. This is exemplified through the fact that consumers are purchasing more and more goods and services on the internet, and is used to having little physical interaction. Further, chatbots have a proven track record of meeting user expectations earlier, as a shopping assistant. Customers considered the chatbots easy to use and evaluated they make their life easier (Shawar & Atwell, 2007), which yet again can help explain why our results differ from those of Moon & Conlon.

Furthermore, anthropomorphism is a keyword that needs to be brought into this discussion. As mentioned earlier, a positive effect on relationship building is expected when chatbots are given human-like abilities (Araujo, 2018). Moon & Conlon (2002) states that one of the underpinnings thoughts of their research is that decision makers attribute more similarity to themselves with another person, than they do to an object. However, chatbots mimic human behavior. As Wirtz et al. (2018) stated, chatbots can be almost indistinguishable from humans on a text-based platform. Compared to the objects used in Moon & Conlon, the difference between a human and chatbot in our study is minimal. This is in alignment with Van Doorn (2017) research, which suggest that customers are more forgiving during service failures the more human like the agent is. What is interesting is that prior research argued that chatbots can be perceived to be socially restricted and cold (Brave & Nass, 2002; Feine et al., 2019). However, researchers are working on how humans can teach them skills which are more relatable to humans, as for instance empathy (Wilson et al., 2017). Hence, one can assume that the reduction of differences between a chatbot and human chat agent, reduces customers discomfort towards technology. Hence, there is no uncanny valley effect for chatbots in service recovery situations, as Ciechanowski et al. (2018) would have put it.

In addition, there are several cases in which new technologies fail to be adopted by users due to the fact that there is uncertainty regarding its ability to provide more value than already existing initiatives. Indeed, our study show that chatbots have the ability to provide more than enough value for its customers, when they work.

However, when they do not work, satisfaction is heavily affected. H3 and H4 were included in order to see if some element of human interaction included in the encounter would influence customer satisfaction toward the firm and with the recovery. We tested when chatbots and humans worked together interchangeably, where the chatbot handed off the conversation to a human chat agent when it was not able to answer the inquiry itself. The results establish that this type of service recovery attempt lie in the intersection between chatbots and human chat agent as separate entities in relation to our satisfaction constructs.

Hence, our findings indicate that the human touch is still an important factor for a firm to maintain satisfied customers in the service industry. The human touch is not the factor that has the potential to result in the highest level of satisfaction, but is a measure to potentially not have the least satisfied customers. From this we can conclude with that combining the two agents to work together, works as a risk reducing measurement compared to using chatbots alone in service recovery. We assume that this stems from the fact that the conversation is handed off when the chatbot fails to solve the problem, and the problem is handed off to the human chat agent which in turn solves it. Here, we believe that the human chat agent stabilizes the situation by using the information already gathered by the chatbot, and solves the problem quickly. Along these lines, Jarrahi (2018) would say that chatbots acts as an extension of the human chat agent, as it provides data in which the human chat agent can take in use to solve the problem in a more holistic manner. This is consistent with the fact that customized treatments in which humans' agents can perform is important to create long term loyalty (Reichheld, 1993). Similarly, we know that communication-related cues impact customers perception of authenticity (Wünderlich & Paluch, 2017), and perhaps the contrast between a chatbot failing and the human chat agent succeeding creates this risk reducing effect.

In conclusion, the present study has made it apparent that the technological advancements in the service industry have been immense, but that the research regarding it has not followed the same pace. This is reflected by the fact that H1 and H2 were rejected. However, despite this, we have been able to answer our research questions in respect to how satisfaction of a firm and the recovery has been affected by our three encounters. Further, we have highlighted how the different encounters differ from one another. Our thesis shows that chatbots can be a valuable and effective way of managing customer complaints when the technology is working properly. However, in situations where chatbots do not function properly, these can instead drain a company's reputation to a greater extent than a human chat agent.

In the following section detailed thoughts regarding the implications of our findings will be provided for future decision makers.

6.0 Managerial and theoretical implications

Our research is of significant relevance for future managers, as a more in depth understanding of how implementing chatbots in complaint situations is valuable for managers to be aware of in relation to customer satisfaction. Further, our thesis contributes with new and supporting material to the theoretical perspective of service recovery, and chatbots as a tool in the service industry.

Firstly, the findings from our study indicate that using chatbots for service recovery purposes at the present moment is a double-edged sword. When the encounter is successful, chatbots generated the highest levels of satisfaction, which can make it even more tempting for managers to implement it in their business. However, in unsuccessful encounters, chatbots result in the lowest level of satisfaction, compared to human chat agents or a combination of them. As we know, low customer satisfaction is related to constructs like negative word of mouth and switching behavior, all of which have costs related to it. Coupled with this, since the main motivations for managers to implement chatbot technologies are to reduce costs and reallocate resources, it can have the opposite effect when the technology does not work. It is highly beneficial for managers to be aware of this, especially when considering that chatbots are a technology that have not been fully developed yet. Chatbots are constantly evolving, and are reliant on a greater input of data in order to be able to answer various customer inquiries without being programmed to answer it. Further advancement of AI and machine learning will result in chatbots becoming even more similar to humans, in the future. For now, a conclusion to be drawn is that managers should seek to adapt new technologies, but should be aware of the consequences of the technology not being fully functioning.

On the other hand, it is apparent that implementing chatbots has cost-saving implications, meaning that less people are needed to be employed as service agents. Staffing costs are generally one of the largest costs that businesses face (Ro, 2015), and from that perspective chatbots can be beneficial for managers to use in order to reduce this. However, what is important for managers in relation to using chatbots, would be to perform cost/benefit analysis of each outcome related to this. In such way, managers can get a clear economical understanding how this double-edged sword actually swings. At the end of the day, the general findings of our study are mixed regarding chatbots versus human chat agents. Hence, chatbots have the least satisfied customers in situations where it does not function properly, but if it still has positive economical consequences, one would assume that managers should probably take the technology in use.

Supplementary, our thesis provides theoretical evidence to the very limited academic area of chatbots as a tool for service recovery. We provide new perspectives to the extensive literature regarding service recovery, by introducing a new way of managing customer complaints, in chatbots. Firstly, our thesis provides some basic ground theory for this area, as there is very limited literature on chatbots in service recovery. In such a way, we establish a relationship between chatbots and human chat agents, and investigate how these types of encounters affect our constructs of customer satisfaction with firm and recovery.

Our results show that the perception of satisfaction regarding the company and the recovery are stronger and more favourable when facing a chatbot than compared to the other two encounters when the outcome of service recovery is successful. This is line with the previous finding that successful recovery is already linked with a positive word of mouth, repurchase intention and loyalty (Gilly & Hansen, 1992; Ndubisi, Malhotra & Miller, 2013; Magnini et al., 2007; Berry, 1995). In other words, a successful handling by a chatbot can create a better reputation and wording regarding a company than a human chat agent. However, the most interesting finding, is that the effect is opposite in an unsuccessful recovery. Here, the results would affect the company and outcome more negatively with a chatbot, than the two other encounters. A more negatively linked relationship with word of mouth, repurchase intention and loyalty are therefore created which is consistent with the findings of Oliver (1997).

In sum, service firms that are considering replacing human chat agents with chatbots to handle service recovery, should be aware of the impact this has on satisfaction when it does not work properly. Hence, managers should be cautious when implementing new technologies, especially removing the human element completely. Our results shed light on the importance that the human touch still has in a service encounter in a world with constant technological improvements. However, there are clearly also many benefits to using chatbots. Along with this, managers should and will always seek to improve its business, and our findings suggest that a less risky option when adopting new technology, is to combine it with humans and not replace it. For now.

7.0 Limitations

Although the result of our study can provide useful guidance for future managers, there are also some limitations present. Firstly, regardless of the many benefits of using scenario-based experiment as previously mentioned, one drawback would be respondents inability of showing how they actually would respond in a real complaint situation (Magnini et al., 2007). In other words, when reading the scenario, one is asked to imagine yourself in that situation, but how one actually reacts is not measured.

We also controlled for age, as it has been proven that age influence perception and how respondents interact with chatbots (Zamora, 2017). Beforehand, we expected that there would be a generational difference in how customers evaluated the different encounters. However, the majority (85.7%) of our respondents was aged between 20-29 years of age. This is not an adequate representative of the general population of Norway and can be identified as an inclusive sampling bias. Users within this age group are more likely to be early adopters of the technology as the use of messaging applications are more frequent compared to the elderly generation (Brandtzaeg & Følstad, 2017). Also, the sample group was recruited by convenience sampling which raises concerns regarding inferences which can be made and biases that can be introduced as a result of this. Furthermore, chatbot is a new terminology, and customers previous experience with chatbots may be limited, especially in complaint situations as it is uncertain if chatbots is used for this purpose yet. Accordingly, companies that focuses on chatbot technology, such as

Google, Facebook and Slack, prioritize the US market (Brandtzaeg & Følstad, 2017). Aligned with this, our study is conducted in Norway, which may induce biases because the institutional context in Norway is different to the USA.

Secondly, there is a threat against the construct validity of this study, as satisfaction with firm and recovery can be identified as too similar to each other. Further, in the factor analysis, a possible confound is present, as our items load on one component. However, all items were gathered from the same scales used in Johnson's et al. (2001) and Maxham & Netemeyer (2002) articles to measure overall satisfaction, satisfaction with recovery and satisfaction with firm.

Lastly, chatbot is a collective term, meaning that there are several types of chatbots. Some are intelligent, some have more anthropomorphic cues attached to it, and some are simpler, designed to answer frequently asked questions. The design of chatbots in our study is not a universal chatbot, thus generalizability would need to be considered. In addition, it is plausible to think that even the way the chatbots hand off the conversation to a human chat agent will affect customer perceptions. Hence, it is apparent that there are extensive areas of research within this topic which needs to be explored further. These will be explained in greater detail in the following section.

8.0 Suggestions for further research

Chatbots are a relatively new terminology in the academic world, especially in relation to service recovery. The present study has shed light upon and outlined tendencies upon how customers react in complaint situations when faced with a chatbot or a human chat agent. Consequently, as chatbots are a relatively new terminology in the world of academics, there are other areas of research that need to be included into the literature regarding the use of chatbots in the service industry.

8.1 Going beyond scenarios

Firstly, an extension of our findings would be to measure customers actual responses faced with a real chatbot, instead of imagining the situation with the help of scenarios. In such cases, respondents would for instance write and possibly adjust their language when talking to a chatbot compared to a human agent. On the other hand, time is included as a variable. Chatbots provide instant answers, whereas human chat agents will spend some time typing. Hence, it is of interest to evaluate the effect the inclusion of these variables will have on customer satisfaction in a complaint situation. Furthermore, as previously mentioned, chatbots is not a common term. In other words, there are great variations regarding what a chatbot is and what purpose it is intended to serve. Previous research has proven that the use of chatbots has been successful in helping to answer straightforward questions or taking payment details in a transaction deal (B. Marr, 2018). Radziwill & Benton (2017) also talk about the successful use of chatbots in harmful ways, which inflate the follower count and instigating the spreading of fake news and rumors. The present study investigated the e-commerce industry and the use of chatbots in complaint handling. However, the bank and finance industry is one of the biggest industries implementing chatbots in their customers service, and therefore it would be interesting to investigate the differences between the industries. Will the perception of chatbots be the same?

Additionally, anthropomorphism seems to be a very delicate concern to handle in relation to chatbots. It has been previously proven that too much anthropomorphic cues are bad, and too little are also bad (Duffy, 2003; Araujo, 2018). Thus, it is necessary to investigate this topic further by including variables that take anthropomorphism, design and usage more into account.

8.2 Chatbots and the justice dimensions

A topic which is highly compelling to research further in the lens of chatbots and AI, is the framework presented by Tax & Brown (1998). Their research found that as much as 85% of the satisfaction is generated by the justice dimensions of the service recovery process (Tax & Brown, 1998). The main takeaway from the study is that customers expect fairness in the actions of the company when the customers have expressed their concern. The justice dimension consist of three parts: (1) procedural justice, (2) interactional justice and (3) outcome justice.

Outcome justice relates to the form of compensation the customer receives during a complaint situation (Tax & Brown, 1998). In our study, the service recovery strategy taken into use was addressing the problem and correcting it. Compensation, discounts, apologies and refunds are examples that there are many other ways one can go about in a complaint situation (Forbes et al., 2005). It would be intriguing to see how customers respond in situations where they are given compensation in the forms of gifts, vouchers or money back by chatbots compared to human chat agents. Another interesting aspect would be to look at the chatbot and see if it is able to distinguish between the different service recovery strategies and use in the correct settings.

Furthermore, Tax & Brown (1998) continue with procedural justice, which relates to how fair the customer feels they have been treated (Tax & Brown, 1998). Would a chatbot perform well in procedural justice because they provide answers instanously? Or would customers feel that they are not important, or worthy of being handled by humans, and be dissatisfied? These are questions future researchers should aim to solve.

The last dimension is interactional justice, which explains the level of justice customers feel in human interactions with employees during the recovery attempt (Rashid, Ahmad & Othman, 2013). The keyword here is human. Chatbots are computers that mimic human behaviour, how would this affect the justice dimension, which again affects customer satisfaction? Together with anthropomorphism and other factors, it would be interesting to see how chatbots create interpersonal relationships in a service encounter. Coupled with perception of fairness, research suggest that trust and commitment are factors that are strongly associated with complaint handling (Kelley & Davis, 1994; Holloway & Beatty, 2003). When customers have high trust for the service company and it fails to deliver on an explicit promise (e.g., “I guarantee you”), customers would feel betrayed and become angrier (Wan, Hui, and Wyer 2011). In this case, an important and interesting factor to investigate is trust and see how this impacts the outcome of an interaction with a chatbot.

8.3 Severity of failure

Wang et al. (2011) mention several types of failures that can cause a service recovery situation, and different degrees of severity a failure can have. Hence, an extension to our findings would be to include these types of variables in a study and check whether or not a chatbot manages to answer according to the various levels of severity and failure, and compare this with a human chat agent.

9.0 References

Books

Berry, L. (1995). *On Great Service: A Framework for Action*, The Free Press, New York, NY

Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2014). *Multivariate Data Analysis: Pearson New International Edition (7th ed).* Harlow

Janssens, W., Wijnen, K., De Pelsmacker, P. & Van Kenhove, P. (2008). *Marketing research with SPSS*. Harlow: Pearson Education Limited

Journal Articles and Web-articles

Anderson, R.E., & Srinivasan, S.S. (2003). E-satisfaction and E-loyalty: A Contingency Framework. *Psychology & Marketing*, 20(2), 123-138

Andreassen, T.W. (2000). Antecedents to satisfaction with service recovery. *European Journal of Marketing*, 34(½), 156-175

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, 85, 183-189.

Arcand, K. (2017). Can Chatbots Fully Replace Humans? Not Yet. *Customer relationship Management,* 21(6), 4

Mimoun, M.S., Poncin, I. & Garnier, M. (2012). Case study—Embodied virtual agents. An analysis on reasons for failure. *Journal of Retailing and Consumer Services*, 19(6), 605–612

Bhandari, M.S., & Polonsky, M. (2011). Variation in Consumer Outcomes: Does the Type of Failure (Process or Outcome) Matter?

Biocca, F., & Harms, C. (2002). Defining and measuring social presence: Contribution to the networked minds theory and measure. *in Proceedings of PRESENCE, 1-36*

Bitner, M. J., Booms, B. H. & Tetreault, M. S. (1990). The service encounter: Diagnosing favorable and unfavorable incidents. *Journal of Marketing,* 54(1), 71–84

Bowen, J.T. & Chen, S-L. (2001). The relationship between customer loyalty and customer satisfaction, *International Journal of Contemporary Hospitality Management*, 13(5), 213-217

Brandtzaeg, P., & Følstad, A. (2017). Why people use chatbots. *In International Conference on Internet Science*, 377-392.

Brave, S., & Nass, C. (2002). Emotion in human-computer interaction. In: Jacko, J.A., Sears, A. (eds.) The human-computer interaction handbook, pp. 81–96. L. Erlbaum Associates Inc, NJ, USA

Campbell-Miller, S. (2017). Chatbots being tested for new state website. *Arkansas Business, 34*(39), 21.

Choudhury, S. R. (2016). SoftBank's Pepper Robot Gets a Job Waiting Tables at Pizza Hut. *CNBC*, Retrieved from https://www.cnbc.com/2016/05/24/mastercard-teamed-up-with-pizza-hut-restaurants-asia-to-bring-robots-into-the-pizza-industry.html

Ciechanowski, L., Przegalinska, A., Magnuski, M. & Gloor, P. (2018). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, 92, 539-548

Cyr, D. (2008). Modeling Website Design Across Cultures: Relationships to Trust, Satisfaction and E-loyalty. *Journal of Management Information Systems*, 24(3), 47-72.

Collins, R. (2018). The human-machine symbiosis. *Management issues*. Retrieved from: https://www.management-issues.com/opinion/7298/the-human-machin-symbiosis/

Coyle, J. R. & Thorson, E. (2001). The effects of progressive levels of interactivity and vividness in web marketing sites. *Journal of advertising*, 30(3), 65-77.

Dabholkar, P. A. (1996). Consumer evaluations of new technology-based self-service options: an investigation of alternative models of service quality. *International Journal of research in Marketing*, *13*(1), 29-51.

Dale, R. (2016). The Return of the Chatbots. *Natural Language Engineering*, 22(5), 811–817.

De Angeli, A., Johnson, G. I. & Coventry, L. (2001). The unfriendly user: exploring social reactions to chatterbots. in *International Conference on Affective Human Factors Design*, Asean Academic Press: London.

Devaney, E. (2016). Infographic: Rise of the chatbots. Retrieved from https://www.drift.com/blog/chatbots-infographic/

Dong, B., Evans, K.R. & Zou, S. (2008). The Effects of Customer Participation in Co-Created Service Recovery. *Journal of the*

Academy of Marketing Science, 36(1), 123-137.

Duffy, B.R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, 42(3-4), 177-190.

Fan, Y.W, Wu, C. C & Wu, W.T. (2010). Department of Information Management, National Central University, Taiwan. *The 2010 International Conference on Innovation and Management,* Penang, Malaysia

Feine, J., Morana, S. & Gnewuch, U. (2019). Measuring Service Encounter Satisfaction with Customer Service Chatbots using Sentiment Analysis, In Proceedings of the 14th International Conference on Wirtschaftsinformatik (WI2019), Siegen, Germany, February 24–27.

Ferrara, E., Varol, O., Davis, C., Menczer, F. & Flammini, A. (2016). The Rise of Social Bots. *Communications of the ACM*, 59(7), 96-104.

Forbes, L. P., Kelley, S. W. & Hoffman, K. D. (2005). Typologies of e-commerce retail failures and recovery strategies. *Journal of Services Marketing*, 19(5), 280-292.

Froehle, C.M. & Roth, A.V. (2004). New measurement scales for evaluating perceptions of the technology-mediated customer service experience. *Journal of Operations Management*, 22(1), 1-21

Gilly, M.C. & Hansen, R.W. (1992). Consumer complaint handling as a strategic marketing tool. *Journal of Consumer Marketing*, 2(4), 5-16.

Goodwin, C. & Ross, I. (1992). Consumer Responses to Service Failures: Influence of Procedural and Interactional Fairness Perceptions. *Journal of Business Research*, 25(2), 149-163, Doi: 10.1016/0148-2963(92)90014-3

Gremler, D.D. & Brown, S.W. (1999). The loyalty ripple effect: Appreciating the full value of customers. I*nternational Journal of Service Industry Management*, 10(3), 271-293.

Grönroos, C. (1988). Service quality: the six criteria of good perceived service. *Review of Business*, 9(3), 10-30.

Grönroos, C. & Voima, P. (2013). Critical service logic: making sense of value creation and co-creation. *Journal of the Academy of Marketing Science*, 41(2), 133-150.

Halstead, D. & Page, T.J. (1992). The effects of satisfaction and complaining behavior on consumer repurchase intentions. *Journal of Consumer Satisfaction*, 5, 1-11.

Hill, J., Ford, W. R. & Farreras, I.G. (2015). Real Conversations with Artificial Intelligence: A Comparison Between Human–Human Online Conversations and Human–Chat-bot Conversations. Compute. Hum. Behave.
49, 245–250.

Hoffman, D. K., Scott, K. W. & Rotalsky, H.M. (1995). Tracking Service Failures and Employee Recovery Efforts. *Journal of Services Marketing*, 9(2), 49-61.

Holloway, B. B. & Beatty, S. E. (2003). Service failure in online retailing: A recovery opportunity. *Journal of Service Research*, 6(1), 92-105.

Huang, M-H. & Rust, R.T. (2018). Artificial Intelligence in Service. *Journal of Service Research*, 21(2), 155-172.

Hyken, S. (2017). AI and Chatbots Are Transforming the Customer Experience. *Forbes*, Retrieved from https://www.forbes.com/sites/shephyken/2017/07/15/ai-and-chatbots-are-transforming-the-customer-experience/#414adc4241f7

Jarrahi, M.H. (2018). Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organizational Decision Making. *Business Horizons*, 61(4), 577-586.

Johnson, M.D., Gustafsson, A., Andreassen, T.W., Lervik, L. & Cha, J. (2001). The evolution and future of national customer satisfaction index models. *Journal of Economic Psychology*, 22(2), 217-245.

Jones, M. A., Mothersbaugh, D. L. & Beatty, S. E. (2000). Switching Barriers and Repurchase Intention in Services. *Journal of Retailing*, 76 (2), 259-74.

Jones, M. A. & Suh, J. (2000). Transaction-specific satisfaction and overall satisfaction: an empirical analysis. *Journal of Services Marketing*, 14(2), 147-159

Julia, N. (2018). E-commerce chatbots: 22 ways to increase sales, conversions & Retention. *Shopify,* Retrieved from https://www.shopify.com/enterprise/ecommerce-chatbots

Kau, A. K. & Loh, E. W. (2006). The effects of service recovery on consumer satisfaction: A comparison between complainants and non complainants. *The Journal of Services Marketing*, 20 (2), 101-111.

Kim, J-H. & Jang, S. C. (2014). A scenario-based experiment and a field study: A comparative examination for service failure and recovery. *International Journal of Hospitality Management*, 41, 125-132.

Kelley, S. W. & Davis, M. A. (1994). Antecedents to customer expectations for service recovery. *Journal of the Academy of Marketing Science*, 22(1), 52-61.

Kelley, S. W., Hoffman, K. D. & Davis, M. A. (1993). A Typology of Retail Failures and Recoveries. *Journal of Retailing*, 69, 429-452.

Komunda, M. & Osarenkhoe. A. (2012). Remedy or cure for service failure?: Effects of service recovery on customer satisfaction and loyalty. *Business Process Management Journal*, 18(1), 82-103.

Larivière, B., Bowen, D., Andreassen, T. & Kunz, W., Sirianni, N., Voss, C., Wünderlich, N. & De Keyser, A. (2017). "Service Encounter 2.0": An investigation into the roles of technology, employees and customers. *Journal of Business Research*, 79, 238-246.

Magnini, V.P., Ford, J.B., Markowski, E.P. & Honeycutt Jr, E.D. (2007). The service recovery paradox: justifiable theory or smoldering myth?. *Journal of Services Marketing*, 21(3), 213-225.

Marr, B. (2018). The Amazing Ways Chinese Tech Giant Alibaba Uses Artificial Intelligence And Machine Learning. *Forbes.* Retrieved from https://www.forbes.com/sites/bernardmarr/2018/07/23/the-amazing-ways-chinese-tech-giant-alibaba-uses-artificial-intelligence-and-machine-learning/#7a89d3be117a

Maxham, J.G & Netemeyer, R.G. (2002). Modeling customer perceptions of complaint handling over time: the effects of perceived justice on satisfaction and intent. *Journal of Retailing,* 78(4), 239-252, doi: https://doi.org/10.1016/S0022-4359(02)00100-8.

Moon, H. & Conlon, D.E. (2002). From Acclaim to Blame: Evidence of a Person Sensitivity Decision Bias. *Journal of Applied Psychology*, 87(1), 33–42.

MyClever, Drift, Salesforce & SurveyMonkey Audience. (2018). State of Chatbots report, Retrieved from https://mk0drift0ho9g7wbfexi.kinstacdn.com/wp-content/uploads/2018/01/2018-state-of-chatbots-report.pdf

Nadella, S. (2016). The Partnership of the Future. *Slate*, Retrieved from https://slate.com/technology/2016/06/microsoft-ceo-satya-nadella-humans-and-a-i-can-work-together-to-solve-societys-challenges.html

Ndubisi, N.O., Malhotra, N.K. & Miller, G. L. (2013). Customer Reactions to Conflict Management: A Review and Empirical Evidence from Two

Service Industries, in Naresh K. Malhotra (ed.). *Review of Marketing Research, 10* Emerald Group Publishing Limited, 63 - 96.

NIST/SEMATECH. (2013). *e-Handbook of Statistical Methods*, Retrieved from http://www.itl.nist.gov/div898/handbook/, 23.04.2019 (*1.3.5.11.Measures of Skewness and Kurtosis)*

Nordea. (2017). Ask Nova - Get a taste of the future. Retrieved from https://www.nordea.com/en/press-and-news/news-and-press-releases/the-digital-hub/2017/2017-06-28-ask-nova-get-a-taste-of-the-future.html

Oliva, T. A., Oliver, R. L. & MacMillan, I. C. (1992). A catastrophe model for developing service satisfaction strategies. *Journal of Marketing*, 56, 83-95.

Oliver, R.L. (1997). Satisfaction: A behavioral perspective on the consumer. McGraw Hill, New York

Pugh, S.D. (2001). Service with a smile: Emotional Contagion in the Service Encounter. *The Academy of Management Journal*, 44(5), 1018-1027.

Radziwill, N. & Benton, M. (2017). Evaluating Quality of Chatbots and Intelligent Conversational Agents.

Reichheld, F. F. (1993). Loyalty-Based Management. *Harvard business review,* 71(2), 64-73.

Ro, S. (2015). Labor accounts for 60% of corporate expenses, and it’s only getting more expensive. *Business Insider.* Retrieved from: *https://www.businessinsider.com/citi-levkovich-rising-labor-costspressure-profit-margins-sp600-2016-2015-12?r=US&IR=T*

Rui Sousa, R. & Voss, C.A. (2009). The effects of service failures and recovery on customer loyalty in e-services: An empirical investigation. *International Journal of Operations & Production Management,* 29(8), 834-864.

Saarem, A. C (2016). Why would I talk to you. Received from: https://www.ntnu.no/design/forskning/artikkel

Shawar, B. A. & Atwell, E. (2007). Chatbots: are they really useful? *LDV Forum, 22*, 29-49.

Shostack, L.G. (1985). “Planning the Service Encounter,” in The Service Encounter: Managing Employee/Customer Interaction in Service Businesses, John A. Czepiel, Michael R. Solomon, and Carol F.

Suprenant,

eds. New York: Lexington Books, 243-254.

Smith, A. K., Bolton, R. N., & Wagner, J. (1999). A model of customer satisfaction with service encounters involving failure and recovery. *Journal of Marketing Research, 36(3)*, 356–372.

Smith, A. K. & Bolton, R. N. (1998). An Experimental Investigation of Service Failure and Recovery: Paradox or Peril?. *Journal of Service Research*, 1(1), 65-81.

Solon, O. (2018). Google's robot assistant now makes eerily lifelike phone calls for you. *The Guardian.* Retrieved from https://www.theguardian.com/technology/2018/may/08/google-duplex-assistant-phone-calls-robot-human

Spreng, R. A., Harrell, G. D. & Mackoy, R. D. (1995). Service recovery: Impact on satisfaction and intentions. *Journal of Services Marketing*, 9 (1), 15-23.

Statistics Solutions. (2013). *Normality*. Retrieved from http://www.statisticssolutions.com/academicsolutions/resources/directory-of-statistical-analyses/normality/

Tax, S. S. & Brown, S. W. (1998). Recovering and Learning from Service Failure. *MIT Sloan Management Review*, 40(1), 75-88.

Tripathy, A. (2018). Why Humans Need To Find Symbiosis With AI. *Forbes*. Retrieved from https://www.forbes.com/sites/forbestechcouncil/2018/01/29/why-humans-need-to-find-symbiosis-with-ai/#6c407d35407a

Todorovic, A. (2015). Has the Turing test been passed? Retrieved from http://isturingtestpassed.github.io

Vegesna, A., Jain, P. & Porwal, D. (2018). Ontology based Chatbot (For E-commerce Website). *International Journal of Computer Applications*, 179(14), 51-55.

Wan, L. C. & Chan, E. (2019). Failure is Not Fatal: Actionable Insights on Service Failure and Recovery for the Hospitality Industry. *Boston Hospitality Review.* Retrieved from https://www.bu.edu/bhr/files/2019/02/Failure-is-Not-Fatal-Actionable Insights-on-Service-Failure-and-Recovery-for-the-Hospitality-Industry-Lisa-Wan-and-Elisa-Chan.pdf

Wang, Y. S, Wu, S. C, Lin, H. H. & Wang, Y. Y. (2011). The relationship of service failure severity, service recovery justice and perceived switching costs with customer loyalty in the context of e-tailing. *International Journal of Information Management*, 31(4), 350-359.

Wilson, H. J., Daugherty, P. R. & Morini-Bianzino, N. (2017.) Will AI Create as Many Jobs as it Eliminates? MIT Sloan Management Review. Retrieved from http://sloanreview.mit.edu/article/will-ai-create-as-many-jobs-as-it-eliminates/

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S. & Martins, A. (2018). Brave new world: service robots in the frontline, *Journal of Service Management*, 29(5), 907-931.

Wüenderlich, N.V., & Paluch, S. (2017). A Nice and Friendly Chat with a Bot: User Perceptions of AI-Based Service Agents. *ICIS*.

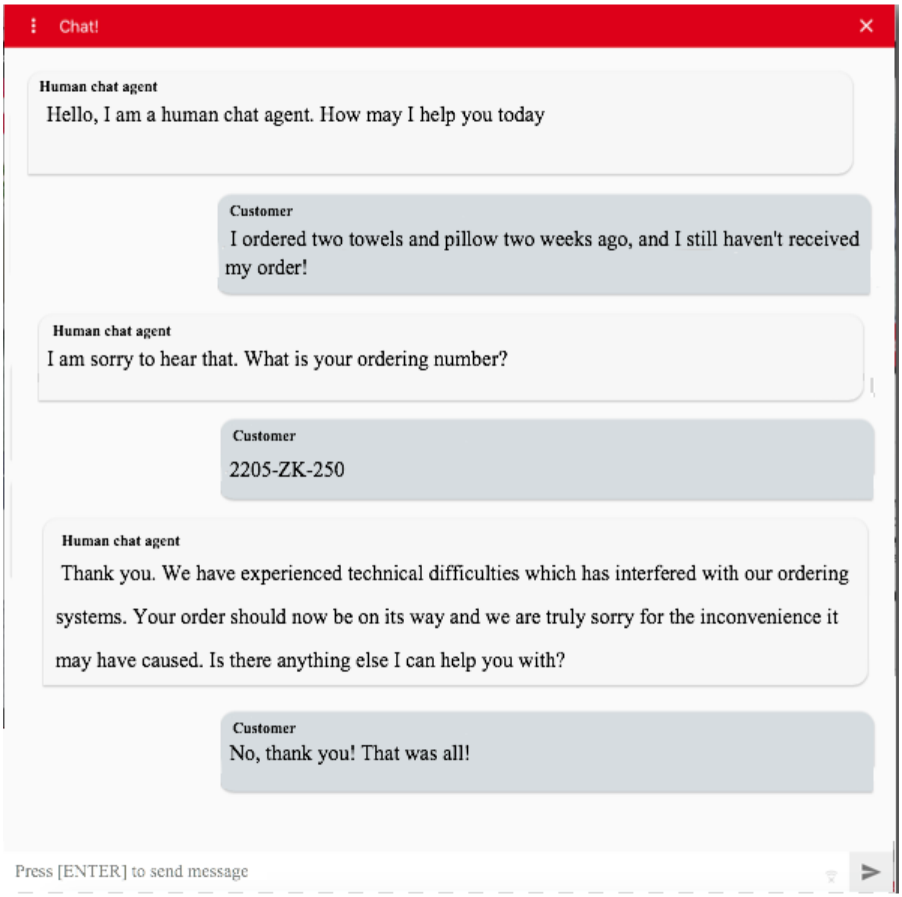
Zamora, J. (2017a). Rise of the Chatbots: Finding A Place for Artificial Intelligence in India and US. *Conference Paper.* Doi: 10.1145/3030024.3040201

Zamora, J. (2017b). I'm Sorry, Dave, I'm Afraid I Can't Do That: Chatbot Perception and Expectations. *Conference Paper*, Doi: 10.1145/3125739.3125766

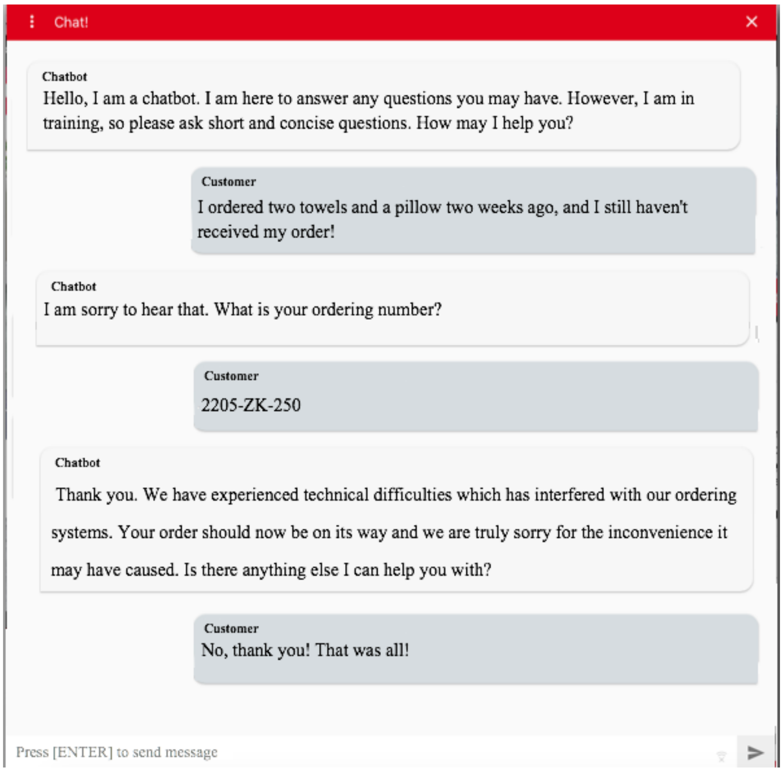
10.0 Appendices

Appendix 1: Service recovery scenarios

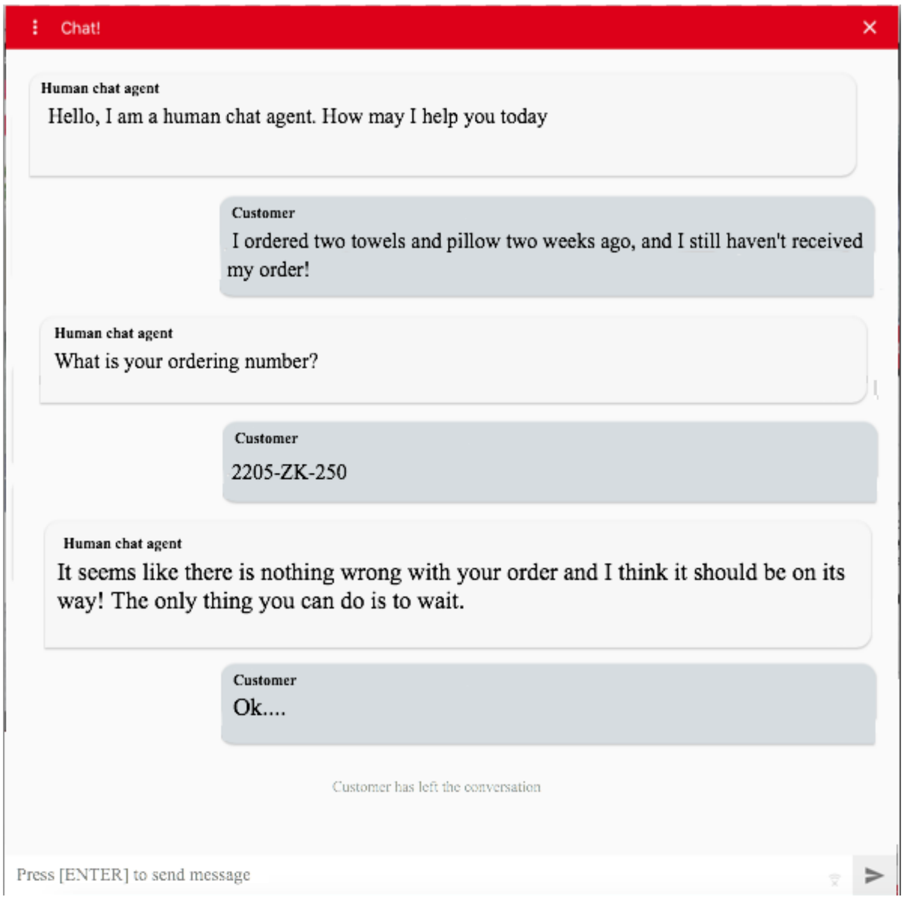
Successful service recovery with human



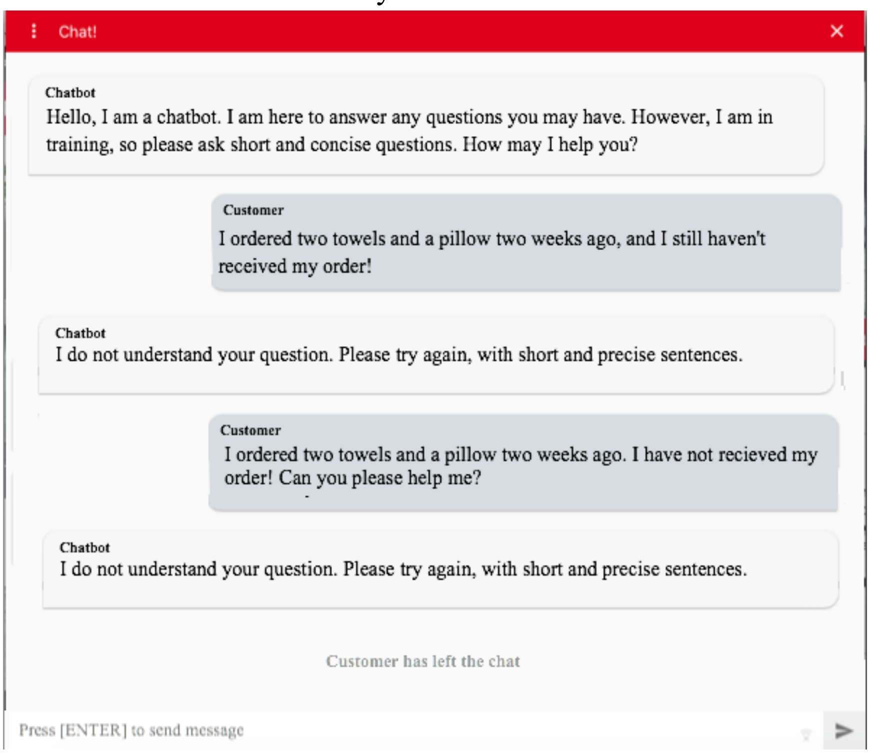
Successful service recovery with chatbot



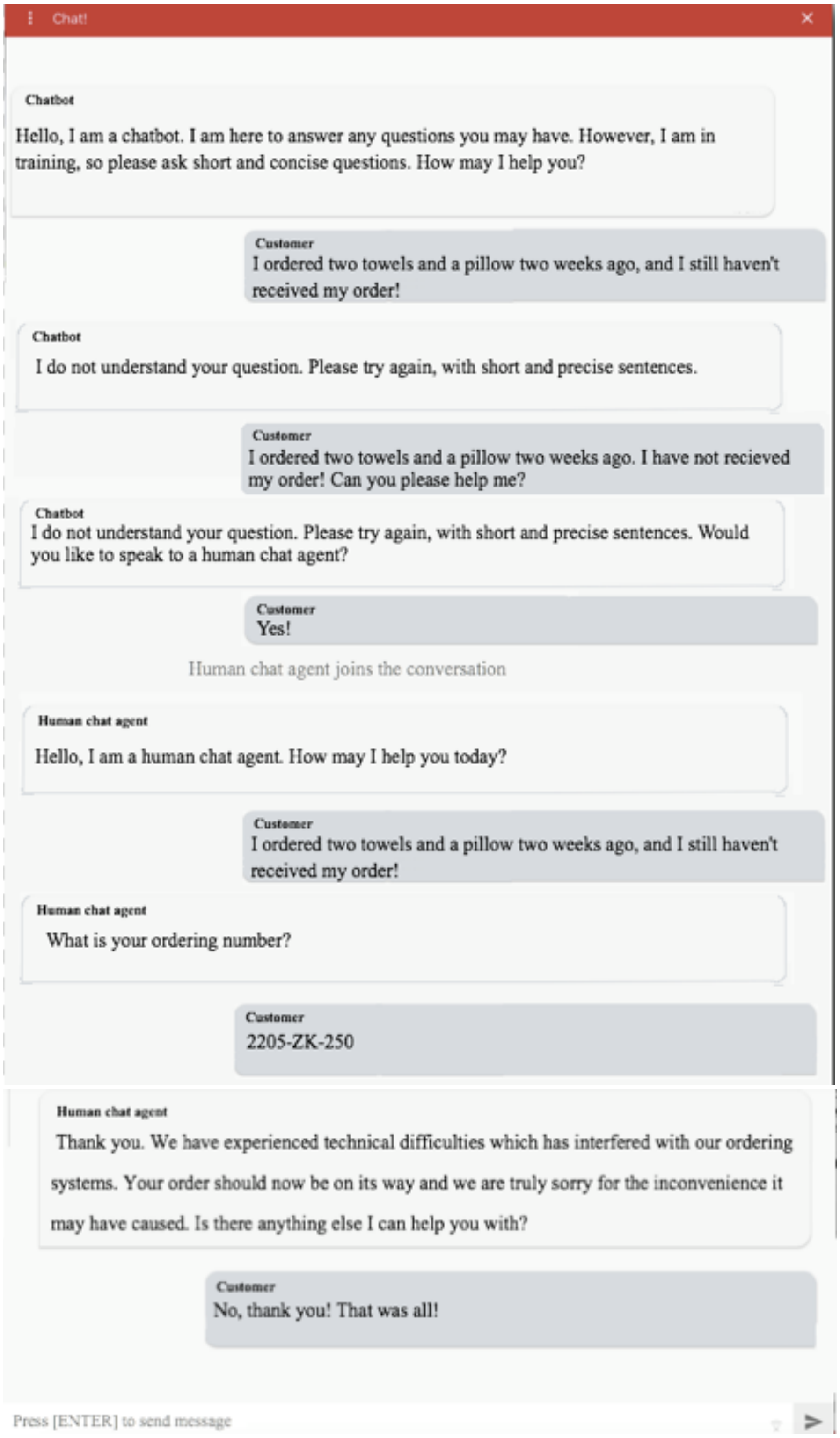
Unsuccessful service recovery with human



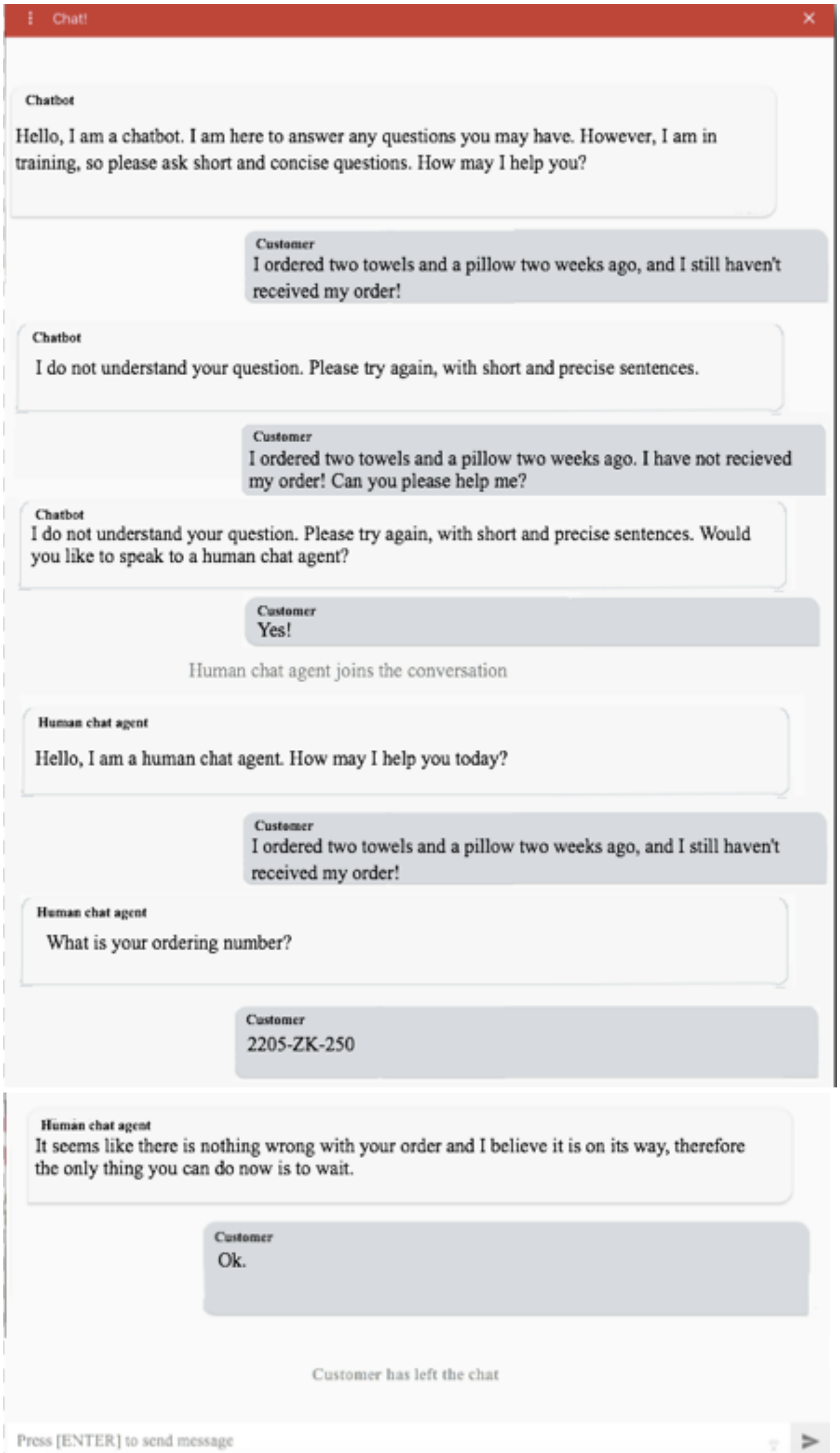
Unsuccessful service recovery with chatbot



Successful service recovery with both chatbot and human



Unsuccessful service recovery with both chatbot and human



Appendix 2: Questionnaire

Hello!

This survey is made in context with our final master thesis of our Master of Science degree in Business at BI, Norwegian Business School.

Participation in this study is voluntary and there is no right or wrong answer, we are merely interested in your honest opinion. Every answer will be anonymous, and will only be used in purpose of our study. We will not gather any personal information such as e-mail, IP-address or names.

You will first get a question regarding your use of chatbots and then a scenario on the next page. Please take your time to read this carefully before answering the questions afterwards. The survey will take approximately 5 minutes to conduct.

We appreciate you taking the time to help us answer our survey!

Thank you!

Arne & Peter

"A chatbot is a computer program that simulates human conversation through voice commands or text chats or both".

Have you ever interacted with a chatbot?

- ◯ Yes
- ◯ No
- ◯ Don't know

If yes, how often have you used a chatbot?

- ◯ 1-3 times
- ◯ 4-6 times
- ◯ 7-9 times
- ◯ 10 or more
- ◯ Never

Please read the following passage. While reading, think about how you feel and how you would react in this situation and answer the attached survey.

Imagine that you are in this situation:

You have ordered two towels and a pillow from the e-commerce store of Beds & Pillows Inc. On the website it said: guaranteed delivery within 10 days. You have waited for two weeks and the order has still not arrived, and you have received no updates or information from Beds & Pillows Inc. regarding your order. Therefore, you decide to take matter into your own hands and decide to contact the store. You decide to enter their website's chat interface:

Think about the problem you had with the order, and how the problem was handled:

How satisfied are you with the resolution to your problem?

Not at all						Very satisfied
1	2	3	4	5	6	7
○	○	○	○	○	○	○

To what extent does the resolution of your problem meet your expectations?

Not at all						To a great extent
1	2	3	4	5	6	7
○	○	○	○	○	○	○

If you imagine the ideal resolution to this problem. What is your opinion about the resolution?

Not very close to ideal						Very close to ideal
1	2	3	4	5	6	7
○	○	○	○	○	○	○

Think about the company overall:

How satisfied are you with Beds & Pillows Inc?

Very dissatisfied						Very satisfied
1	2	3	4	5	6	7
○	○	○	○	○	○	○

To what extent does Beds & Pillows meet your expectations?

Not at all						To a great extent
1	2	3	4	5	6	7
○	○	○	○	○	○	○

If you imagine the ideal customer service for a home essential shop. What is your opinion on Beds & Pillows Inc services?

Not very close to ideal provider						Very close to ideal provider
1	2	3	4	5	6	7
○	○	○	○	○	○	○

Would you prefer to speak to an agent or use automation (chatbot), if the outcome and time were equal?

- ○ To speak to an employee
- ○ To use automation/self-service
- ○ Have no preferences

What is your preferred method for contacting a firm?

- ◯ Website self-service/mobile app
- ◯ Call the call center
- ◯ Email
- ◯ Visit the store/office
- ◯ Social media
- ◯ Web chat
- ◯ No preference
- ◯ Don't know/Not applicable

Age?

- ◯ <20
- ◯ 20-29
- ◯ 30-39
- ◯ 40-49
- ◯ 50-59
- ◯ 60-69
- ◯ 70+

Gender?

- ◯ Male
- ◯ Female
- ◯ Other

Marital status?

- ◯ Single, never married
- ◯ In a relationship
- ◯ Married, or domestic partnership
- ◯ Divorced
- ◯ Widowed

What is your yearly income?

- ◯ <200.000
- ◯ 200.000 - 499.000
- ◯ 500.000 - 799.000
- ◯ 800.000-1.099.000
- ◯ 1.100.000+
- ◯ Don't want to answer

Occupation

- ◯ Student part-time
- ◯ Student full-time
- ◯ Unemployed
- ◯ Employed part-time
- ◯ Employed full-time
- ◯ Pensioner

What is the highest degree or level of school you have completed?

- Elementary school
- High School
- Bachelor's Degree
- Master's Degree
- Doctorate (e.g. PHD)
- Other (please specify)

Appendix 3: Factor analysis

KMO and Bartlett's Test		
Kaiser-Meyer-olkin Measure of Sampling Adequacy		.868
Bartlett's Test of Sphericity	Approx. Chi-Square	1138.198
	df	15
	Sig.	.000

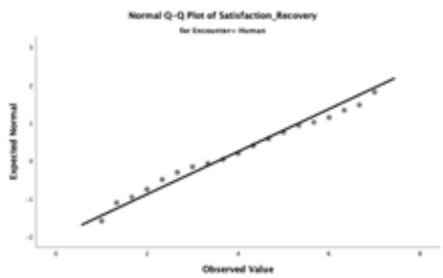
Total Variance Explained

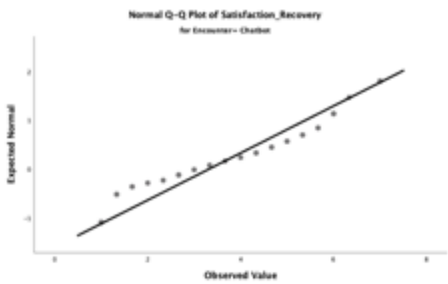
	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.608	76.795	76.795	4.608	76.795	76.795
2	.563	9.382	86.177			
3	.350	5.833	92.009			
4	.218	3.638	95.647			
5	.151	2.515	98.162			
6	.110	1.838	100.000			

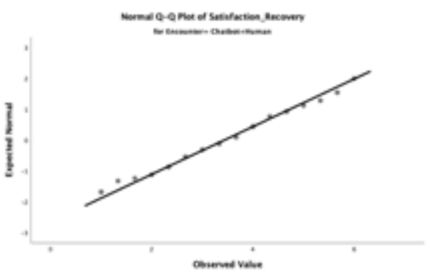
Component Matrix	
	Component
	1
Satisfaction_Recovery_2	.903
Satisfaction_Recovery_1	.896
Satisfaction_Firm_3	.893
Satisfaction_Firm_1	.890
Satisfaction_Firm_2	.871
Satisfaction_Recovery_3	.800

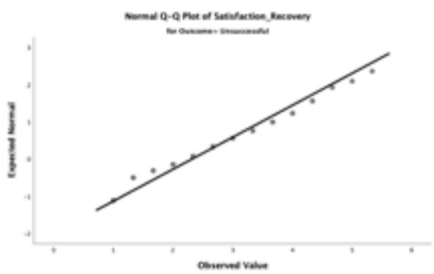
Appendix 4: Q-Q plots

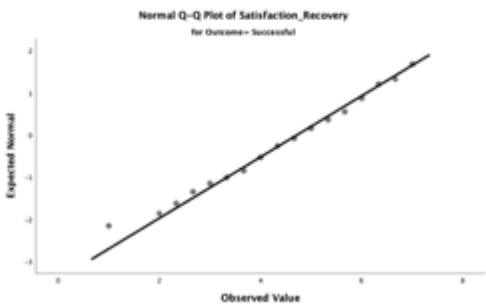
Dependent variable: Satisfaction with Recovery



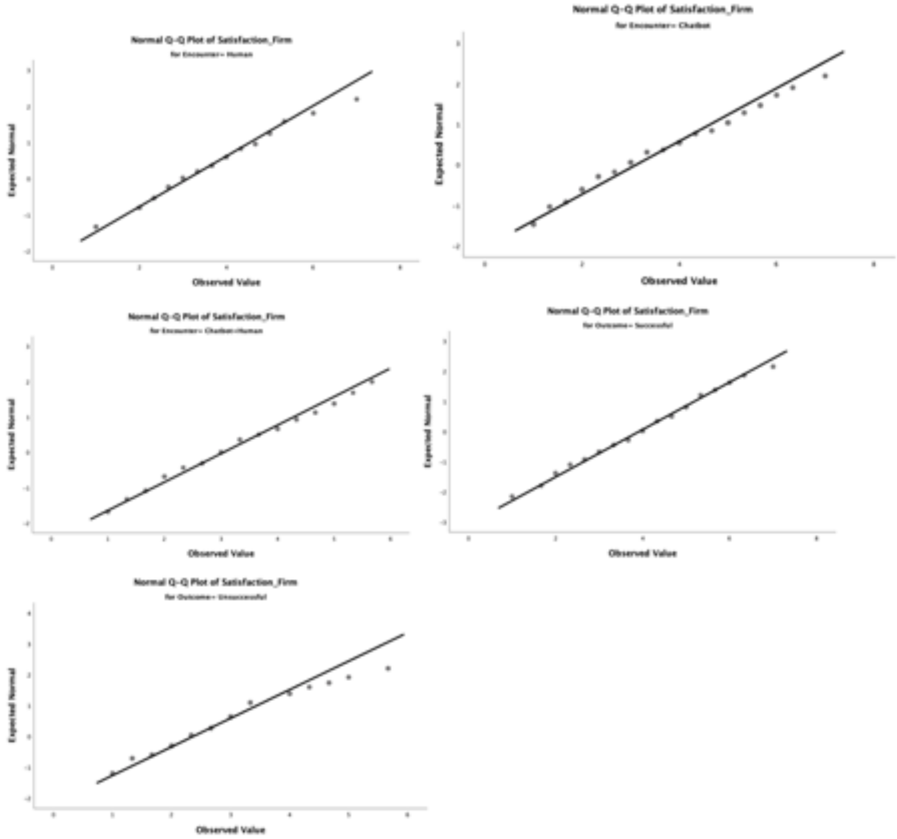








Dependent variable: Satisfaction with Firm



Appendix 5: Manova - Test of Between Subjects Effects

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Satisfaction_Recovery	334.694	5	66.939	46.738	.000	.543
	Satisfaction_Firm	140.113	5	28.023	21.332	.000	.351
Intercept	Satisfaction_Recovery	2529.719	1	2529.719	1766.310	.000	.900
	Satisfaction_Firm	2001.997	1	2001.997	1524.009	.000	.886
Outcome	Satisfaction_Recovery	286.923	1	286.923	200.337	.000	.504
	Satisfaction_Firm	120.153	1	120.153	91.465	.000	.317
Encounter	Satisfaction_Recovery	1.339	2	.670	.467	.627	.005
	Satisfaction_Firm	1.494	2	.747	.569	.567	.006
Outcome * Encounter	Satisfaction_Recovery	40.752	2	20.376	14.227	.000	.126
	Satisfaction_Firm	17.505	2	8.753	6.663	.002	.063
Error	Satisfaction_Recovery	282.145	197	1.432			
	Satisfaction_Firm	258.787	197	1.314			

a. R Squared = .543 (Adjusted R Squared = .531)

b. R Squared = .351 (Adjusted R Squared = .335)