



The effect of chatbot introduction on user satisfaction.

Master Thesis
Frank Hendriks
May 2019

THE EFFECT OF CHATBOT INTRODUCTION ON USER SATISFACTION.

Master Thesis in Information Management for MSc. IMMIT

Author:

Frank Hendriks

Supervisor Tilburg University:

Prof. dr. X. (Carol) Ou

Supervisor KPMG:

Sander Bockting

Words:

22.126

May 2019

PREFACE

How is it possible that a well-designed chatbot by one of the largest tech companies turns into a racist and sexist in less than a day? (Horton, 2016; Price, 2016).

In 2016 Microsoft started an experiment by developing a chatbot named Tay. Tay was a self-learning chatbot, based on artificial intelligence, and released on Twitter. Tay could interact with the other human users of twitter. The goal was to build a chatbot that could learn way of speech and interaction from humans, and experiment with conversational understanding (Wolf, Miller, & Grodzinsky, 2017).

The learning part worked well, as Tay started having pleasant conversations and shouting nice tweets. However, Tay learnt quickly. She was fed with input from other users from Twitter, and she learnt automatically. And as the saying goes: ‘garbage in, garbage out’, she soon started to spill out sexist and racist tweets. She picked these lines up from other users on Twitter.

If we trace back the root cause of these racist tweets, we see that Tay wasn’t really much more than a parrot with an internet connection. It was very simple to manipulate her and to let her tell whatever you wanted (Vincent, 2016).

Nevertheless this experiment teaches us a lesson on the use and expectations of artificial intelligence. We sometimes think that AI will solve our current problems and will enhance efficiency on various aspects.

But, we should never forget that all the technology we use always learns from ourselves. This holds for all technology we implement, but especially for the artificial intelligence – the self-learning technology. All these technologies will definitely make our lives easier, but they will never erase flaws in humankind, which we deal with every day. Just as Tay wasn’t able to overcome the racist tweets.

I hope you’ll enjoy reading this thesis just as much as I enjoyed making it, and I hope it will increase your understanding about chatbots, and their future impact. But with this preface I also like to prove that technology itself isn’t always the solution: technology always needs humans to determine the way we want to use it.

ACKNOWLEDGEMENTS

This thesis has been constructed with the utmost effort of the author. Nevertheless, without the help of a lot of people, the quality of the delivered work would have been much lower. Therefore I would like to start with thanking several people for their contributions.

First of all, I would like to thank my supervisor from Tilburg University, Prof. Dr. Carol Ou. Throughout the process Carol has always been available to ask questions and inquire for advice. Her attitude to think along and help where possible have severely impacted the quality of this thesis in a good way. I have enjoyed our cooperation and would like to thank Carol for all the fruitful discussions and feedback sessions.

I also like to thank my supervisor from the company I have written my thesis for: Sander Bockting. He has offered professional support throughout the writing process. His aim was to keep an overview of the planning and progress, and aid with finding my way through KPMG, as well as giving academic support where necessary. His advice has definitely increased my efficiency in the writing process, and therefore I would like to thank Sander.

Next to this, I have had very helpful support from all the colleagues within KPMG, and especially all the colleagues from the ES academy, who provided very productive meetings. These regular meetings have helped me to keep on track, and provided critical looks at my research design and purpose. This helped in keeping a focused approach which has enhanced the end result. A big shout out to all the colleagues involved in this project.

Also a special thanks to all the other interns for their input, feedback and advice. Their advice, expressed in the many coffee meetings we had, was very helpful.

Also a final and big thanks to everyone who spend 7 minutes of their time filling in my thesis survey!

MANAGEMENT SUMMARY

This thesis research aims to answer the research question: Which chatbot introduction yields the highest user satisfaction? It does so by executing an experiment where participants are exposed to various chatbot introductions and by measuring their satisfaction afterwards.

A chatbot is an online conversational partner, which allows internet users to have a conversation online with a robot. This conversation takes place by sending and receiving text written message, a so called chat. Their use is expected to bring huge advantages to customer service and web care environments. Online web shops are already using chatbots to handle customers' questions regarding product choice, product return and answering frequently asked questions.

Chatbots are the new next trend. Their presence online is growing vastly, and more and more online users have frequent interaction with a chatbot. Technical development is fast, and implementing a chatbot has become accessible and uncomplicated.

But next to this technical development, there is also the side of human interaction. Chatbot's acceptance rate, customer satisfaction, and alike factors are highly depending on the way a chatbot interacts, and how 'human' this interaction is perceived by the user. This research focusses on this interaction element.

Relevance is extracted by the fact that chatbots are getting smarter, and technical development makes it possible for a chatbot to perfectly imitate a human being. This would make it impossible for users to distinguish their conversational partner from either a chatbot or a human being.

In the future it will therefore be important for a chatbot to properly introduce itself, since users can only derive the identity of their conversational partner by the way this partner introduces itself. This makes it important to know how different users' reactions are on various ways of introducing and which introduction method will yield the highest satisfaction for the user.

Literature research is mainly focused at retrieving similar research and describing the leading theories in this field. The theoretical part mainly builds on existing theories in social sciences focused on communication. By using various paradigms in human-computer interaction, these theories from another research discipline are relayed to this topic. This approach is executed due to the lack of relevant theories available in chatbot research due to the novelty of this topic.

The literature research mainly shows two important theories: the CASA paradigm and CMC theories. CASA stands for Computer Are Social Actors, and research to this subject reveals that social theories are also applicable to human-computer interaction, instead of only human-human interaction. This was proven in an experimental way, by replacing one social actor in communication theories by a computer. The second most important finding in the literature research was research in CMC, Computer Mediated Interaction. These theories describe the difference of interaction between humans among themselves, or their interaction via or with computers.

The third chapter elaborates on the theoretical framework, which was derived by looking at similar research. The aim was to provide measurable constructs in order to determine the impact various introductions have on users. By looking at similar research, three usable constructs have been extracted: Social Presence, Perceived Humanness and Service Encounter Satisfaction. These three constructs are used in the final experiment, and measure respectively: the satisfaction with the held conversation; the satisfaction with the conversational partner; and the satisfaction with the given advice and treatment.

In the final experiment, participants first watch a video showing an excerpt of a chatbot conversation, and according to one of the three experiment groups the participant is in, are exposed to a unique introduction, which is the manipulated variable. The three different experiment groups watch three different introductions, which range from a human being, an undefined conversational partner, to a chatbot.

Based on this research, the following conclusions can be derived. First of all, users prefer to talk to a real human being in all cases, as this leads to a higher satisfaction on all constructs. Second, if users talk with a chatbot, it is better for a chatbot to introduce itself to the user in an undefined way, instead of explicitly stating its true identity (e.g. “Hello, how can I help you?” is better than “Hello, I’m a chatbot, how can I help you?”). This undefined introduction leads to a higher rating on social presence and perceived humanness, which translates in a higher satisfaction with the conversation and with the conversational partner. Thirdly, this thesis concludes that for the last measured construct, service encounter satisfaction, users are indifferent of the measured satisfaction. This was displayed by a non-significant difference between the measured results on this construct. So for the satisfaction with the final advice and treatment, the users do not mind if they are talking to a chatbot or a real human, nor the way this conversational partner introduces itself.

TABLE OF CONTENTS

| | |
|--|-----|
| PREFACE..... | I |
| ACKNOWLEDGEMENTS..... | III |
| MANAGEMENT SUMMARY | IV |
| TABLE OF CONTENTS..... | VI |
| LIST OF FIGURES | IX |
| LIST OF TABLES..... | X |
| 1 INTRODUCTION..... | 1 |
| 1.1 Introduction..... | 1 |
| 1.2 Promising Areas..... | 2 |
| 1.3 Current State of Literature | 3 |
| 1.4 Research Gap | 4 |
| 1.5 Research Question..... | 5 |
| 1.6 Sub Questions | 5 |
| 1.7 Research Design..... | 5 |
| 1.8 Relevance..... | 6 |
| 2 LITERATURE RESEARCH..... | 9 |
| 2.1 Turing Test..... | 9 |
| 2.2 What is a Chatbot? | 11 |
| 2.3 Theories on Chatbot Interaction..... | 12 |
| 2.4 Limitations of the CASA paradigm concerning Chatbot interaction..... | 15 |
| 2.5 Related research in Chatbot Communication..... | 16 |
| 3 THEORETICAL FRAMEWORK..... | 25 |
| 3.1 Used Constructs in Literature | 25 |
| 3.2 Social Presence | 30 |
| 3.3 Perceived humanness | 31 |
| 3.4 Service Encounter Satisfaction | 33 |

| | | |
|------|---|----|
| 3.5 | Difference between Social Presence and Perceived Humanness..... | 34 |
| 3.6 | Social Information Processing Theory..... | 35 |
| 3.7 | Relationships..... | 37 |
| 3.8 | Conceptual Model..... | 41 |
| 4 | RESEARCH DESIGN..... | 43 |
| 4.1 | Experiment Design..... | 43 |
| 4.2 | Vignette Research | 44 |
| 4.3 | Manipulation | 46 |
| 4.4 | Condition Design | 48 |
| 4.5 | Experimental Control..... | 49 |
| 4.6 | Pre-test | 50 |
| 4.7 | Power Analysis | 52 |
| 4.8 | Measures | 53 |
| 4.9 | Participants..... | 54 |
| 4.10 | Data gathering..... | 54 |
| 4.11 | Procedure | 56 |
| 4.12 | Analysis..... | 57 |
| 4.13 | Ethics..... | 58 |
| 5 | RESULTS..... | 59 |
| 5.1 | Error Rate and Dropout Rate | 59 |
| 5.2 | Basic Demographical Information..... | 60 |
| 5.3 | Manipulation Check..... | 60 |
| 5.4 | Scale Reliability | 61 |
| 5.5 | Significance..... | 61 |
| 5.6 | Results..... | 62 |
| 5.7 | Regression..... | 63 |
| 5.8 | Hypothesis Testing..... | 68 |
| 6 | CONCLUSION, LIMITATIONS & FUTURE RESEARCH..... | 71 |
| 6.1 | Discussion | 71 |
| 6.2 | Conclusion | 72 |
| 6.3 | Academic Implications | 73 |

| | | |
|-----|------------------------------------|-----|
| 6.4 | Practical Implications..... | 74 |
| 6.5 | Limitations | 75 |
| 6.6 | Future Research..... | 76 |
| 7 | BIBLIOGRAPHY | 79 |
| 8 | APPENDIX | 89 |
| 8.1 | Questionnaire | 89 |
| 8.2 | Messages from the video..... | 96 |
| 8.3 | Message Delay Calculation..... | 99 |
| 8.4 | Extended Statistical Results | 100 |

LIST OF FIGURES

| | |
|---|----|
| Figure 1: Schematical depiction of the Turing Test. From: Pinar Saygin et al., 2001 | 10 |
| Figure 2: Conceptual Model. | 41 |
| Figure 3: Amended Conceptual Model for Regression purposes. | 63 |

LIST OF TABLES

| | |
|--|----|
| Table 1: Table elaborating on the terms HHI, CMC and HCI. Freely based on: Appel, von der Pütten, Krämer, & Gratch (2012) and Spears & Lea (1992). | 13 |
| Table 2: Schematical overview of the origin of related theories (University of Twente, 2019)..... | 14 |
| Table 3: Table of comparable research. | 19 |
| Table 4: Used constructs in comparable research. | 26 |
| Table 5: Social Presence Taxonomy Table..... | 30 |
| Table 6: Perceived Humanness Taxonomy Table. | 31 |
| Table 7: Service Encounter Satisfaction Taxonomy Table..... | 33 |
| Table 8: Social Information Processing Theory Taxonomy Table..... | 35 |
| Table 9: Required mean sequence of the manipulation check question for manipulation success..... | 47 |
| Table 10: Schematical design of the experiment conditions, groups and definition. | 49 |
| Table 11: Formula to calculate the dynamical message delay. All figures represent seconds. | 50 |
| Table 12: Means Comparison Pre-test 1..... | 51 |
| Table 13: Means Comparison Pre-test 2..... | 52 |
| Table 14: Overview of the used input parameters for power analysis..... | 53 |
| Table 15: Overview of the used measures and detailed information..... | 54 |
| Table 16: Overview of distribution methods and target groups for the experiment...60 | |
| Table 17: Manipulation Check Means Comparison Main Experiment. | 61 |
| Table 18: Manipulation Check One-Way Anova Table. | 61 |

| | |
|--|----|
| Table 19: Calculated Cronbach's Alpha for all the constructs. | 61 |
| Table 20: Anova table for the Social Presence Construct..... | 61 |
| Table 21: Anova table for the Perceived Humanness Construct. | 62 |
| Table 22: Anova table for the Service Encounter Satisfaction Construct. | 62 |
| Table 23: Mean overview Social Presence Construct..... | 62 |
| Table 24: Mean overview Perceived Humanness Construct. | 62 |
| Table 25: Mean overview Service Encounter Satisfaction. | 62 |
| Table 26: Coefficients Table for simple regression. Dependent variable: Service Encounter Satisfaction..... | 64 |
| Table 27: Regression model summary for Social Presence..... | 65 |
| Table 28: Regression model for Perceived Humanness. | 65 |
| Table 29: Regression model summary for Perceived Humanness..... | 66 |
| Table 30: Regression model for Perceived Humanness. | 66 |
| Table 31: Regression model summary for Service Encounter Satisfaction..... | 67 |
| Table 32: Regression model for Service Encounter Satisfaction..... | 67 |
| Table 33: Hypothesis testing process for hypothesis 1..... | 69 |
| Table 34: Hypothesis testing process for hypothesis 2..... | 69 |
| Table 35: Hypothesis testing process for hypothesis 3..... | 69 |

1 INTRODUCTION

The terms *chatbot*, *bot*, *conversational agent*, and *conversational user interface* will be used interchangeably in this research for readability purposes. Denotation will, however, be identical and the same throughout this thesis.

1.1 Introduction

Chatbots are the next big thing (Gartner, 2017). The technological institute Gartner (2011) has predicted that in the future, interaction between companies and customers will be for 85% automated, and without human interaction. Research conducted by Aspect (2016) says that 49% of the consumers prefer the use of chat and messaging channels to perform customer service interactions with companies, if the company would implement it properly. The same study acknowledges that 69% of the consumers interacts with a conversational agent at least once a month. Especially the younger generation quickly gets acquainted with chatbots, and sees the advantages for their use (Salesforce, 2018).

Looking at it from the end-users perspective, we see a positive future outlook. The demand for these conversational agents will be growing during the coming years. If we switch to the technical perspective, we also see a bright future. Technology is emerging fast, and designing and developing a chatbot becomes less complex (Moore, Arar, Ren, & Szymanski, 2017). Chatbots can be created faster and easier, due to the fact that low-coding platforms are developed (KPMG, 2019). This requires less technical expertise for a successful chatbot to be built. Big players in the market, such as IBM and Microsoft, with their respective products Watson and LUIS, are becoming pioneers in this segment, and their technologies can be regarded as leading in the market (Mind Bowser, 2017).

For companies, the use of a chatbot brings several advantages. A case study performed by Scheepers, Lacity, & Willcocks (2018) at a digital university shows that the implementation of a chatbot can greatly enhance experiences from staff, customers and the institution itself. This study identified widely ranging advantages, including: cost reduction, freeing up staff for more critical tasks, improved competitive positioning, better service delivery and better content governance. Side note is that these advantages occur in well executed projects, and no auspicious future needs to be expected straightaway. Early adopters need to have realistic

expectations, because these bot platforms need training and the creation of them involves effort and time (Davenport, 2015).

1.2 Promising Areas

Chatbots offer serious advantages to companies and users, but not in every sector will these advantages have the same flourishing effect. Different sectors respond differently on this new technology. However, some very promising areas can be highlighted, wherein a prosperous future is expected.

This prosperous future is firstly depending on the role. A chatbot can take many roles and forms, as reflected in the wide spectrum they are active at this moment. PwC Digital Services (2017) show that consumers mainly see a chatbot as an adviser, and this role is expected to grow in within the coming five years. This view is confirmed in the use case from Deakin University, where a chatbot was used as a help desk for answering student's questions (Scheepers et al., 2018). A taxonomy created by Robinson, Gray, Cowley, & Tan (2017) shows three main types of chatbots, to be: informational, transactional and advisory. The first is mainly designed as a FAQ chatbot, capable of answering simple questions based on a frequently asked questions database. This form requires the lowest form of intelligence, and is therefore relative simple to create. However, proper data governance, in the form of a single source of truth database is necessary, as noted in the Deakin University case study (Scheepers et al., 2018).

An informational, or FAQ, chatbot is for now the most widely available. This is directly linked to the promising areas for these bots. The ideal places for a chatbot are areas where many, relatively simple interactions take place, in a questions-answer form. Research by Srinivasan, Nguyen, & Tanguturi (2018) confirms this, as they show after sales, customer service, marketing and sales as such areas for chatbots to add value.

1.3 Current State of Literature

Chatbot research can generally be divided in two parts: the technical part and the interaction part. The research related to the technological part is the most all-encompassing. However this is the least relevant for this thesis, it still shows that technology is advancing fast. New technologies in terms of artificial intelligence and machine learning continuously improve the text understanding and natural language processing (Nimavat & Champaneria, 2017). Since technology and knowledge is mostly centralized by a few large players, advancement is quick.

Subjects in this technical research field come down to the following architectural topics, as identified by Cahn (2017): natural language processing, including dialogue recognition and intent identification; Response generation, which could be, among others, rule-based, information retrieval or machine learning; knowledge base creation, which involves the creation of a corpora existing of usable data; and dialogue management, which include strategies for human imitation and communication strategies. For virtual assistants which use speech, also speech-to-text conversion and text-to-speech conversion are included. As the pioneers in the chatbot industry, such as IBM and Microsoft, keep innovating, scientific literature follows closely to measure the impact and possibilities this brings.

But not only technological factors determine the impact chatbots have. An evenly important factor is the interaction between users and chatbots (Qiu & Benbasat, 2009). The main reason for this is the fact that, even though people know they are interacting with a chatbot, they still apply the social rules and expectations from normal human to human conversations in their interaction with chatbots that use natural language or show their human characteristics (Nass, Steuer, & Tauber, 1994). Thus, if we see a chatbot as a social actor, we must also apply appropriate theories and research.

In research, these factors influencing interaction have been neglected for long time (Brandtzæg & Følstad, 2016). Research in this area mainly comes down to the factors that influence the human-computer interaction. Not all factors have been researched yet, and in this field, still a lot of work is yet to be done. Some remarkable research in this area has however given helpful insight in the topics of: chatbot response time (Gnewuch, Morana, Adam, & Maedche, 2018), Communication style (Verhagen, van Nes, Feldberg, & van Dolen, 2014), degree of interactivity (Schroeder & Schroeder, 2018; Schuetzler, Grimes, Giboney, & Buckman, 2014), perceived agency (Appel, von der Pütten, Krämer, & Gratch, 2012), presence of virtual character (Von Der Pütten, Krämer, Gratch, & Kang, 2010), smiling (Verhagen et al., 2014), tone awareness (Hu et al., 2018a), and the use of different typefaces (Candello,

Pinhanez, & Figueiredo, 2017). This thesis anticipates on adding knowledge and improve the progress of research in this specific area.

1.4 Research Gap

Chatbots are growing vastly, but only little is known to the way people react to their use. Technological abilities are expanding, but the knowledge about the interaction with a chatbot is not expanding coequally. Building a successful chatbot depends on both these factors, technological and social. This means that more knowledge is needed on interaction with chatbots.

Many people are unaware of the fact that they could be talking to software, when they expect a real human being on the other side. What would the user's reaction to this be, and how does it influence the conversation if the user is in doubt? Is it important to know on beforehand if you are talking to a chatbot or a human being? And what factors are influencing the answer to this question? Is there a possibility that this influences the engagement with the chatbot, user satisfaction, or perceived competence of the agent? These questions are all very relevant, as we see chatbots rising more and more.

As indicated in the current state of literature paragraph, many factors involving the human-computer interaction are determinant for the eventual satisfaction and evaluation of the conversation. It is also shown that research on these specific factors lags behind (Følstad & Brandtzæg, 2017). This opens a research gap in literature. As technology improves quickly, over time, it will become harder to tell if we are talking to a chatbot or an actual human being. The impact of this change on end users is of utmost importance, since it directly affects the user satisfaction and thus user adoption of chatbot experience. Without the insights on the interactive part, the technical developments for chatbots are in vain, since a successful chatbot depends on both parts. This research aims to understand the attitude towards chatbots, whenever they do or do not introduce themselves as being a chatbot.

1.5 Research Question

Research performed in this thesis will lead to the answer to the following research question:

Which chatbot introduction yields the highest user satisfaction?

1.6 Sub Questions

Sub questions are created in order to aid in a structured answering of the main research question. The following sub questions have been identified:

1. What is a chatbot?
2. What relevant theories concerning chatbots are available?
3. What measures are suitable for measuring the impact of a chatbot's introduction?
4. Which options are available for a chatbot to introduce itself?

1.7 Research Design

This thesis aims to unravel the impact of knowing a chatbot's identity on the user satisfaction over the held conversation. It will use these various introductions as dependent variable, and measure the effect with relevant measures. This research will be conducted via an online experiment, where participants will be exposed to various chatbot environments. Results in the form of participants' opinions will be gathered afterwards, in the form of a questionnaire. Also other demographic information will be asked, such as age, education level and previous experiences with similar technology. These will be used as controlling variables in the research.

This research is in the form of a survey experiment, making use of a vignette approach (Atzmüller & Steiner, 2010; Gaines, Kuklinski, & Quirk, 2007). This enables participants to be exposed to a manipulation variable and afterwards expressing their opinion and intentions in a survey form.

This thesis will start with a literature research. Key theories and constructs available in current scientific literature will be highlighted and their impact on this research will be

1. Introduction

explained. All the concepts used in this research will be thoroughly explained and related research will be touched upon.

For a reliable outcome, the survey needs a wide diversity of respondents to test for all the different factors which have influence. The participants will be led to a test environment, in which they will randomly be assigned to a testing group. Participants will be asked to read a chatbot conversation, and afterwards answer questions on the quality and satisfaction of the conversation, and the rating of the bot.

Statistical methods will determine the impact of the various cultural and personal factors on the rating of the conversational partner, depending on the statement of the identity of the agent.

1.8 Relevance

Chatbot related technology is advancing fast. It is to be expected that in the near future chatbots are able to perfectly imitate human beings in text chats (Vlek, 2014). From that moment onwards, users are unable to tell if they are talking to a chatbot or a human being. Introduction will then start to play an important role in chatbot interaction. Already studying on the effect of various introductions will provide for a decent research baseline for the future. Both for academic as for practical purposes this thesis extracts relevance, which is further elaborated upon in their respective paragraphs below.

1.8.1 Academic Relevance

Knowing the effect of various introductions by chatbots to users will provide a solid research ground which is important for now and in the near future. For now, this thesis will provide an important insight in the research to chatbots. It will improve the research funnel towards chatbot interactions and will add value to knowing the impact of various introductions. It will enhance the satisfaction of end users on chatbot interaction.

But also in the future is this thesis of value. The final result of this thesis will reinforce future research on this topic, and provide necessary stepping stones for advances in literature. This is reflected by the literature research this thesis executed, which provides an overview of the available theories and research and summarizes the current position of scientific literature.

1.8.2 Practical Relevance

Practical relevance from this thesis is mainly extracted for designers and creators of chatbots. This thesis will provide practical advice based on scientific literature to enhance chatbots. Programmers in the field of chatbots are currently unaware of the best way to let a chatbot introduce itself to the user they are interacting with, since no existing scientific literature provides these answers. Presumably they improvise their current introductions, but with the final result of this thesis they will be provided with a solid scientific base to build their chatbots on.

Especially for the currently emerging low-coding chatbot design platforms the result from this thesis will be valuable. For example the currently developed Digital Advisor platform, designed by KPMG (KPMG, 2019).

1. Introduction

2 LITERATURE RESEARCH

The main goal of this second chapter is to answer the first two sub questions stated in chapter one: explaining to what a chatbot is, and what relevant theories concerning chatbots are available.

In this chapter all relevant literature and research will be elaborated upon, in order to explain and clarify all the constructs, terms, theories and other important factors relevant for the execution of this research. It will retrieve already performed research which is relevant to the substantiation of the research questions and assumptions which underlie this thesis's research design.

The structure will be as follows. The first part will elaborate on what literature defines as a chatbot. What do we refer to if we use this term, and what conditions need to be met in order to fulfill to this term. The second part will elaborate on the interaction possibilities with chatbots. Which theories are on the basis of this interaction, and in what discipline do these theories find their basis. The third part will take a look at comparable research. It mainly questions which factors influence this interaction. In a table form it will be presented how similar research has been conducted, and will show the various constructs, methodologies and outcomes from these articles.

2.1 Turing Test

Chatbots are getting smarter and smarter. The technique is promising, and quickly developing. But this increasing popularity also comes with a downside. Where the business demand increases, the IT departments who are responsible for building these chatbots, become overloaded, and sometimes lack the proper skills to execute these requests (KPMG, 2019).

Mechanical intelligence has already been the precursor of artificial intelligence for a long time. It was already studied in the 1940s by, among others, Alan Turing. He studied the mathematical implications of artificial intelligence, but also the implications on social level, the element of interaction (Cooper & Leeuwen, 2013). Alan Turing came up with the Turing Test, which is still very actual today (Turing, 1950). The test is as follows, as written by Pinar Saygin, Cicekli, & Akman (2001): an interaction between a human and two other entities, one is a human and the other is a computer. Both try to impersonate a human, and the only way they have contact is through written text. After a given time, the human as to make a statement

2. Literature Research

on the identity of the entities he or she is interacting with. This test is schematically depicted in Figure 1.

In the test, a human person, the tester, has a chat conversation with both a machine and another human being. The tester does not know who the other human being is, and who the machine is. After a given length of time, and having held an extended conversation, the tester is asked to tell which conversation was held with the machine, and which conversation was held with the other human being. If the tester chooses wrong, or cannot tell the difference, the machine passes the test. At that point, the machine identically mimics a human, and thus is indistinguishable from a real human being.

So far, no computer program has really passed the Turing Test. One program though claims to have passed, but there are still doubts about the actual setup of the experiment (Vlek, 2014). Especially doubt is about the actual length of the conversation, as Turing (1950) does not prescribe a specific length in his test layout. The shorter the test is, the easier it is for the machine to pass.

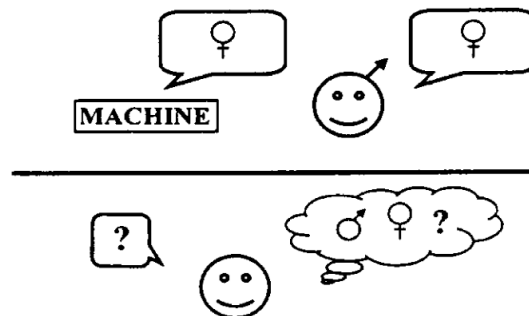


Figure 1: Schematic depiction of the Turing Test. From: Pinar Saygin et al., 2001

So, why is this important for this thesis? The Turing Test can be seen as the last boundary between man and computer. Right now it is fairly easy to know whether we are talking to a bot or a human, everyone who ever interacted with a chatbot can tell.¹ Up until now, technical flaws give away the real identity of the conversational partner. Examples as bad grammar and misunderstanding of figurative way of speech reveal the bot. But as Vlek (2014) shows, we are currently in a period where, with the current techniques, passing the Turing Test will be

¹ Websites to interact with a chatbot are:

Eliza: <https://www.masswerk.at/elizabot/> One of the earliest chatbots, founded in 1960s.

Mitsuku: <https://www.pandorabots.com/mitsuku/> Currently regarded as the best chatbot, closest to human interaction.

imminent. If we combine this with insights of the rapid development of artificial intelligence, it will only be a matter of time before chatbots start to pass for the Turing Test. Large scale passing of this test by software will mean that it is impossible to tell whether we are talking to a human or a chatbot. This will create a new era in text based interaction, where computers, who are impersonating humans, are indistinguishable from real humans.

2.2 What is a Chatbot?

Conversational agents are software agencies which interact with humans on the most natural basis possible (Mujeeb, Hafeez, & Arshad, 2017). Literature has given this term a clear demarcation, however some inconsistencies between articles remain. The first important factor is to know the difference between a chatbot and any other software program. Franklin & Graesser (2005) have proposed a taxonomy which gives insight in this difference. Their essay compares various literature sources and weighs of the different definitions currently existing. The main properties identified are the fact that an agent is: reactive, thus able to respond timely to changes in the environment; autonomous, meaning having control over its own behavior; goal-oriented, so not just a simple environment responsive design; and temporally continuous, meaning that the process runs constantly (Franklin & Graesser, 2005).

Conversational Agents make part of this taxonomy, and fall in the specific group of agents being able to communicate to humans. They communicate with human beings in a way that mimics the human-human interaction in the best way possible (Mujeeb et al., 2017). Even within this group of communicative agents, subcategories exist. There are for example varieties in the way we communicate, either by speech or written text, but also varieties in the objectives, which ranges from asking a question, to just having a small talk conversation.

But also differences in agency exist. Research by Appel, von der Pütten, Krämer, & Gratch (2012) has given insights in the effect of perceived agency on the social behavior. They identify various types of agency, such as chatbots with a virtual character, chatbots with just a profile picture, or just an agent, which converses only with chat text. This degree of humanness in the agency has impact on the way people interact with a chatbot, and the amount of social cues can be send and picked up.

This thesis mainly focusses on the text-based chatbots, and does not necessarily focus on chatbots involving speech interaction or avatars. However, the research done into these kinds of chatbots will be useful for construction the theoretical framework for this thesis.

2.3 Theories on Chatbot Interaction

Chatbot interaction combines two research disciplines which are normally separated from each other. Chatbot interaction namely involves interaction between a human being and an autonomous computer, using natural language (Dale, 2016; McTear, Callejas, & Griol, 2016). Social and communication sciences are focused at communication between human beings, in a social way, but not interaction between man and computer. Computer science on the contrary is focused at communication between computers, but this normally involves technical communication protocols, such as internet. Computer science also inhabits interaction between humans and computers, but expresses this mainly in user interfaces, and not in social interactions which make use of natural language, as is the case with chatbots.

This raises the question where to look for relevant theories: computer science, or social sciences. But this also raises the question where to find theories about interaction between human and computers in a social way. There is no single research discipline covering all the aspects of chatbot interaction. But also due to the novelty of this topic, the combination between these different research disciplines is relatively uncommon.

This paragraph explains the theories which have the highest relevance for this research, and describes two main paradigms which are critical for the combination of these different research disciplines.

The important basic theories underlying the way we use chatbots, are traced back to social sciences and communication sciences. These two are the hosting research disciplines of these theories. In this research discipline, various theories have emerged, which are, for example, related to organizational communication, communication processes and interpersonal communication and relations (University of Twente, 2019). One main cluster within this research discipline is of importance: communication and information technology.

Within this cluster the CMC related theories are found. CMC stands for computer mediated communication, and involves the communication between persons, via a computer, and not with the computer itself (Spears & Lea, 1992). For example this includes the way we use Skype; we communicate with another person, but this communication is mediated by the computer software created by Microsoft.

HCI is on the other side of this spectrum. It stands for human-computer interaction, and is a research field in which many relevant research has been done for this thesis. It mainly compares the interaction that takes place between a human and an autonomous software agent, in

2. Literature Research

comparison to the interaction between two or more human entities, so called human-human interaction (HHI)(Appel et al., 2012).

These three theories create a spectrum which varies in the way communication takes place. HHI is interaction between two humans, in a face to face setting. CMC is also a form of communication between humans, but communication is mediated by computer software. Finally HCI is an interaction between a human and a computer program.

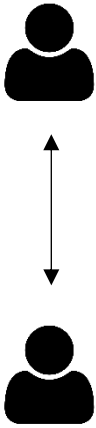


| HHI | CMC | HCI |
|--|--|--|
| Human Human Interaction | Computer Mediated Communication | Human Computer Interaction |
|  |  |  |
| Social Orientation | | Technical Orientation |

Table 1: Table elaborating on the terms HHI, CMC and HCI. Freely based on: Appel, von der Pütten, Krämer, & Gratch (2012) and Spears & Lea (1992).

2. Literature Research

| Research Discipline | Research Area | Cluster | Theories |
|---------------------|-----------------------|--|---------------------------------|
| Social Sciences | Communication studies | Communication and Information Technology | Computer Mediated Communication |
| | | | Social Presence Theory |
| | | | Media Richness Theory |
| | | | Reduced Social Cues Theory |
| | | | Social Response Theory |

Table 2: Schematical overview of the origin of related theories (University of Twente, 2019).

This paragraph shows the related theories in the discipline of social sciences. But so far, no connection has been made to their use in computer interaction. A research gap remains between the applications of communication theories in computer science. This gap must be bridged, before we can use theories based on human-human interaction, and apply them on human-computer interaction.

The bridging of this gap starts with the research of Nass et al. (1994). Their initial research showed that interaction between a human and a computer shows similarities with interaction between humans. These similarities appear in both their original research (Nass et al., 1994) as well as in their follow-up research (Nass & Moon, 2000), which both complete the paradigm.

This research entails the confirmation that social sciences theory are applicable to human-computer interaction. Their methodology covered taking a social science theory and replacing the word human with computer. After this, with experimental methods, they verified the results and noticed a clear match in outcome with the original theory. They argue therefore that computers are seen as social actors. This CASA paradigm, which stands for Computers As Social Actors, has been widely adopted after.

Their research included, by conducting several experiments, that factors such as gender stereotypes and politeness, are all applied to computer in a social environment. Follow-up research performed by Nass & Moon (2000) confirmed these findings and extended these factors with reciprocity, in group versus outgroup and ethnicity.

Both these researches tell us that we should see computers as being social actors, just like any other human. Therefore we are able to apply social and behavioral scientific theories also to a chatbot and other HCI situations, as confirmed by research from Tourangeau, Couper, & Steiger (2003). This research field is particularly interesting for this thesis. The work shows that social theories from social sciences are applicable to computer interaction, and thus chatbot conversations.

2.4 Limitations of the CASA paradigm concerning Chatbot interaction

As the previous paragraph explained, the CASA paradigm, and the underlying research, has proved that theories from social sciences, which are valid for human-human interaction, are also applicable to human-computer interaction (Nass et al., 1994). This however does not mean that interaction happens on exactly the same way. Basic theories and constructs are interchangeably applicable, but details in the interaction differ. This paragraph continues on the paradigm described in paragraph 2.3, but shows that this paradigm is not unifiable applicable, and some details differ.

Research by Pearson, Hu, Branigan, Pickering, & Nass (2006) shows that users expectations and beliefs of a system influences the way people interact. For example changing their use of language. This effect is also seen by the study of Hill, Randolph Ford, & Farreras (2015), which shows that people send more messages to chatbots than to human conversational partners, but that the messages sent contain considerably less words. These results were obtained by analyzing conversations from both chatbot as human interaction, while people were fully aware with whom or what they were chatting.

But, since not all people are the same, nor is the way they interact, or wish to interact with a conversational agent. Various factors influence this, and are, among others, related to people's age (Heerink, Kröse, Evers, & Wielinga, 2006), social generation with which they identify themselves (Salesforce, 2018), profession (PwC Digital Services, 2017), all in combination with the different purpose of the agent itself (Bickmore, Caruso, & Clough-Gorr, 2005).

But also the chatbot itself is important. Research done by Iacobelli & Cassell (2007) shows that ethnic identity in an embodied conversational agent influences the engagement with that specific bot. This confirms the thought that the chatbot itself influences the user experience.

This is in line with the growing trend of so called embodied conversational agents. Isbister & Doyle (2002) define an embodied conversational agent as a program or software which represents a human being. Basically a chatbot which pretends to be a human being, by showing characteristics from a human being. This could either be in the form of showing facial expressions in an animated face, or showing emotions and personality (Cassell, 2000).

These chatbots are considered to be part of the Web 2.0 technologies, in an trend of moving toward Enterprise 2.0 (Seo & Rietsema, 2010). These concepts are already widely adopted by various companies, especially in web care environments. Examples are the virtual assistants of various online web shops, such as Bol.com and Coolblue.com.

Since chatbots are using the common channels for inter-human contact, such as WhatsApp and other messenger channels, and are entering the sphere of the personal domain (Nimavat & Champaneria, 2017), it becomes harder to tell if we are talking to a real human being, or a robot. This is fueled by the fact that technology is improving, and that it becomes harder to tell with what or whom you are actually talking. Neuroscientific research performed by Ciechanowski, Przegalinska, Magnuski, & Gloor (2019) showed that the more we see a chatbot as human, the more competent we believe it is. This is directly linked to the expectancy we have of the identity of a conversational partner – being a human, or a chatbot.

2.5 Related research in Chatbot Communication

The previous paragraphs have been an introduction to chatbot interaction theory. These sections showed which basic theories are underlying the use of chatbots, and that theories from social sciences are uniformly applicable to human computer interaction, by using the CASA paradigm. This next section looks at the appliance of these theories in actual research. By means of a table, it shows research which is similar to the research approach from this thesis. The table gives insight in subject, method, measures, used constructs, and results of this similar research.

The goal of this second part of the literature review is to provide an overview of relevant research which has already been conducted in this area. This overview also forms the basis of the next chapter, which builds a theoretical framework based on this table.

The overview gives information about the used theories, constructs and methodologies that can be used, and more importantly, how they can be practically used in research. It forms the translation between the theories described in the first part of this literature review, and the actual use in research. Value is extracted from comparing various research approaches and comparing the usefulness of the methodologies, theories and constructs for this thesis.

The reason to use a table for this overview is because a table format gives a clear overview, in a structured way, which makes similarities and differences visible. The table should therefore be seen as a depiction of the current research landscape, highlighting the various research areas and publications.

The table on the following page shows all relevant research for this thesis in a structured approach. The various columns show the most important information about each of the article.

2. Literature Research

The table starts by identifying the article based on the author and publication year. Thereafter it shows the basic setup of the research, namely which factor in chatbot interaction it researches, and the according research question. Also, if an article focusses on a specific sector, this is displayed. This is mostly important for the generalizability of the results and show a potential demarcation for using the research results in other sectors.

The following columns focus on the research design and approach. They show which theory has been used, and which constructs appeared in the research. Also the methodology is shown. It shows how the research is done, which research design is used, and how the results are gathered. The final columns shows briefly the results of the research.

The previous sentences showed information about the outline of the table. A discussion about the content of the table – the research itself – can be found underneath the table.

2. Literature Research

| | Authors, year | Factor | Research Question | Context | Theory | Construct | Methodology ² | Conclusion |
|---|-----------------------|---|--|---------------------|--|---|--|--|
| 1 | Gnewuch et al., 2018 | Response Time. | “How do dynamically delayed responses affect users’ perception of a customer service chatbot as compared to near-instant responses?” ³ | N/S ⁴ | Social Presence theory; Media Synchronization Theory; Social Response Theory. | Social Presence; Perceived Humanness; Service Encounter Satisfaction. | Survey Experiment; Structured interaction. | “Dynamically delayed responses positively affect users’ perception of chatbots.” |
| 2 | Verhagen et al., 2014 | Friendliness; Expertise; Smiling. | The effect of friendliness, expertise and smiling on Social Presence, Personalization and Service Encounter Satisfaction. | N/S | Implicit personality; Social Response; Emotional Contagion; Social Interaction; Personalization. | Social Presence; Personalization; Service Encounter Satisfaction. | Survey Experiment; Structured interaction. | “[...] evaluation of an agent’s friendliness and expertise elicits social presence and personalization and in turn, social presence and personalization have a strong effect on service encounter satisfaction.” |
| 3 | Candello et al., 2017 | Typefaces. | RQ1: “Are machine-like typefaces (such as OCR) more perceived as machines in a chat?” RQ2: “Are typefaces which mimic human handwriting (such as Bradley) more perceived as human in a chat?” | Financial Services. | HCI; Language Processing Theory. | Perceived Humanness. | Survey Experiment; Vignette. | RQ1: “Yes, machine-like typefaces (such as OCR) are more perceived as machines in a chat.” RQ2: “No, typefaces that mimic human handwriting (such as Bradley) are not more perceived as human in a chat.” |
| 4 | Hu et al., 2018 | Tone-Awareness. | What is the effect of a tone-awareness chatbot on user experience? | Customer Care. | HCI; Customer Satisfaction; User attitude. | User Experience. | External data analysis. | “A tone-aware chatbot generates as appropriate responses to user requests as human agents.” |
| 5 | Appel et al., 2012 | Agency (virtual/embodied); number of social cues. | Analyzing the importance of social cues and perceived agency of a computer system of the emergence of social reactions during human computer interaction. | N/S | HCI; CASA ⁵ ; Social Presence; Social Response. | Person Perception; Social Presence; Rapport factors. | Survey Experiment; Structured interaction. Conversation (meta) data. | “Subjects in the virtual character conditions (high number of social cues) had a stronger feeling of social presence [...] than subjects in the text conditions.” |

2. Literature Research

| | Authors, year | Factor | Research Question | Context | Theory | Construct | Methodology ² | Conclusion |
|---|--|--|---|---------|--|--|--|---|
| 6 | Schuetzler et al., 2014 | Chatbot Dynamicity; Message veracity. | RQ1: "How does a chat bot that provides dynamic, rather than static, responses influence user perceptions, responses, and behavior in a real-time chat environment?" RQ2: "How does a chat bot that provides dynamic, rather than static, responses influence user behavior when engaging in deception?" | N/S | Social Presence; Deception. | Perceived Humanness; Perceived Engagement; Response Latency. | Survey Experiment; Structured interaction; Conversation (meta) data. | RQ1: "The research found that people perceive a dynamic chat bot to be both more engaging and more human-like than a static interview." RQ2: "When in a deception situation, people interacting with a static chatbot alter their behavior, while with a dynamic chatbot response behavior is more natural." |
| 7 | Schroeder & Schroeder, 2018 | User/machine Modality (talk/type); Gender. | The effect of the mode of interaction on willingness to share personal information with machines. | N/S | Human Uniqueness scale. | Anthropomorphism; Behavioral Trust. | Survey Experiment; Structured interaction. | Speech as user expression modality yields higher trust. Response modality and gender gave no effect. |
| 8 | Gourov, 2019 | Sentiment orientation; Personalization. | "What characteristics should a chatbot contain in order to affect the User Experience?" | N/S | User gratification theory; Technology readiness. | Effectiveness; Efficiency; User satisfaction; Novelty. | Survey Experiment; Vignette. | Users' perception on aggregated user experience do not depend on the different designs of chatbot in the research. |
| 9 | Murgia, Janssens, Demeyer, & Vasilescu, 2016 | Agency (Human/Machine). | "To what extent a bot can emulate a human on a question and answer website and what feedback it receives." | FAQ | N/S | Reputation Rate. | Experiment, Free interaction; Conversation (meta) data. | "Humans do not completely trust suggestions provided by a machine; or they have such high expectations from machine that its answers have to be significantly better than the ones provided by a human." |

Table 3: Table of comparable research.

2. Literature Research

² For summarizing reasons, these terms are appointed by the author of this thesis based on the expressions made in the article.

³ Text between quotation marks are literally cited from the article. Non-quoted text is summarized by the author of this thesis, based on the article.

⁴ N/S (not specified) means that information in this cell is not directly and literally expressed in the article, and therefore nonexistent in the table.

⁵ Computers are Social Actors framework, designed by: Nass et al. (1994).

2. Literature Research

Table 3 on the previous pages shows the various research that has already been conducted in this area. This section will elaborate on the contents of this table. It will discuss the various research that have showed up in this table and debate their choices.

2.5.1 Criteria for eligibility as comparative research

First of all the main criteria for research ending up in this table. This choice is based on a relation between the dependent and independent variables, which is similar to this thesis's research approach. All research in the table manipulates a factor in chatbot experience, such as reaction time (Gnewuch et al., 2018) or font type (Candello et al., 2017). The similarities in this research are also reflected in the measurement of the dependent variable. So, all research in Table 3 changes a factor in chatbot experience, and measures the effect of this change in a construct related to the interaction experience. Not all research focusses on the same construct, differences for example appear from the degree of anthropomorphism (Schroeder & Schroeder, 2018) to a construct involving effectiveness (Gourov, 2019). Consistent factor however remains the manipulation of a factor influencing the interaction, and measuring the effect of this manipulation on this interaction.

2.5.2 The used constructs

In general we can see three different constructs which occur frequently: social presence, perceived humanness and an aggregated form of user experience. This latter has a wide interpretation in literature, and therefore is measured on a different scale and based on different theories. Examples from the table show names as user experience (Hu et al., 2018b), user satisfaction (Gourov, 2019) and perceived engagement (Schuetzler et al., 2014). In practice, the meaning of these different constructs are relatively close together.

In smaller form, something similar can be seen in the constructs of perceived humanness and anthropomorphism, which in both cases measured the amount of humanness shown by the conversational partner.

A more in depth analysis of the various constructs will take place in the next section.

2.5.3 Comparable Methodology

On terms of methodology, research design and data collection (in the table: methodology) we see also quite similar results. Most research has pursued an experimental setup, wherein they have set up different groups in which different treatments were tested. In the table this is referred to as: survey experiment. All research, except for the one testing on tone-awareness (Hu et al., 2018b) have used this experimental design.

The approaches in experimenting were also similar. Researches have set up different treatment groups, depending on the amount of constructs tested. In the table this ranges from only one treatment group, up to eight treatment groups. All this research, except for the one performed by Murgia et al. (2016) has used randomization to determine the placement in the different groups, but made sure control factors, such as gender, were equally spread over the groups. As said, the research by Murgia et al. (2016) has used a different approach, and had the two treatment groups running consecutively to each other, and not simultaneously. Their research involved a live chatbot for people to use, therefore this research is also the only research which allowed free, thus unstructured, interaction with the chatbot. Despite the highly interesting research scope, relevant for this thesis, the experiment setup used in this research is lacking the required robustness. This includes the presence of various biases which are unaccounted for, and an early termination of the experiment by the hosting provider.

Most other research in the table has made use of a structured interaction with a chatbot. People knew they were participating in a research, and interaction took place in a controlled setting. Structured interaction meant a given starting situation before interaction took place. For example an explanation to the participant that their telephone bill was too high, and that their goal was to gain information and buy a new phone subscription, after which chatbot interaction started (Gnewuch et al., 2018).

Some research also used the vignette approach (Candello et al., 2017; Gourov, 2019). In this kind of research a short excerpt from a chatbot interaction was shown, either in static form or in a video. Participants were invited to experience this shown conversation as if they were the one having the conversation. This research design lead to the elimination of many biases, because the conversations were set and very structured. This way only the effect of varying the independent variable between the groups could be measured very closely.

Almost all research in the table also used a questionnaire as data collection method. This is because the constructs were depending on the user's opinion. Only few research used the data created directly from the conversation, referred to in the table as: conversation (meta) data. In

2. Literature Research

no research this data was used as a single source to draw conclusions on; it was only used together with questionnaire data. Examples of this form of data were the length and amount of messages sent (Schuetzler et al., 2014), the amount of unfinished sentences or the used pause fillers, such as ‘ehm’ and ‘uhh’ (Appel et al., 2012).

The research performed by Hu et al. (2018a) is the only research who used a specific data collection method. They collected historical data from real chatbot conversations and analyzed these. In the table this is referred to as external data analysis. Their research was not done in a lab setting, and the participants who were having the conversations, did not know their data was going to be used in research later on.

Altogether, the most occurring methodology is the survey experiment. This is a research form which combines an experiment with a survey afterwards (Gaines et al., 2007).

2.5.4 Context

Much research has focused on the general use of chatbots, this is displayed in Table 3 as N/S – non-specified – in the context column. Only few research has specifically focused themselves on a sector, such as customer care of FAQ chatbots. This, however, does not mean that the result provided by research not focused on a sector, is not generally applicable to chatbots in other contexts. This is certainly possible. The context only explains the research environment in which the experiment took place.

2.5.5 Implications for this research

Table 3 shows the various factors influencing chatbot experience. Nonetheless, the changes in the independent variables are all in the domain of chatbot-human interaction. This means that in all cases only the impact of changes in the chatbot settings were measured. No research has examined the changes between the domain of chatbot-human interaction and human-human interaction. So no research investigated the differences between a chatbot talking to a human, or a human talking to a human. Research on this area is still in an early stage, and no comparison research is yet present in scientific literature. As an exception on this we look at the research from Murgia et al. (2016). It measures the effect of a chatbot talking to a human, and a chatbot

2. Literature Research

proposing to be a human, talking to a human. This is not the same, but for research design purposes, the similarities do occur.

This means though that no research has been conducted which uses a human-human interaction either as treatment or as control group. In a certain way, this thesis will conduct research in uncharted territory. The research exposed in Table 3 is of very good use to see various options in research design related to constructs, variables, and design, but no such specific research has preceded this thesis.

3 THEORETICAL FRAMEWORK

This chapter of the thesis builds on the previous literature review, and makes a profound analysis of this theory, in order to turn it into a usable framework. The goal of this chapter is to answer the second two sub questions. It answers on how to measure the impact of a chatbot introduction, and which options are available for a chatbot to introduce itself.

3.1 Used Constructs in Literature

In order to give more concrete insights on the approaches of these comparable articles, the next table, Table 4, has been created.

This table transforms the exact same articles as used in Table 3, but aligns them according to the used constructs. The columns respectively then show the articles in which the construct was used, the measurement of this construct, the analysis that has been used, and the effect for this thesis's research approach.

The next section will make clear on which constructs in literature this thesis will focus itself on. This choice is based on the effect of these constructs in other research. In order to get a clear overview of these constructs, the next table has been created.

This table is important because based on this table, the usability of various constructs will be measured, which are suitable for answering this thesis' research question.

3. Theoretical Framework

| | Construct | Authors, year | Measures | Analysis | Effect in this thesis |
|---|--------------------------------|--------------------------|-------------------------------------|--|-----------------------|
| a | Social Presence | Gnewuch et al., 2018 | 7-point Likert scale | Descriptive statistics; Welch's two-sample t-test. | Highly usable. |
| | | Appel et al., 2012 | 5-point Likert scale | Descriptive statistics. | |
| | | Verhagen et al., 2014 | 7-point Likert scale | Partial least squares. | |
| b | Perceived Humanness | Gnewuch et al., 2018 | 9-point semantic differential scale | Descriptive statistics; | Highly usable. |
| | | Candello et al., 2017 | N/A | Dichotomy. | |
| | | Schuetzler et al., 2014 | 6-point semantic differential scale | Welch's two-sample t-test. | |
| c | Service Encounter Satisfaction | Gnewuch et al., 2018 | 7-point Likert scale | Descriptive statistics. | Highly usable. |
| | | Verhagen et al., 2014 | 7-point Likert scale | Partial least squares. | |
| d | User Experience | Hu et al., 2018 | Text measures. | Linear Regression analysis. | Potentially Usable. |
| e | Person Perception | Appel et al., 2012. | 7-point semantic differential scale | Welch's two-sample t-test. | Not usable. |
| f | Rapport Factors | Appel et al., 2012. | 8-point Likert scale | Combined Factor Analysis. | Not usable. |
| g | Perceived Engagement | Schuetzler et al., 2014. | 7-point Likert scale. | Welch's two sample t-test. | Not usable. |
| h | Response Latency | Schuetzler et al., 2014. | Absolute measurement | Descriptive statistics. | Not usable. |
| I | Effectiveness | Gourov, 2019. | 7-point Likert scale | Regression analysis. | Not usable. |
| j | Efficiency | Gourov, 2019. | 7-point Likert scale | Regression analysis. | Not usable. |
| k | User Satisfaction | Gourov, 2019. | 7-point Likert scale | Regression analysis. | Not usable. |
| l | Novelty | Gourov, 2019. | 7-point Likert scale | Regression analysis. | Not usable. |
| m | Reputation Rate | Murgia et al., 2016 | Received up-and down votes. | Baseline comparison. | Not usable. |
| n | Personalization | Verhagen et al., 2014 | 7-point Likert scale | Partial least squares. | Not usable. |

Table 4: Used constructs in comparable research.

3. Theoretical Framework

In Table 4 we see again a table of comparable research, this time sorted on the used constructs, together with their way of measurement and used analysis. In the last column, the usability for this thesis is denoted.

The analysis of items in this table will now be structured column-wise. Meaning that the contents of the columns will be discussed, in left to right order.

3.1.1 Constructs

We see that some constructs have been used more frequently than others. The constructs social presence, perceived humanness and service encounter satisfaction have been used the most frequent in research which is comparable to this thesis. We also see that these constructs are most frequently combined within one single research. This is clearly depicted in the research from Gnewuch et al. (2018), as their research uses all three constructs. Also the research from Verhagen et al. (2014) uses two out of these three constructs.

Similarly, the measurements used for these constructs are alike. For social presence and service encounter satisfaction a Likert scale is mostly used, while for perceived humanness a semantic differential scale is preferred.

Although the similarity in the measures can be seen, their statistical methods used for analysis differ. For the social presence construct, Gnewuch et al. (2018) for example make use of basic statistics such as mean and standard deviation, which are called descriptive statistics in the table. Their objective was to show a difference between humanness and social presence between the two chatbot conditions, which had a significant result. On the other hand, the research of Verhagen et al. (2014) measured the same construct but used different analyses methods. They used a partial least squares method, since this allowed them to test their moderating effects via multiple group analysis.

Altogether, the use of social presence as a construct is highly usable for this thesis's research. This is due to the fact that social presence has an extensive background in literature, a clear demarcated way of measuring, and has been used before in comparable research, in a successful way.

An also frequently used construct in research is perceived humanness. This construct's use in research is more dispersed, meaning that less articles use this construct in a combination with other frequently used constructs, as was with social presence. Perceived humanness is

3. Theoretical Framework

thus more used as a single measure, and is often combined with a metadata measurement, such as response latency. This was for example done by Schuetzler et al. (2014) and Candello et al. (2017), who also make use of the construct perceived humanness. Their way of measurement is not exposed, and the only information that can be traced back is the use of a form of dichotomy. They measured the percentage of cases perceived as either men or machine in a yes-no way.

Nevertheless, this construct is highly usable for this thesis. This is due to the fact that the measure perceived humanness exactly measures an important element for this thesis, and that its use is reflected in comparable research.

The last construct which can be traced back in more than one comparable article is service encounter satisfaction. This construct is used in the research of Gnewuch et al. (2018) and Verhagen et al. (2014), and has therefore close relationships with the constructs social presence and perceived humanness, since they are also used in at least one of these articles. Measurement is done in similar way, namely with a 7-point Likert scale, but analysis differs. This is probably due to the different degree in comprehensiveness in both articles.

Service encounter satisfaction, or in a shorter form service satisfaction or just satisfaction, is closely related to constructs used in other research, such as user experience (Gourov, 2019; Hu et al., 2018b) and user satisfaction (Gourov, 2019). All constructs measure, in a comparable way, the final opinion of the user on the interaction with the chatbot.

Although we see that the measures used for service encounter satisfaction are exactly similar, scientifically substantial, and based on prior scientific literature, the questions used measure selectively on the held interaction with the chatbot, the quality of the advice given, and the way of treatment by the chatbot. Gourov's (2019) measurement by contrast measures user satisfaction more in a way of willingness to use a chatbot in the future, based on the held interaction.

Side notes on this short analysis are that detailed information about measurement and questions in the research of Hu et al. (2018a) is not mentioned, and therefore not taken into account in this paragraph. Second note here is that Gourov (2019) measures an aggregated form of the construct user experience, whereof user satisfaction is a component.

This short analysis on comparable constructs is important for the usability for this thesis. Since the construct service encounter satisfaction has a more profound scientific basis, higher reproducibility, and clearer measurements, we see the construct Service Encounter Satisfaction as highly usable for this thesis, were other similar constructs are usable to not-usable.

3.1.2 Construct Selection Criteria

In the previous table we can see various constructs used in similar research. This paragraph will elaborate on the last column in the table: the effect in this thesis. This mainly comes down to the usability in this research.

As said before, this literature review's goal is to find suitable items which can be used in answering the research question.

Criteria for weighing of the usability of constructs is based on the following:

- i. Construct maturity, as reflected in the previous use in research.
- ii. Validity of the primary measures of the construct.

First argument for weighing of construct usability is its maturity. The maturity of a construct is determined by how frequently the construct used this construct is in previous research. A higher frequency comes with a better defined construct based on theory, and lowers the risk of unexpected results. Examples of a high construct maturity is an established measuring scale, or by a previous calculated Cronbach's Alpha (α). Also, all constructs come with according measurement variables and a measurement scale. These measurements have been used before in research, and it is clear what and how they measure. This leads to a high validity. Again, choosing renowned measurements lowers the risk of unexpected results and lowers the chances of biases in the research.

The second argument of using this construct in this thesis is a high validity. It is important that what the constructs measurements measure, is relevant for this thesis. Therefore this thesis must choose measures accordingly. An important selection criteria is therefore the scope of the construct measurement.

3.1.3 Other Constructs

All other constructs seen in comparable research have no similarity or relation to other constructs, and are not used more than once in research articles. This means that there is no overlap in the measures of these constructs, and that no two constructs measure the same thing, even though they would have been named differently. These constructs are also not used in more than one research, and are specifically suitable for one specific research topic. This also means that they are topic-wise further apart from this thesis's research topic, and therefore not applicable and relevant in this thesis's research approach.

3.2 Social Presence

| Theory | Construct | Variable | Definition |
|------------------------|-----------------|-------------------|--|
| Social Presence Theory | Social Presence | Human Contact | The degree to which the communication has been experienced as natural. |
| | | Personalness | The quality or state of being personal, and being appealing to the individual. |
| | | Sociability | The degree to which the quality, state, disposition or inclination of being sociable has been experienced. |
| | | Human Warmth | The sensation of friendliness, kindness or affection. |
| | | Human Sensitivity | The quality of being tender, sympathetic and sensitive, and responding to signals expressing these. |

Table 5: Social Presence Taxonomy Table.

Social presence theory was first defined by Short, Williams, & Christie in 1976 as: “the degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships” (p. 65). In 1995 this definition has been clarified by Gunawardena as: “the degree to which a person is perceived as a ‘real person’ in mediated communication” (Gunawardena, 1995, p. 151). In general, the social presence theory defines the perception of the conversational partner, in terms of humanness. The theory is related to CMC – computer mediated communication, which means interaction between humans making use of computer technology – and made an effort to explain the effect a communication medium can have on the way people communicate. Short et al., (1976) see social presence as a key factor of such a communication medium, which influences the way people communicate with each other. They state that different media of communications have different degrees of social presence. So, video communication has a higher social presence than audio communication. Lowenthal (2011) sees that the definition of social presence slightly changes per application, as the definition remains interpretable and varies per research. However, this research has made great effort to make social presence measurable, over a continuum scale.

In the scope of this research, social presence has mainly been used as measurement in research questionnaires. Herein the scale developed by Lowenthal (2011) is frequently used. These scales have been practiced by comparable research, which has led to a set of questions, with accompanying scales and validity test values. The questions mainly ask on feelings as sensing human contact, personalness, sociability, human warmth, and sensitivity with the conversational partner (Gnewuch et al., 2018; Verhagen et al., 2014). These constructs have been measured on 7-point Likert scales, and lead to high reliability values.

3.3 Perceived humanness

| Theory | Construct | Variable | Definition |
|---|---------------------|----------------|---|
| Three-factor theory of anthropomorphism | Perceived Humanness | Human likeness | The degree to which an actor acts as and shows similarities with a human being. |
| | | Skill level | The amount and level of expertise displayed. |
| | | Thoughtfulness | The amount of consideration shown to others. |
| | | Politeness | The showing of good manners in behavior and expression as commonly accepted by society. |
| | | Responsiveness | The speed and broadness of reaction. |
| | | Engagement | The scope and depth of possible interaction. |

Table 6: Perceived Humanness Taxonomy Table.

The construct perceived humanness finds its origin in the three-factor theory of anthropomorphism (Epley, Waytz, & Cacioppo, 2007). This three-factor theory of anthropomorphism defines anthropomorphism as: “the tendency to imbue the real or imagined behavior of nonhuman agents with humanlike characteristics, motivations, intentions, or emotions.” (Epley et al., 2007, p. 1). This can be either to animals, gods or objects, where chatbots fall in this latter category. Epley et al. (2007) divide their theory in three main factors:

- i. Elicit Agent Knowledge
- ii. Effectance motivation
- iii. Sociality motivation

The first item, elicit agent knowledge, is about the accessibility and applicability of the knowledge of anthropomorphism. Effectance motivation explains the motivation to understand other agents’ behavior. The latter, sociality motivation describes the desire for social contact.

Anthropomorphism in the attribution of human characteristics to something which is inherently non-human. One of the constructs is Perceived Humanness.

Perceived Humanness is a construct frequently used in Human Computer Interaction (HCI). It is a construct which is merely created while executing experiments, and is based on a continuum reaching from machine like to human like, as seen in various research (Candello et al., 2017; Gnewuch et al., 2018; Schuetzler et al., 2014). This continuum could be formatted as a scale, ranging from six to nine points (Gnewuch et al., 2018; Schuetzler et al., 2014) or as a dichotomy, ranging from simply yes to no (Candello et al., 2017).

Perceived Humanness has been around since the introduction of HCI, Human Computer Interaction. Kim (2015) has described that within HCI user experience is currently the most

3. Theoretical Framework

important factor. However, he states, that it is important to know the preferences of the user, which is communicating as natural as possible. Therefore the perception of the interaction partner is important, as is its perceived humanness.

The current research executed uses the scales developed by Holtgraves & Han (2007), which simply consist of an nine point semantic differential scale, ranging from extremely nonhuman to extremely human. This scale was accompanied by a similar scale, also nine point and semantic differential, where the participant was invited to rate their conversational partner from definitely a human to definitely a bot. This forms the basis use of perceived humanness that can be traced back into literature.

More recent studies have expanded this construct with items consisting of the human likeness, skill level, thoughtfulness and engagement (Gnewuch et al., 2018). Schuetzler et al. (2014) even refer back to the already named Turing Test (Turing, 1950) as a basic construct of perceived humanness. The better a piece of software scores on the Turing Test, the higher the perceived humanness is.

This still gives no clear notion of a theory of what perceived humanness actually is. This is due to the fact that perceived humanness did not originate from a scientific designed theory, but more from practical usage. This means that the measures are clearly described in articles, but the actual thing that should be measured is not clearly described.

For the sake of this thesis we will from now on use perceived humanness as its intended use, based on the comparable research which uses this construct.

3.4 Service Encounter Satisfaction

| Theory | Construct | Variable | Definition |
|---------------------------------------|--------------------------------|----------------------------------|---|
| Service Encounter Satisfaction Theory | Service Encounter Satisfaction | Advice Satisfaction | The degree of satisfaction with the proposed advice. |
| | | Treatment Satisfaction | The level of satisfaction with the way of treatment by the other party. |
| | | Overall Interaction Satisfaction | The level of satisfaction experienced in the overall experience of the interaction. |

Table 7: Service Encounter Satisfaction Taxonomy Table.

Service encounter satisfaction is a construct closely related to measuring and understanding customer satisfaction. It mainly emerged from the research area of marketing, in order to understand customer's perception. This was firstly expressed by Rushton & Carson (1989) in order to measure satisfaction over the goods-service continuum. Services are more intangible, and therefore harder to evaluate, they state. The construct of service encounter satisfaction is based on this difficulty, and aims to unravel the underlying processes relevant for reviewing a product or service in terms of satisfaction.

Walker (1995) has published a conceptualization for the construct service encounter satisfaction. This article states that it is generally agreed upon that consumer satisfaction is the result from a weigh off between the expected and perceived attribute levels. Expectations prior to the encounter and perceived evaluations after the encounter, related to performance of the product or service, form the basic comparison on which the theory is settled. This results in three outcome situations, based on a higher expectation than reality, an equal expectation to reality, and a lower expectation than reality. This leads respectively to a negative disconfirmation, a neutral confirmation or a positive disconfirmation. This then leads to different satisfactory outcomes.

Apart from this conceptualization study, we see a more practical application of this theory. This practical implementation in various scientific studies has led to a measurement and scale applicable in research. This is reflected in various research, related to service and product evaluation (Barger, Grandey, Barger, & Grandey, 2006) but also seen in chatbot interaction situations (Gnewuch et al., 2018; Gourov, 2019; Verhagen et al., 2014).

3.5 Difference between Social Presence and Perceived Humanness

On the first glance, the two constructs Social Presence and Perceived Humanness, which are both explained in this paragraph, may look similar. Their measurements are showing similar characteristics, and the actual construct they are measuring also looks similar. There is however an important difference between both.

Social Presence measures the degree to which a person is perceived as a real person (Gunawardena, 1995), this perception is altered due to the method which people are using to communicate. It is measured with aspects of that communication, such as human warmth and sensitivity. The theory says that the more social cues are transferred during the communication, the higher the social presence will be. Social presence is thus more about the method of communication and the interaction itself, and thus computers and humans can theoretically achieve the same level of social presence.

Perceived Humanness is more focused on the conversational partner, instead of the conversation. The rating on this scale is directly coupled to the identity of the partner with whom somebody is interacting. In a theoretical similar situation a human will always achieve a higher score than a machine. This effect is still altered through the method of communication, but more in the form of distorting effects.

Since in a chatbot situation interaction only consists of written text messages, many social cues are deleted. This leads to the situation where we cannot directly assess the level of humanity of our partner, and thus we measure the perceived humanness as a construct.

3.6 Social Information Processing Theory

| Theory | Construct | Variable | Definition |
|--------------------------------------|---------------------|-----------------------------|---|
| Social Information Processing Theory | Hyperpersonal Model | Selective Self-Presentation | The ability by the sender to present itself in a controlled way. |
| | | Idealization of the Sender | The perception that the receivers makes of the sender, which is an idealized perception due to the reduced social cues. |
| | | Channel Management | The nature of the used communication channel. |
| | | Feedback | Behavioral confirmation, reaffirming the original prediction of the individual's expectation |

Table 8: Social Information Processing Theory Taxonomy Table.

The Social Information Processing Theory was originally founded by Joseph B. Walther in 1992 (Berger, Roloff, & Walther, 2016) and explains “The development of interpersonal impressions and relational communication via computer-mediated communication (CMC)” (Berger et al., 2016, p. 1). In general the theory explains how relations are created via computer-mediated communication. Building personal relations through CMC was deemed impossible based on the existing theories, due to the severe differences between CMC and human-human interaction (HHI).

Three theories already present during the creation of the social information processing theory were the social presence theory, the media richness theory and the reduces social cues approach (Daft & Lengel, 1983; Short et al., 1976; Tanis & Postmes, 2003). These theories combined described a paradigm in CMC where building personal relations was arduous. This was due to the fact that CMC had little bandwidth to express complexity in messages (Daft & Lengel, 1983), consequently leading to reduced transfer of verbal and nonverbal social cues (Tanis & Postmes, 2003), and thus leading to a reduced feeling of human aspects such as warmth and sensibility, as described in the social presence theory (Short et al., 1976). Basically established literature sketched CMC as an environment much less social as real face to face communication.

The Social Information Processing theory acknowledges these findings in existing theories, and reacts on these differences in two ways. The theory explains that relationships can be built up, based on verbal cues, and the extended time.

The Social Information Processing theory states, according to Berger et al. (2016), that the reduced social cues that are transmitted over CMC, are replaced by verbal cues. What users do not convey with body language and other non-verbal signs, are transformed into spoken cues

3. Theoretical Framework

and conveyed in a written way. This reasoning of interchangeability of cues, is of the essence for the social information processing theory. Therefore the absence of nonverbal cues in CMC cannot solely be seen as a deficit in the expressive capability leading to a comparative reduction in the quality of social interaction (Berger et al., 2016).

The second part of the theory described the extended time necessary for a relationship to build. Whereas the theory acknowledges the lesser amount of richness of media transferred, it states that a longer duration in length of time compensates for the lack of transferred social cues. It compares the building of relationships with sipping versus gulping. In the normal face to face conversation in human-human interaction, people tend to gulp all the impressions that are formed all at once, where on the contrary in CMC, the impressions are formed at a reduced rate, comparable to taking smaller sips (Griffin, 2012).

The social information processing theory thus states, that although the situation is different, due to the lack of social cues transferred, personal relationship building it still possible due to the translation into verbal cues and the extended time. This theory is also strongly related with some other relevant theories described before, such as the social presence theory.

The theory also introduces various constructs, such as the impersonal, interpersonal and hyperpersonal model. Especially this latter is of importance, since it consists of interesting variables. The hyperpersonal model consists of four variables, and selective self-presentation is relevant for this thesis.

Griffin (2012) describes selective self-presentation as the opportunity to make and sustain an overwhelmingly positive impression. That is because the sender can control which factors he or she shares and thus can make a controlled impression. Self-presentation is highly related to self-disclosure, where this last one has a higher impact on the entire relationship, where self-presentation focusses more on the beginning of the relationship.

3.7 Relationships

The following paragraph will explain the various relations between the constructs as shown in the conceptual model in Figure 2.

3.7.1 *Selective Self-Presentation and Social Presence*

Selective self-presentation, as part of the social information processing theory is highly related to social presence. Both theories acknowledge the fact that in CMC there is a reduced number of social cues. The social information processing theory explains that these non-verbally expressed cues are transformed into verbal cues in written communication (Berger et al., 2016). This created the opportunity for a selective self-presentation. The social presence theory says that these reduced social cues lead to a lesser extend of social presence in the conversation (Short et al., 1976).

The evaluation of social presence of a conversational partner is based on the interactions in the held conversations. In a CMC environment these interactions consist of the messages send between the conversational partners. As social presence focusses on the degree of salience and the degree to which a conversational partner is perceived as 'real' (Gunawardena, 1995; Short et al., 1976), these perceptions are based on the conversational information and social cues expressed by the conversational partner.

Selective self-presentation is an important social cue expressed in interactions between conversational partners. The theory acknowledges the importance of social cues in conversations, and focusses on the development of interpersonal impressions via computer-mediated-communication (Berger et al., 2016). This means that selective self-presentation should be seen as a social cue expressed through written communication in computer-mediated communication.

By means of logically deductive reasoning we can therefore elaborate on the relationship between both constructs. The construct of social presence is based on social cues, sent messages and expressions of salience and 'realness' of conversational partners. The variable selective self-presentation is such a social cue, a sent message and an expression of salience and 'realness' of the conversational partner. The relationship between both can therefore be logically deducted as that social presence is, partially and among others, based on selective self-presentation.

3. Theoretical Framework

Based on this concluded relationship between both constructs, we can also give a direction to this relationship. Selective self-presentation can be seen on a level. A high level of self-presentation comes with a true, all-encompassing and fully open self-presentation of the conversational partner. The more information the conversational partner reveals about himself, bound to the condition that the information is true, will lead to a higher level of selective self-presentation. This means an expression of more social cues and information to the other conversational partner to base its evaluation of social presence on. Therefore we can state that the relationship between both constructs is positively correlated.

For sake of this thesis a high level of self-presentation will be called an identity revealing self-presentation, and a low level of self-presentation will be identified as neutral self-presentation.

Based on the conclusions of this paragraph, we can formulate the following hypothesis:

H₁: A chatbot with an identity revealing self-presentation will yield a higher experienced social presence than a chatbot with a neutral self-presentation.

3.7.2 Selective Self-Presentation and Perceived Humanness

The two constructs selective self-presentation and perceived humanness are closely linked together. As said in the previous paragraph, selective self-presentation is a social cue expressed through written messages in computer-mediated communication (Berger et al., 2016).

Perceived humanness is extracted from the three-factor theory of anthropomorphism. This theory is defined as the tendency to imbue behavior from nonhuman agents with humanlike characteristics (Epley et al., 2007). This process of imbuelement is based on the expression of these humanlike characteristics, such as motivations, intentions or emotions. The computer-mediated theories prescribe that in CMC the expressions of these characteristics will happen through verbally expressed social cues and written messages (Tanis & Postmes, 2003).

With the use of logically deductive reasoning, we can extract the relationship between these two constructs. In this rule of reasoning, two conditional statements and one antecedent can be found in order to conclude on the relationship.

First the conditional statement that perceived humanness is based on imbued human behavior of nonhuman agents (Epley et al., 2007). This definition is extracted from the theory's literature. The second conditional statement is that in computer-mediated communication

3. Theoretical Framework

behavior is expressed by social cues and verbal written messages (Tanis & Postmes, 2003). The antecedent is that selective self-presentation is a social cue and a verbal written message aimed at expressing social behavior. Therefore it can be concluded that, by using CMC theories, the imbued human behavior which formulates the evaluation of perceived humanness is, at least partially and among others, based on selective self-presentation from the conversational partner.

Based on this relationship, a direction can be added. As expressed before, selective-self presentation can be seen on a level, leading from high to low. For use in this thesis this level will be identified as identity revealing and neutral self-presentation. Since a higher expression of self-presentation will lead to the production of more social cues and written messages which indicate a higher disclosure of the conversational partner's identity and more social cues will lead to a higher imbuement of perceived humanness, we can state that the correlated relationship between both constructs is positive.

Therefore we can formulate the following hypothesis:

H₂: A chatbot with an identity revealing self-presentation will yield a higher experienced perceived humanness than a chatbot with a neutral self-presentation.

3.7.3 Selective Self-Presentation and Service Encounter Satisfaction

Service Encounter Satisfaction and Selective Self-presentation are linked together. The service encounter satisfaction mainly focusses on the satisfaction of the customer on the experienced service. It mainly focusses on the way this experience is brought, and important variables are the way of treatment and the overall interaction satisfaction. Also the final advice plays part in the rating on this scale. The theory at the basis of this construct says that the measure of service encounter satisfaction is used for reviewing a product or service in terms of satisfaction (Rushton & Carson, 1989).

Anderson, Pearo, & Widener (2008) have identified the various components of service encounter satisfaction. They identified in the paradigm of service-dominant logic, the drivers of operational performance, service interactions and physical setting. This paper indicates that service encounter satisfaction is based on the interaction that takes place in this service, which is based on communication between service provider and service recipient.

3. Theoretical Framework

The CMC theories state that in computer-mediated communication there is a lower presence of verbally expressed social interaction (Kim, 2015). Instead, these interactions are transformed into written messages. In the situation of computer-mediated communication there is a reduction in the social cues expressed and transmitted.

The social information processing theory acknowledges this reduction of social cues and states their transformation into written expressions (Berger et al., 2016). This theory has put forward the construct selective self-presentation as part of the hyperpersonal model to make up for the loss of non-verbally expressed social cues.

By making use of the logically deductive reasoning method we can express the relationship between both constructs, while making use of the CMC theories in a supportive role. The service encounter satisfaction theory says that the construct is based on the service interaction that takes place (Anderson et al., 2008). The social information processing theory states that the construct selective self-presentation is expressed in the domain of interaction. The supportive CMC theories say that these interactions are expressed in written form via written messages. Therefore we can state that the constructs selective self-presentation and service encounter satisfaction are related.

Although this relationship strength is not as high as with the other constructs, it is harder to give a sound explanation of the direction of this relationship. Nevertheless with logical reasoning we can propose a positive relationship between both. This reasoning is based on the fact that the better you know a person, the more you will like him or her. This is backed up by the social information processing theory (Berger et al., 2016). We can therefore also assume that in the situation of providing a service, knowing the service provider better will make you more like the service. Therefore we can assume a positive relationship, which leads to the formulation of the following hypothesis:

H₃: A chatbot with an identity revealing self-presentation will yield a higher experienced service encounter satisfaction than a chatbot with a neutral self-presentation.

3.8 Conceptual Model

This finalizes the conceptual model as follows.

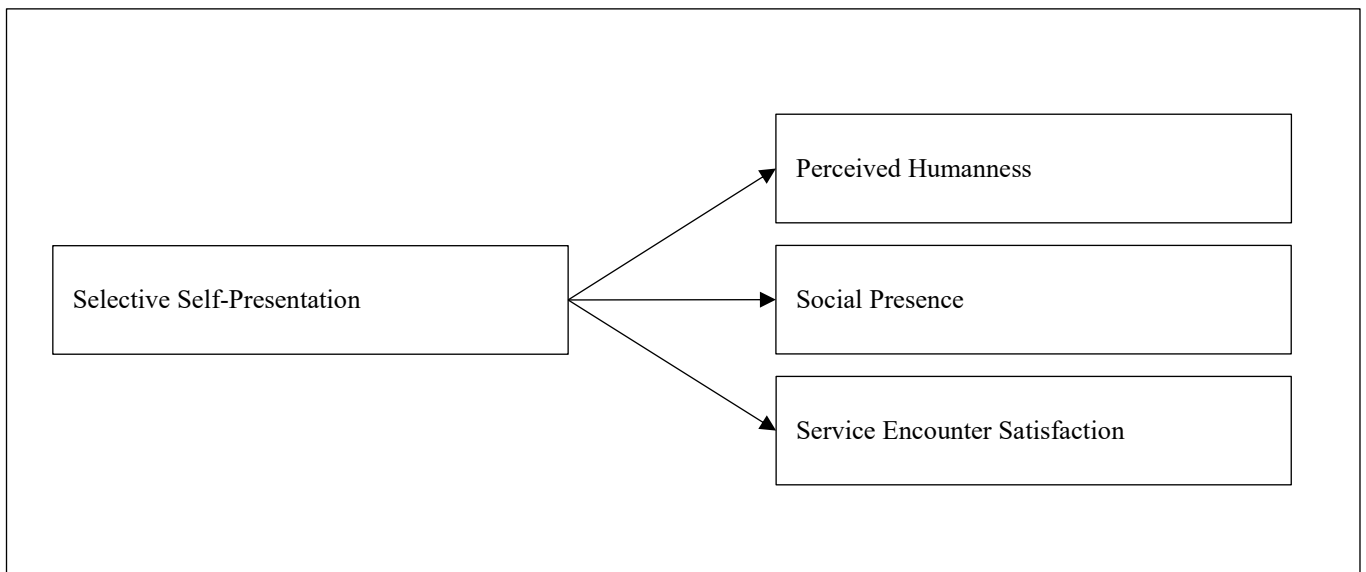


Figure 2: Conceptual Model.

3. Theoretical Framework

4 RESEARCH DESIGN

This chapter in the research explains about the followed methodology for this thesis. It will explain the followed steps and will justify each decision taken.

4.1 Experiment Design

This research aims to answer the research question: Which chatbot introduction yields the highest user satisfaction?

In order to do so, an experimental design was set up. The reason to use an experimental design approach is because the research examines a relationship that varies between groups. The research question of this thesis researches the relationship between the introductory message of a conversational partner and the three constructs Social Presence, Perceived Humanness and Service Encounter Satisfaction. Since theory prescribes a potential causal relationship, an experimental research design is chosen for this thesis. This experiment falls under a true-experimental design, as the environment and the independent variable were completely controlled.

By means of a one-way, between groups experiment, this relationship was examined. It involved a posttest only design, and only manipulated one variable. Therefore an ANOVA statistical procedure was used to check significance, and mean comparison was used in interpret results. Also, to determine the size of the impact in the relationship, several regression analyses have been executed.

This research is in the form of a survey experiment (Gaines et al., 2007). This is reflected by the build framework from the previous chapter, which also showed this form of methodology as the most common in the chatbot interaction research field. Therefore it has also been chosen for this research.

The justification for this approach is based on inherent factors of this approach, as well as the fact that this approach has frequently been used before in comparable research. Inherent factors contain the fact that it is an experiment, and thus enables the use of various research groups and testing various manipulations of the variable. A survey afterwards enables the participant to show their intentions and feelings, which can be used in statistical analysis. This combination enables the research of various different introductions and expose the participant

to these manipulations. The survey enables the measurement of the three used constructs to base calculations and conclusions on. This is in line with the research question of this research.

The use of this survey experiment approach is fortified by the fact that previous and comparable research also commonly uses this approach, and that the used constructs in this research have more frequently been used in this kind of methodology. Therefore the maturity and reliability of this approach is higher. This ensures robustness of the methodology and reliability of the final results.

4.2 Vignette Research

This experiment makes use of a vignette research style. Atzmüller & Steiner (2010) define a vignette as: “a short, carefully constructed description of a person, object, or situation, representing a systematic combination of characteristics.” (p. 2). A vignette research is always accompanied by a survey afterwards, and thus consist of two components. Participants are first exposed to the vignette itself, and afterwards are asked to answer questions for the measurement of respondent-specific characteristics.

In this research the vignette approach was operationalized by a fictional situation where after the participants were requested to show their feelings and intentions in answering survey questions. The vignette in this research consisted of a short video showing an excerpt of a chat conversation. The situation sketch was operationalized by a short introductory message preceding the vignette, and expressed a hypothetical situation regarding a customer care question.

Participants first saw the introductory message, which led them into the situation. The situational sketch explained that the participants bought a non-specified product at a certain web shop, and their intention of returning this product in order to get a refund. The nature of the bought product was non-specified, as it was not relevant for the research of the participants' intention. The situational sketch explained to the participants that in order to gain information about the return policy of the web shop, the participant opened a chat window to have a conversation with the customer care department. In the following part, the participant saw the video which showed the conversation the participant had with the customer care department. Participants were told that, despite they did not actually held this conversation, as it was fictional and pre-recorded, they were to act as if they were having this conversation, and to act on the successive questions accordingly.

4.2.1 Justification for Vignette Research

This thesis followed a vignette research approach because of two reasons:

- i. Reduction of confounding variance.
- ii. High control of treatment variance.

The use of vignette research reduces the effect of confound variance in this particular research. The reduction of confounding factors increases the internal validity (Slack & Draugalis, 2001). This is mainly due to the fact that videos are used in the vignettes. Because of using video excerpts of chats, instead of, for example, letting participants having a real chat interaction, the research approach is more structured. The following confounding variables are mitigated with this approach:

- Level of technology maturity
- Outcome and sequence of the conversation
- Topic of the conversation
- History effects

The level of technology maturity of the participants could introduce a confounding variable in this research. When participants were allowed a free chat interaction, the level of acquaintance with this technology or the experience with contact with customer care departments could influence the sequence or outcome of the chat. By having the conversation and the message structure fixed by a vignette, this confounding effect is eliminated.

Similar approach holds for the outcome and sequence of the conversation. If various chats held by participants would have different outcomes, for example because of different input messages, the answers of the following questionnaire could be biased. By having the sequence and outcome of the messages fixed in a video vignette, this confounding variable is eliminated.

Also for the topic of the conversation, the same logic holds. Because in a free chat interaction, participants could wander off their original goal. Even though participants could end up with the same outcome, the topic could be different during the conversation, for a limited time. This could occur despite the clear introduction and assignment given. By showing a video excerpt in a vignette, this confounding variable is reduced.

Finally the confounding effect of history effects. History effects are the intervening events that could influence measurement outcomes (Christ, 2007). This effect is partially mitigated by vignette approach, and partially by random group assignment to the conditions. The vignette approach mitigates events that can occur outside the research setting but inside the chat experience. These intervening events include errors from the chat partner, for example going

4. Research Design

outside the chatbot's database, or connection problems. By showing a video the confounding effect of history effects is minimized.

The second reason why a vignette research was used is the control of treatment variance. Because the videos show a pre-fixed conversation, the differences between these conversations can be controlled in a detailed way. Also, the differences between the conversations are fixed and do not differ depending on the participant, as it would have been when the survey allowed for a free or structured interaction.

Altogether, the use of a vignette research design leads to a higher internal validity by elimination of confounding factors and leading to a higher manipulation control.

4.3 Manipulation

4.3.1 Manipulated Variable

The manipulated variable was selective self-presentation, which originated from the social information processing theory, as explained in paragraph 0. The operationalization of this manipulation was based on designing different introduction messages which corresponded with the conditions designed.

4.3.2 Manipulation Check

In the research a manipulation check was present. This check comes highly recommended, as it improves the stability of the framework and verifies the implementation of the controlled factors (Foschi, 2014). Therefore it is an essential part of an experiment.

In this research the manipulation checks have been operationalized by a control question in the questionnaire. Since the manipulated variable was the introduction of the conversational partner, reaching from chatbot to human, the control question involved the participant's opinion on the identity of the conversational partner. The question was asked on a semantic differential scale, leading from 'definitely a chatbot' to 'definitely a human being'. As advised by Hoewe (2017), the manipulation check questions were geared towards the participant's understanding of the cognizance regarding the condition the participant was exposed to.

4. Research Design

The logic behind this setup of questions was that participants would rate a chatbot as more chatbot on this scale, and respectively a human being more as a human being. The conversational partner with an undefined introduction would have to be in-between these two for the logic to withstand.

The success of the manipulation was measured with the statistical results of the answers on the manipulation check question. A predetermined scattering of the mean answers on this question from the various conditions was used as a measure for success of the manipulation. If these means were in order and with enough separation, the manipulation check was successful.

The required scattering was a sequential order of the means of respectively the chatbot condition, undefined condition, and the human being condition. Possible non-symmetrical skewness of the distribution was accepted as long as the mean separation between the conditions was large enough.

Practically this meant that the chatbot condition mean should score very low on this scale, and the human being condition should score very high on this scale. The undefined condition should score in the middle of both means from the other conditions. Skewness of the means distribution was deemed irrelevant, as long as the sequence was intact and the mean separation of the three conditions was enough.

$$\bar{x}_{(\text{Control Group})} > \bar{x}_{(\text{Treatment group 2})} > \bar{x}_{(\text{Treatment group 1})}$$

Table 9: Required mean sequence of the manipulation check question for manipulation success.

The reason for these specific criteria of success of the manipulation check was due to the indicative nature of the pre-test. No significant results were to be expected as the sample size of the pre-test was small ($n = 19$). The pre-test had as a goal to check the potential success of the manipulation, as an indication of the significance of the main experiment's results. The required mean sequence and separation were an indication to the effect in the main experiment. Therefore, significance indicators of the pre-test's results were neglected.

4.3.3 Location of Manipulation Check Questions in the Survey Sequence

For the pre-testing phase, the manipulation check question was asked immediately after the manipulation. This is in accordance with Aronson & Carlsmith (1968), as they advise this to avoid a reduction in the participant's ability to fully describe its reaction to the manipulation. For the main experiment, the manipulation check questions were asked after the measurement

4. Research Design

of the dependent variables, in contrast to the pre-testing phase, were they were asked before the dependent variables measure.

The reason for this research setup is due to the different goals of both the pre-test and the main experiment. The pre-testing phase was aimed at checking if the maximization of the treatment variance was sufficient to have a significant effect. Therefore it was important to ask the manipulation check as early as possible in the experiment (Aronson & Carlsmith, 1968). For the main experiment, the outcome of the manipulation check is less important, and the focus is on the results of the measurement of the dependent variables. The manipulation checks are then moved to the end of the survey, to avoid biases and confounding effects (Summers & Perdue, 1986).

4.4 Condition Design

This research design is a one-way design with three conditions based on three different manipulations of the independent variable. Two conditions were considered as treatment groups, one condition was a control group, and considered as baseline measure. The manipulated variable was based on the construct Selective Self-Presentation, extracted from the Social Information Processing Theory.

A control group in this research design was necessary because of the absence of an objective measure of the baseline level of the independent variable. Therefore this control group had the design of a normal situation, wherein a situation of human-human interaction was displayed.

4. Research Design

| Treatment group name | Treatment Group 1 | Treatment Group 2 | Control Group |
|---|--|---|--|
| Introduction Manipulation | Introduction as Chatbot | Introduction Undefined | Introduction as Human Being |
| Visible Connection Status Metadata ² | Setting you up with a customer care chatbot. Please wait while we set up the connection. Conversation is set up with 'Customer Chatbot' | Setting up your conversation with customer care. Please wait while we set up the connection. Conversation is set up. | Setting up your conversation with a customer care employee. Please wait until an expert is available to speak to you. Conversation is set up with 'Peter' |
| Intro message in vignette | Hello, you are talking with a customer care chatbot. How can I help you? | Hello, how can I help you? | Hello, you are talking with a customer care employee. How can I help you? |

Table 10: Schematical design of the experiment conditions, groups and definition.

4.5 Experimental Control

4.5.1 Randomization

Participants were randomly allocated to one of the various conditions. Randomization increases the internal validity and decreases possible confounding factors in the experiment (Senthil Mahesh, Rodrigues, & Campbell, 2013). Randomization in this research was done automatically, and non-discriminant on any factor. Despite the random allocation, all conditions were evenly shown. This was to ensure an even distribution among the conditions. The randomization algorithm was used before the movie was showed, and therefore the dropout rate was not taken into account.

² This manipulation was only present in the second pre-test and main experiment. It was not part of the first pre-test.

4.5.2 Response Latency

The response latency of the sent and received messages was set to a dynamical delay. This is in line with the research from Gnewuch et al. (2018). To make sure no confounding variable was created, the delay times were held exactly the same in all the conditions.

| |
|---|
| $\text{Message response delay (s)} = 2 + (\text{Amount of words} * 0.35)$ |
|---|

Table 11: Formula to calculate the dynamical message delay. All figures represent seconds.

As can be seen in Table 11 the response delay was dynamically delayed based on the amount of words the message contained, combined with a pre-set delay of two seconds. The exact delay times for each of the shown messages can be found in the appendix (8.3).

4.6 Pre-test

In order to provide a decent and robust research setup, a pre-test has been executed. This pre-test had as goal to verify the assumed results, and eliminate potential design errors. In particular, this pre-test focused on the manipulation between the conditions. Since this research involves a small effect size, resulting in a small manipulation between groups, a pre-test was deemed necessary.

This paragraph reports about the procedure and setup of the held pre-tests. Since the pre-tests focused on the manipulation, more detailed information about the organization of manipulation checks and justifications can be found in paragraph 4.3.

The main goal of the pre-test was to check if the maximization of the treatment variance was sufficient to have a significant effect in the eventual main experiment. This effect of the treatment variance was assessed by a survey question measuring the effect of the manipulation on the participants.

In total, two pre-tests have been executed. The reason for this was a failure in the first pre-test to report results that were in line with the requested sequential mean scattering and mean separation between conditions. Therefore a second pre-test was held with adjusted specifications.

Both pre-tests were held with a sample size of 19 people ($n = 19$), where in the second pre-test one response was deleted because of a failure to watch the videos. The participant selection

4. Research Design

for both pre-tests were on invitation basis. By means of paper cards people were invited to take part in this pre-test. The condition assignment was based on random assignment.

In the first pre-test, the results did not show a success of the manipulation, and therefore amendments were made before a second pre-test was held.

4.6.1 Results First Pre-test

The focus of both pre-tests was mainly to check the success of the manipulation. Therefore the focus was mostly on the manipulation check question. In both pre-tests this question was asked on a 9-point semantic differential scale, and asked directly after the manipulation.

The analysis of the results was as follows. First, the group data was put into a dummy variable, which was a nominal value. The question answers were in numbers ranging from one to nine, and were considered as interval variables. Second, the means of each group and their corresponding standard deviations were calculated. These can be seen in Table 12. Third step was to see if the mean sequence and separation was in accordance with the pre-determined expectations.

| Group | Mean | N | Std. Deviation |
|---------------|--------|----|----------------|
| 0 Human being | 3,1667 | 6 | 2,31661 |
| 1 Chatbot | 4,5000 | 6 | 1,97484 |
| 2 Non-intro | 4,2857 | 7 | 1,79947 |
| Total | 4,0000 | 19 | 2,00000 |

Table 12: Means Comparison Pre-test 1.

As can be seen, the results from the first pre-test showed results which were not in line with the expected and required results to continue. The employee introduction was regarded as less human ($\bar{x} = 3,167$) than the actual chatbot ($\bar{x} = 4,500$) or undefined intro group ($\bar{x} = 4,286$). Therefore it could be concluded that the manipulation was not a success in this experiment.

4.6.2 Results Second Pre-Test

The second pre-test was similar to the first pre-test. It had the same questions and analysis. Only the manipulation has been made bigger, to enhance manipulation success.

| Group | Mean | N | Std. Deviation |
|---------------|--------|----|----------------|
| 0 Human being | 4,0000 | 7 | 2,70801 |
| 1 Chatbot | 2,8000 | 5 | 1,78885 |
| 2 Non-intro | 3,0000 | 7 | 1,73205 |
| Total | 3,3158 | 19 | 2,10957 |

Table 13: Means Comparison Pre-test 2.

As can be seen in Table 13, the group mean sequence and separation is more in line with the expected and required results. The sequence of the means is in line with the manipulation intention. Namely, the employee group has been rated most human-like ($\bar{x} = 4,000$), the chatbot group has been rated the least human ($\bar{x} = 2,800$) and the undefined intro group lays in the middle of both the other groups ($\bar{x} = 3,000$).

With these results, the manipulation has been a success, and the manipulation question answers confirm this success in the treatment. The sequence is correct, as it is in accordance with Table 9. The scattering is acceptable, as a greater mean separation would have been agreeable, but not strictly necessary. Also, a greater separation might be expected in the main experiment, as the sample size increases.

Based on these results it was decided to put the questionnaire to the next research phase, which entails the main experiment.

4.7 Power Analysis

In order to determine the required sample-size of the questionnaire, a power analysis has been conducted (Erdfelder & Buchner, 2003). For executing the calculations the software G*Power was used, version 3.1.9.4.

The nature of the experiment involves a one-way design with three groups. Therefore an ANOVA test will be used. For this reason, the parenting test family used for this Power Analysis will be an f-test.

The input variables used to calculate the total sample size can be seen in Table 14.

4. Research Design

| | |
|--------------------------------------|------|
| Effect Size | 0.25 |
| Error Probability (α) | 0.05 |
| Power (1- β error probability) | 0.80 |
| Number of groups | 3 |
| Total Sample Size (n) | 159 |

Table 14: Overview of the used input parameters for power analysis.

The effect size was set on 0.25. This involved a medium effect size, and is based on the rule of thumb (Wassertheil & Cohen, 2006). Since no relevant data on the effect size is available from previous research nor from similar articles, was the rule of thumb method used. The error probability – significance level or p -value – was set to .05. This is corresponding to a .95 significance level which is generally accepted in behavioral sciences. The power was set on .80, which means an 80% chance to find an effect. .80 Is also a generally accepted power estimation in behavioral sciences (Wassertheil & Cohen, 2006). The number of groups was set to 3, as this corresponds with the research setup.

Based on these input parameters, G*Power calculated the necessary sample size on a total of 159 respondents. This equals 53 respondents per condition.

4.8 Measures

The measures which determine the outcome of this experiment were asked after the manipulation. The measures were in the form of questions, which were answered by the respondents themselves. This placed the measures of this research in the self-report measure category, and made the question cognitive self-reports.

Questions in the experiment were aligned with predefined questions which apposed to the used constructs. All the constructs in this research are measured with previously used and proved reliable questions. This ensures a high reliability (Cronbach's Alpha).

All questions in the main experiment were measured on a 7-point scale. This could either be a Likert scale or a semantic differential scale. This is in line with the response measures as they were obtained from theory, and it shows consistency throughout the experiment, which is favorable for participants.

The following table shows more information about the used measures.

4. Research Design

| Construct | Number of items | Scale |
|--------------------------------|-----------------|-------------------------------|
| Social Presence | 5 | 7-point Likert |
| Perceived Humanness | 6 | 7-point semantic differential |
| Service Encounter Satisfaction | 3 | 7-point Likert |
| Manipulation Check | 1 | 7-point semantic differential |

Table 15: Overview of the used measures and detailed information.

4.9 Participants

The target group of this research is: ‘users and potential users of web care chatbots’. This could be explained as all people who use the internet, and have the possibility to get in touch with a chatbot. This is a fairly large sample group, and knows very little restrictions which could exclude respondents from participating. The fact that this survey was web based was also the threshold for participating. This means that the fact that people were able to reach and fill in the survey via a web browser automatically made them eligible for participation, since by doing this, they would fall in the target group.

Nevertheless, in order to get sound and representative results, a divers sample was required. Therefore the distribution was aimed at reaching divers people in terms of age, gender and education level.

4.10 Data gathering

The survey for this research was set out with online questionnaires, which could be filled in over the internet. Only a modern web browser was needed to take part in this research. Due to the fact that a movie needed to be watched for a successful participation, some participants’ response was dropped. Mainly browsers from Microsoft (Internet Explorer and Edge) were unable to handle the contained movies. Therefore it was strongly recommended to every participant to use Google Chrome or a Chromium based browser. This was clearly stated to every invitation to participate to the research. When people were unable to see the movie, they were withheld from completing the questions, as this would pollute research results. Instead, these participants were required to restart the survey in another browser, accompanied with instructions how to do so.

4. Research Design

The reasons to choose for this web based approach is due to the low intensity spreading and higher numbers of more diverse respondents. It was easier to reach a large sample group via web based distribution. This increased the amount of respondents, and also the diversity of the sample. This diversity was in accordance with the required spread for the research.

4.10.1 Distribution

The questionnaire was distributed via separate channels, in order to get the most divers sample group possible. Respondents were approached either via electronical means or personal. In both cases they were provided with a specific link to the survey, so it could be traced back via which channel and thus which target group the respondents came.

Distribution happened via three ways, and three general target groups were approached. The main distribution channels were: electronical link, personal approach, and network distribution.

With the electronical link, people were invited via a personal message which was sent to them. This could happen individually or in a group. Personal approach is when participants were invited by means of personal contact. They were approached and asked to take part in the research. As a reminder, they were given a paper with instructions to enter the research. Network approach is when the invitation to the survey was posted on a network place, including several social media or company's intranet.

The three main target groups included university students, relatives and business professionals. University students are students which are either in their bachelor or master study. Relatives are people in close relation to this thesis's author. Business professionals are employees of KPMG The Netherlands, which have filled in the survey on personal behalf.

In Chapter 5, Table 16 on page 60, an overview can be seen for each of the target groups with their corresponding distribution method. In this table, also the segment rates of each of the target groups, as well as their response rate can be seen.

4.11 Procedure

The survey procedure contained a web based survey for participants to fill in. The survey started with a short introductory message. Explanations were made about the technical procedure of the survey, such as the length, preferred browser use, and the fact that it contained a movie, but no explanations regarding the content of the research was made. This was to prevent possible confounding variables due to the fact that participants had prior knowledge to the research topic. Also because it was explained in advance that participants were going to watch a movie, but no audio devices were necessary.

The first questions of the survey were basic demographical questions. Questions such as age, gender and educational level were asked. Also the proficiency in the English level was measured. Participants were asked to rate their own skills in English on a 5-scale level. This question was interesting since the whole survey was in English, and it was expected that a large group of non-native English speakers would take part in the research. This question could check if potential distorted research results could have originated from a language barrier.

After this section, again some instructions were given over the layout of the chat vignette. Detailed information about the video was given, and an example of the chat messages were shown.

Directly after, the vignette part started. First, consumers were introduced to the situation they were going to be exposed to. The background and nature of the situation was explained, and after, participants could continue to the video. The contents of the conversation shown in the video can be read in the appendix (8.2).

After the video, the construct questions started. All the questions related to one construct were shown in table form on the same page. Each construct was shown on a new page. After the three construct pages, the manipulation check question was shown, and a final closing message was consecutively shown.

More detailed information about the survey layout can be found in the appendix (8.1).

4.12 Analysis

Data analysis was done by using the SPSS Statistics package. The following functions were used, in order:

1. Significance analysis for manipulation check (ANOVA)
2. Reliability analysis for all three constructs (Cronbach's Alpha)
3. Aggregating construct measures
4. Significance analysis for construct measures (ANOVA)
5. Mean comparison for construct measures
6. Regression Analyses

First, the significance of the manipulation check was analyzed. This was important because it shows the success of the manipulation of the experiment. A significant result ($p < .05$) is important because it shows that the participants have experienced the treatment correctly, and that conclusions based on the constructions measure are grounded by a successful experiment.

Second, the measures for the constructs were checked on reliability. For this the Cronbach's Alpha calculation was used. The acceptance rate was that the Cronbach's Alpha was above .70, as this is a generally accepted level for an acceptable reliability (Bland & Altman, 1997; Gliem & Gliem, 2003).

Thirdly, when the Cronbach's Alpha was above the accepted level, the constructs were aggregated. This was done by taking the numerical mean of all the construct's items. Note that this action was only undertaken if the Cronbach's Alpha was sufficient.

In order to draw conclusions, the final step was undertaken. Here the means of the experiment conditions were compared. Also with the use of regression, the impact of the constructs in the relationship could be established.

4.13 Ethics

This research has limited impact on breach of generally accepted ethical principles. This is due to the research design and subject, which are not near a boundary or grey area of what is generally accepted as right or wrong. Therefore only basic ethic practices are taken into account for this research. These involve the informed consent, letting the participant know the fact that they are taking part in a research experiment, and letting them know the general objective and purposes of the research. Also, the anonymity and confidentiality of the participants and their personal data will be safeguarded. Results will be based on generalized results, and not on personal factors.

The deception effect is going to be minimized for all participants. Participants are not going to be actively deceived, but for the sake of the research and experiment outcome, some information will deliberately be held back. This includes the information concerning the research group the participant is on, or the actual design of the research progress.

The effects and risk of breaching ethical principles will be minimal. Nevertheless will this research and the author be constantly aware of ethics issues and their impact on this research. This includes a continuing ethics assessment and having ethics principles in mind when designing, executing and finalizing this research.

5 RESULTS

This chapter of the thesis will elaborate about the results and findings of the experimental research. All used syntaxes from SPSS can be found back in the appendix (8.4.3).

5.1 Error Rate and Dropout Rate

In total 277 respondents have filled in the survey ($n = 277$). However, a great percentage dropped out, mainly in two cases. Firstly, people dropped out because they terminated the survey while they were making it ($n = 74$; 26.71%). This dropout rate could have various reasons, which this thesis will not try to clarify. Secondly, there is an error rate. This includes the people who were not able to watch the movie, ($n = 53$; 19.13%) and therefore were not allowed to answer the construct questions. These people were invited to re-take the survey with instructions to make it work.

Between the error rate and dropout rate, some overlap exists. However it can be seen that when people were able to actually watch the movie, they almost always finished the survey. Only 2 persons still dropped out after being able to watch the movie (.72%).

Both the error rate and dropout rate participants are not taken into account in the final calculation of results. Therefore this dataset ends with 159 active participants (57.40%) to base its calculations on. This amount corresponds to the predetermined quantity in a power analysis to determine the needed number of participants. Data gathering was stopped as soon as this number was reached.

In Table 16 the used distribution methods and target groups, as specified in Chapter 4, can be seen. The distribution rate shows the segmentation in the total sample ($n = 277$), and the response rate shows the response calculated as a percentage from the total group of addressed people. Two remarks on this table are that the percentages are based on links respondents used to enter the survey, and not on the dataset. Second, the calculation from the response rate of relatives was not possible due to an unknown size of the total addressed group.

From this part onward, all shown calculations and results are based on this set of 159 participants.

5. Results

| Distribution method | Target group | Distribution Rate | Response rate |
|--|------------------------|-------------------|----------------|
| Electronical Link; Network distribution | University Students | 41% | 51% |
| | Relatives | 17% | <i>Unknown</i> |
| Personal Approach; Network distribution | Business Professionals | 40% | 34% |
| Other | Other | 2% | |

Table 16: Overview of distribution methods and target groups for the experiment.

5.2 Basic Demographical Information

The dataset consisted of 98 males (61.64%) and 61 females (38.36%). The sample was fairly high educated, but the educational levels of respondents still showed a decent spread for representative results. An extended table showing all the education levels can be found in the appendix.

The age spread in the sample was also adequate. Average respondents' age was 28 ($\bar{x} = 28.12$; $\sigma = 8.07$). The minimum age was 20, and the maximum age in the dataset was 70. A more detailed view of the age distribution can be found in the appendix.

To overcome bias due to the existence of language barriers, a self-control questions was inserted in the questionnaire. People had to rate their own mastery of the English language. The response on this question did not raise any doubts for the presence of a bias, since the vast majority rated their English skills as 'very good' (64.15%) and an even larger group 'above average' (94.34%). This does not imply any verdict about the participants' English skills, but it gives an indication of the general level of English understanding by the participants, and showed that the language barrier posed no problems.

5.3 Manipulation Check

In the main experiment, the manipulation check question was asked at the end of the survey. Therefore minimizing potential biases. As the results from the pre-tests showed, the experiment setup presented indicated an adequate mean spread and distance between the experiment conditions, which was stated in Table 9 (p. 47). Results from similar tests in the main experiment showed to following distribution.

5. Results

| Group | Mean | N | Std. Deviation |
|---------------|--------|-----|----------------|
| 0 Human being | 4.0566 | 53 | 2.307 |
| 1 Chatbot | 3.0556 | 54 | 1.827 |
| 2 Non-intro | 3.9038 | 52 | 1.624 |
| Total | 3.6667 | 159 | 1.980 |

Table 17: Manipulation Check Means Comparison Main Experiment.

| | Sum of Squares | df | Mean Square | F | Sig. |
|----------------|----------------|-----|-------------|-------|--------------|
| Between Groups | 31.151 | 2 | 15.575 | 4.131 | .018* |
| Within Groups | 588.183 | 156 | 3.770 | | |
| Total | 619.333 | 158 | | | |

Table 18: Manipulation Check One-Way Anova Table.

5.4 Scale Reliability

In order to determine the scale reliability of the construct measures, the Cronbach's Alpha measure was used. The following table shows the results for each construct.

| Construct | Cronbach's Alpha (α) |
|--------------------------------|-------------------------------|
| Social Presence | .906 |
| Perceived Humanness | .750 |
| Service Encounter Satisfaction | .886 |

Table 19: Calculated Cronbach's Alpha for all the constructs.

5.5 Significance

After a sufficient Cronbach's Alpha, the construct measures were combined by taking the numerical average. Anova tables for each of the constructs then reports the following.

| | Sum of Squares | df | Mean Square | F | Sig. |
|----------------|----------------|-----|-------------|-------|----------------|
| Between Groups | 23.371 | 2 | 11.686 | 8.383 | .000*** |
| Within Groups | 217.465 | 156 | 1.394 | | |
| Total | 240.836 | 158 | | | |

Table 20: Anova table for the Social Presence Construct.

5. Results

| | Sum of Squares | df | Mean Square | F | Sig. |
|----------------|----------------|-----|-------------|-------|--------------|
| Between Groups | 3.236 | 2 | 1.618 | 3.500 | .033* |
| Within Groups | 72.104 | 156 | .462 | | |
| Total | 75.340 | 158 | | | |

Table 21: Anova table for the Perceived Humanness Construct.

| | Sum of Squares | df | Mean Square | F | Sig. |
|----------------|----------------|-----|-------------|-------|------|
| Between Groups | 3.189 | 2 | 1.594 | 1.151 | .319 |
| Within Groups | 216.008 | 156 | 1.385 | | |
| Total | 219.196 | 158 | | | |

Table 22: Anova table for the Service Encounter Satisfaction Construct.

5.6 Results

The following tables show the means from the different construct measures.

| Group | Mean | N | Std. Deviation |
|---------------|--------|-----|----------------|
| 0 Human being | 5.1434 | 53 | 1.31989 |
| 1 Chatbot | 4.3185 | 54 | 1.08124 |
| 2 Non-intro | 5.1115 | 52 | 1.12819 |
| Total | 4.8528 | 159 | 1.23462 |

Table 23: Mean overview Social Presence Construct.

| Group | Mean | N | Std. Deviation |
|---------------|--------|-----|----------------|
| 0 Human being | 4.8239 | 53 | .70439 |
| 1 Chatbot | 4.4938 | 54 | .54750 |
| 2 Non-intro | 4.7532 | 52 | .77227 |
| Total | 4.6887 | 159 | .69053 |

Table 24: Mean overview Perceived Humanness Construct.

| Group | Mean | N | Std. Deviation |
|---------------|--------|-----|----------------|
| 0 Human being | 6.1572 | 53 | 1.10671 |
| 1 Chatbot | 6.0432 | 54 | 1.14384 |
| 2 Non-intro | 6.3846 | 52 | 1.27552 |
| Total | 6.1929 | 159 | 1.17784 |

Table 25: Mean overview Service Encounter Satisfaction.

5.7 Regression

Table 26 measures the impact of the three constructs among themselves, in order to express a relationship between them. Here, the construct service encounter satisfaction was placed as a dependent variable, based on perceived humanness and social presence.

The control variables have been added, and in four consecutive models, the various effects of both the independent variables have been measured both individually and combined.

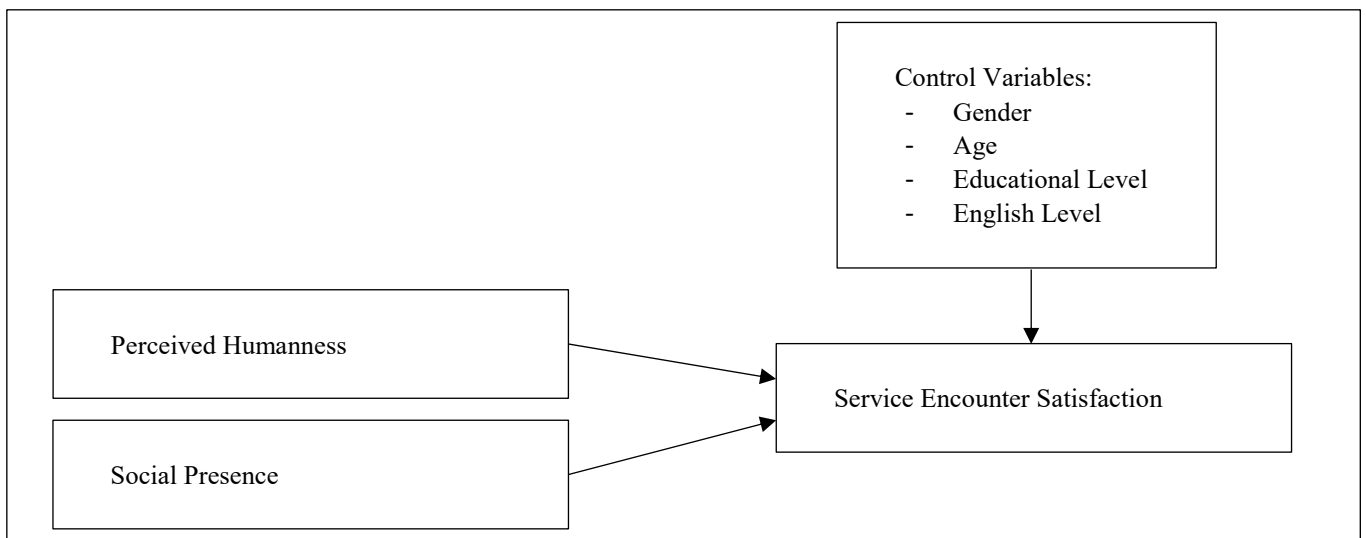


Figure 3: Amended Conceptual Model for Regression purposes.

The following tables, table 27 until 32, show a dummy regression analysis to measure the effect for each of the introductions on the measured constructs. In these models the original conceptual model has been used, as shown in Figure 2.

In order to turn the three experiment groups into a regression, they have been turned into dummy variables. The human being group has in all three regressions used as control group.

5. Results

5.7.1 Regression analyses among measured constructs

| Variables | Model 1 | | | Model 2 | | | Model 3 | | | Model 4 | | |
|------------------------------|---------|---------|----------|---------|---------|-----------------|---------|---------|-----------------|---------|---------|-----------------|
| | SE B | β | <i>p</i> | SE B | β | <i>p</i> | SE B | β | <i>p</i> | SE B | β | <i>p</i> |
| Control variables | | | | | | | | | | | | |
| Age | .012 | -.136 | .091 | .011 | -.109 | .137 | -.010 | -.066 | .325 | .010 | -.068 | .312 |
| Educational level | .055 | -.054 | .500 | .050 | -.051 | .485 | -.027 | -.040 | .553 | .045 | -.041 | .541 |
| Gender | .191 | .127 | .112 | .176 | .061 | .404 | .168 | .070 | .294 | .160 | .059 | .377 |
| English level | .147 | -.063 | .431 | .134 | -.027 | .711 | -.097 | -.053 | .430 | .122 | -.044 | .510 |
| Independent variables | | | | | | | | | | | | |
| Social Presence | | | | .070 | .424 | .000 *** | | | | .082 | .118 | .168 |
| Perceived Humanness | | | | | | | .950 | .557 | .000 *** | .145 | .483 | .000 *** |
| Significance | | | | | | | | | | | | |
| R ² | .043 | | | .216 | | | .345 | | | .353 | | |
| Adjusted R ² | .018 | | | .191 | | | .324 | | | .328 | | |

Table 26: Coefficients Table for simple regression. Dependent variable: Service Encounter Satisfaction.

5. Results

5.7.2 Dummy variable regression analysis for Social Presence

| Model Summary | | | | |
|---------------|------|----------|-------------------|----------------------------|
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | ,191 | ,037 | ,012 | 1,22742 |
| 2 | ,370 | ,137 | ,103 | 1,16921 |

Table 27: Regression model summary for Social Presence.

| Coefficients | | | | | | | | |
|--|--------------------------|-----------------------------|------------|---------------------------|--------|----------------|---------------------------------|-------------|
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | 95,0% Confidence Interval for B | |
| | | B | Std. Error | Beta | | | Lower Bound | Upper Bound |
| 1 | (Constant) | 10,109 | 5,221 | | 1,936 | ,055 | -,205 | 20,422 |
| | Age | -,010 | ,012 | -,064 | -,801 | ,425 | -,034 | ,014 |
| | Educational Level | -,006 | ,058 | -,008 | -,096 | ,923 | -,119 | ,108 |
| | Gender | ,393 | ,201 | ,155 | 1,953 | ,053 | -,004 | ,791 |
| | English level | -,164 | ,155 | -,085 | -1,062 | ,290 | -,470 | ,141 |
| 2 | (Constant) | 10,249 | 4,999 | | 2,050 | ,042* | ,373 | 20,125 |
| | Age | -,013 | ,012 | -,084 | -1,083 | ,281 | -,036 | ,011 |
| | Educational Level | ,004 | ,055 | ,005 | ,069 | ,945 | -,105 | ,112 |
| | Gender | ,404 | ,192 | ,160 | 2,109 | ,037* | ,026 | ,783 |
| | English level | -,159 | ,148 | -,083 | -1,078 | ,283 | -,452 | ,133 |
| | Dummy variable Chatbot | -,837 | ,227 | -,322 | -3,692 | ,000*** | -1,284 | -,389 |
| | Dummy variable Non-intro | -,022 | ,231 | -,008 | -,094 | ,925 | -,479 | ,435 |
| a. Dependent Variable: Social Presence | | | | | | | | |

Table 28: Regression model for Perceived Humanness.

5. Results

5.7.3 Dummy regression analysis for Perceived Humanness

| Model Summary | | | | |
|---------------|------|----------|-------------------|----------------------------|
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | ,165 | ,027 | ,002 | ,68983 |
| 2 | ,269 | ,072 | ,036 | ,67808 |

Table 29: Regression model summary for Perceived Humanness.

| Coefficients | | | | | | | | |
|--|--------------------------|-----------------------------|------------|---------------------------|--------|--------------|---------------------------------|-------------|
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | 95,0% Confidence Interval for B | |
| | | B | Std. Error | Beta | | | Lower Bound | Upper Bound |
| 1 | (Constant) | 5,544 | 2,934 | | 1,890 | ,061 | -,252 | 11,341 |
| | Age | -,011 | ,007 | -,126 | -1,560 | ,121 | -,024 | ,003 |
| | Educational Level | -,011 | ,032 | -,026 | -,325 | ,746 | -,074 | ,053 |
| | Gender | ,145 | ,113 | ,103 | 1,284 | ,201 | -,078 | ,369 |
| | English level | -,020 | ,087 | -,019 | -,236 | ,814 | -,192 | ,151 |
| 2 | (Constant) | 5,670 | 2,899 | | 1,956 | ,052 | -,057 | 11,398 |
| | Age | -,012 | ,007 | -,136 | -1,695 | ,092 | -,025 | ,002 |
| | Educational Level | -,007 | ,032 | -,017 | -,213 | ,831 | -,070 | ,056 |
| | Gender | ,149 | ,111 | ,105 | 1,343 | ,181 | -,070 | ,369 |
| | English level | -,021 | ,086 | -,019 | -,241 | ,810 | -,190 | ,149 |
| | Dummy variable Chatbot | -,328 | ,131 | -,226 | -2,495 | ,014* | -,588 | -,068 |
| | Dummy variable Non-intro | -,042 | ,134 | -,029 | -,313 | ,754 | -,307 | ,223 |
| a. Dependent Variable: Perceived Humanness | | | | | | | | |

Table 30: Regression model for Perceived Humanness.

5. Results

5.7.4 Dummy regression analysis for Service Encounter Satisfaction

| Model Summary | | | | |
|---------------|------|----------|-------------------|----------------------------|
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | ,208 | ,043 | ,018 | 1,16695 |
| 2 | ,250 | ,062 | ,025 | 1,16286 |

Table 31: Regression model summary for Service Encounter Satisfaction.

| Coefficients | | | | | | | | |
|--------------|--------------------------|-----------------------------|------------|---------------------------|--------|--------------|---------------------------------|-------------|
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | 95,0% Confidence Interval for B | |
| | | B | Std. Error | Beta | | | Lower Bound | Upper Bound |
| 1 | (Constant) | 10,451 | 4,964 | | 2,106 | ,037* | ,646 | 20,257 |
| | Age | -,020 | ,012 | -,136 | -1,701 | ,091 | -,043 | ,003 |
| | Educational Level | -,037 | ,055 | -,054 | -,677 | ,500 | -,145 | ,071 |
| | Gender | ,306 | ,191 | ,127 | 1,600 | ,112 | -,072 | ,684 |
| | English level | -,116 | ,147 | -,063 | -,790 | ,431 | -,406 | ,174 |
| 2 | (Constant) | 9,866 | 4,972 | | 1,984 | ,049* | ,043 | 19,688 |
| | Age | -,023 | ,012 | -,156 | -1,936 | ,055 | -,046 | ,000 |
| | Educational Level | -,036 | ,055 | -,053 | -,667 | ,506 | -,144 | ,071 |
| | Gender | ,311 | ,191 | ,129 | 1,630 | ,105 | -,066 | ,687 |
| | English level | -,098 | ,147 | -,053 | -,667 | ,506 | -,389 | ,192 |
| | Dummy variable Chatbot | -,108 | ,225 | -,044 | -,481 | ,631 | -,554 | ,337 |
| | Dummy variable Non-intro | ,282 | ,230 | ,113 | 1,223 | ,223 | -,173 | ,736 |

a. Dependent Variable: Service Encounter Satisfaction

Table 32: Regression model for Service Encounter Satisfaction.

5.8 Hypothesis Testing

In this section, the hypothesis as created in chapter 3 will be accepted or rejected, based on the statistics presented in this chapter.

The testing process is done in five steps:

1. The hypothesis is stated, as it was created in chapter 3.
2. The hypothesis is again stated, but now in mathematical terms. Also the required p -value is shown.
3. The results as presented in this chapter are shown.
4. The actual p -value, as calculated in this chapter, is presented.
5. Based on the preceding steps, the conclusion is formed about the hypothesis.

5. Results

| | | |
|---|--|---|
| 1 | H₁ : A chatbot with an identity revealing self-presentation will yield a higher experienced social presence than a chatbot with a neutral self-presentation. | |
| 2 | $\bar{x}_{(Social\ Presence; Group\ 1)} > \bar{x}_{(Social\ Presence; Group\ 2)}$ | $p = .05$ |
| 3 | $\bar{x}_{(Social\ Presence; Group\ 1)} = 4.3185$ | $\bar{x}_{(Social\ Presence; Group\ 2)} = 5.1115$ |
| 4 | $p < 0.000$ | |
| 5 | H ₁ is rejected. | |

Table 33: Hypothesis testing process for hypothesis 1.

| | | |
|---|--|---|
| 1 | H₂ : A chatbot with an identity revealing self-presentation will yield a higher experienced perceived humanness than a chatbot with a neutral self-presentation. | |
| 2 | $\bar{x}_{(Perceived\ Humanness; Group\ 1)} > \bar{x}_{(Perceived\ Humanness; Group\ 2)}$ | $p = .05$ |
| 3 | $\bar{x}_{(Perceived\ Humanness; Group\ 1)} = 4.4938$ | $\bar{x}_{(Perceived\ Humanness; Group\ 2)} = 4.7532$ |
| 4 | $p = .033$ | |
| 5 | H ₂ is rejected. | |

Table 34: Hypothesis testing process for hypothesis 2.

| | | |
|---|--|--------------------------------------|
| 1 | H₃ : A chatbot with an identity revealing self-presentation will yield a higher experienced service encounter satisfaction than a chatbot with a neutral self-presentation. ³ | |
| 2 | $\bar{x}_{(SES; Group\ 1)} > \bar{x}_{(SES; Group\ 2)}$ | $p = .05$ |
| 3 | $\bar{x}_{(SES; Group\ 1)} = 6.0432$ | $\bar{x}_{(SES; Group\ 2)} = 6.3846$ |
| 4 | $p = .319$ | |
| 5 | Fail to accept nor reject H ₃ due to statistical insignificance. | |

Table 35: Hypothesis testing process for hypothesis 3.

³ For readability purposes Service Encounter Satisfaction is abbreviated to SES.

5. Results

6 CONCLUSION, LIMITATIONS & FUTURE RESEARCH

6.1 Discussion

Results from this experiment show interesting results. This section will give an interpretation of the results, in order to understand the meaning of the outcome. The first part will explain the meaning of the retrieved results. The outcome will be coupled back to the theory in order to derive meaningful information. The last section will elaborate on the justification of this results, wherever possible.

By means of an experimental research design this research has the capability of showing a relationship. By looking at the results, it can be seen that web care consumers prefer a conversation with a real human being over that with a chatbot. This is in line with the expectations. However, it can also be seen that users prefer an undefined conversational partner over a chatbot. Although users are aware of the fact that they do not know the identity of the partner they are talking with, as shown by the manipulation check, they still prefer this partner. Users thus prefer an unknown partner over a chatbot, even though they are aware that their partner could still be a chatbot.

This rationale holds for both the rating of the contentment of the conversation – social presence – as with the comfort of the conversational partner – perceived humanness. However, for the last construct, service encounter satisfaction, another logic seems to hold. Since questionnaire responses on this construct are non-significant, we cannot speak of a serious difference between the various introductions. Therefore logic seems to be true that users are indifferent of the conversational partner they speak with, regarding the satisfaction with the advice and the overall interaction.

Even though this research can prove a relationship between these constructs, it cannot explain nor justify why this relation exists. At first glance, these results might look counter intuitive. Why do people prefer it when they do not know how they are talking to? Scientific literature is at this moment deficient of explaining this unclarity.

However, counter intuitive results are not uncommon in this research area. The paper by Gnewuch et al. (2018) also shows results not in line with previous research. Based on the results from this and similar research, it can be said that the chatbot research area does not directly comply with social norms and expectancies from human-human interactions. Therefore we

argue that more research needs to be done in this area, to give a more complete understanding of this new technology.

Nevertheless, assumptions can be made for the understanding of this result. Why would people prefer talking to an unknown partner instead of a chatbot? Potentially it has something to do with the perceived identity. The fact that users are unaware of the real identity of the partner will make people start to make assumptions about the identity, also called: wishful thinking. People act on what they hope the identity is, because real evidence is lacking. This is reflected in the results on the manipulation check, where users indicate that they do not know the identity of their partner, but still rate the undefined partner closer to a human, than to a chatbot. Since a human conversational partner yields a higher satisfaction, automatically the satisfaction for the unknown partner will also raise.

It must be clearly stated that these attempts to declare the results are on the boundary of what this thesis can explain. This thesis has sufficient data to prove the cause and effect relationships, but cannot declare why this relationship exists. In order to do so, more research is needed to confirm hypotheses and assumptions made to explain these results, and to unravel the underlying causes.

6.2 Conclusion

This thesis research focused on answering the following research question: Which chatbot introduction yields the highest user satisfaction? By means of an experimental research setup a relationship has been established. Based on statistical results based on participant's ratings on three constructs, a conclusion can be derived.

Three constructs have been used to measure: social presence, which focusses on the evaluation of the conversation; perceived humanness, which focusses on the evaluation of the conversational partner; and service encounter satisfaction, which focusses on the evaluation of the advice.

Based on this research, the following conclusions can be derived. First of all, users prefer to talk to a real human being all of the times, as this leads to a higher satisfaction. Second, if users talk with a chatbot, it is better for a chatbot to introduce itself to the user in an undefined way, instead of explicitly stating its true identity (e.g. "Hello, how can I help you?" is better than "Hello, I'm a chatbot, how can I help you?"). This undefined introductions leads to a higher rating on social presence and perceived humanness, which translates in a higher satisfaction

with the conversation and with the conversational partner. Thirdly, this thesis concludes that for the last measured construct, service encounter satisfaction, users are indifferent of the measured satisfaction. This was displayed by a non-significant difference between the measured results on this construct. So for the satisfaction with the final advice and treatment, the users do not mind if they are talking to a chatbot or a real human, nor the way this conversational partner introduces itself.

6.3 Academic Implications

The result of this thesis has various implications to the user perceptions of chatbots. Therefore this research can make a contribution to the parties designing or creating chatbots in customer care situations. But also due to the scientific nature of this research, several relevant academic implications can be made.

First of all this research addresses the fact that chatbot design is two sided. The technical part, which mainly looks from a programming perspective, and the interaction perspective, which mainly looks from an end users' points of view. As Moore et al. (2017) already stated, the technical part of designing a chatbot gets easier, due to the rapid technological developments. Designing a chatbot is not hard, but getting it accepted by users and including the right social and human elements in order to create natural interaction is key for a successful chatbot.

This thesis argues for the creation of a perspective wherein this interaction part is being seen equally important as the technical design of a chatbot. In order to do this, more research in chatbot interaction is necessary.

This thesis has used various theories from the social and behavioral sciences discipline to base its theoretical framework on, mainly due to the lack of existing theories in chatbot research. These social and behavioral theories are not one-on-one applicable to chatbot interaction, but with the help of the CMC theories, they could be relayed to make them suitable for chatbot research. This, however, does not mean that the theories form a tailored fit to this research. The use of theories from another research discipline is due to the scarcity and is not a preferred method. This misfit in theory is reflected by the fact that this research, among other research into chatbots, found counter intuitive results which are not aligned with the theory.

Therefore, this thesis argues that more research and theory building is necessary, and that this research field should be regarded as independent, and not as a component of social and

behavioral sciences, or technical research disciplines. The social and interaction side of chatbots is specialized, and is not conform to previous research in comparable, other disciplines. It should therefore also not be treated as such. This does imply that more research to chatbot interaction is needed.

Altogether, this thesis argues for a perspective wherein chatbot research is unfolded in two different research disciplines: the technical design of a chatbot, and the human interaction side, where this latter part should not be based on social and behavioral sciences' theory, but should be regarded as an independent research field. In order to accommodate the enormous and promising trend of chatbots into literature, more research on both the two sides should be conducted.

6.4 Practical Implications

The practical implications this thesis makes are aimed at improving current chatbot design. Practical implications therefore mainly focus on chatbot designers, and gives them concrete advice to increase the interaction satisfaction with end-users.

The way a chatbot introduces itself has a significant effect on the course of the conversation, and the experienced satisfaction from users. Based on this thesis, it can be advised to chatbot creators to focus on this important part of the conversation. Test results show that users experience the highest levels of satisfaction when they are talking to a real human being, instead of a chatbot. But, whenever they talk to a chatbot, it is better to keep the real identity enclosed. Therefore this thesis advises not to introduce a chatbot as a chatbot, but keep the introduction more indefinite and generic. This allows users to only assume the real identity of their conversational partner, which yields a higher satisfactory score. This holds for both the satisfaction with the conversation itself, as with the conversational partner.

For a deeper level of advice, this section is from this point forward split up for several users practically involved in chatbot interaction.

For companies interested in implementing a chatbot, this thesis advises the following. Implementation of a chatbot in a customer care situation can have positive effects, however this process needs to be undertaken thoughtfully and with care. Nowadays, people still prefer chatting with a real human being over a chatbot in regards of the best conversation and happiness with the conversational partner. But for the satisfaction with the given advice, users are indifferent between a human being and a chatbot.

6. Conclusion, Limitations & Future Research

In customer care environments involving easy and simple transactions, as referred to before as a FAQ situation, a chatbot can be an excellent solution. In these situations the focus is on answering a question and getting advice, and not on having a pleasant conversation. Implementing a chatbot in these situations can be cost saving for the company, while keeping a sufficient level of customer satisfaction. The customer satisfaction might experience a little drop due to a lower satisfaction with the conversation, but since the main goal is giving advice, this drop will be acceptable. Weighed off against the cost saving potential, the tradeoff is acceptable.

For chatbot designers, this thesis advises the following. If a chatbot is being designed for conversational purposes, such as the earlier named Mitsuku, designers should be reticent. Conversing with a chatbot implies lower satisfaction than talking with a real human being. Especially in situations where satisfaction with either the conversation itself or the conversational partner is of importance, these effects are present. This is especially the case for social bots. However, in these situations, this thesis advises to let the chatbot introduce itself in a neutral, thus undisclosed, way. This will yield a higher satisfaction for the user.

Important side note to this advice is that it is only valid when a user cannot derive the actual identity of the conversational partner from anything else than the introduction. This is only the case if the technology is mature enough to perfectly imitate a human being and if leading up to the chat no information about the identity is provided.

6.5 Limitations

This thesis has tried to extensively research the defined research question and tried to find answers. However, this research was also bound by various constraints. This has limited the outcome of the research. This paragraph will describe various limitations this research has.

First of all, this research made use of a vignette approach. This means that in the experiment a video of a chatbot conversation was used. This eliminates confounding variances in the research, but also limits the feeling of real chatbot interaction. The limited interaction possibilities in this research form a limitation.

Also the focus on the introductory message forms a limitation of this research. The underlying assumption is that the user identifies the conversational partner on the introduction message, and not on the further course of the conversation. This implies the assumption that in the course of the conversation the differences between chatbots and humans are nonexistent,

and therefore the conversation itself forms no of identification possibility of the conversational partner.

This point of view is however not aligned with today's technology, and this thesis anticipates at improving chatbot technology which make it impossible to distinguish chatbots from humans based on conversation messages.

Still, this perspective forms a limitation of this research since nowadays people will be able to recognize their conversational partner based on conversational data.

The last major limitation of this thesis can be found in the way the theoretical framework has been constructed. The theoretical framework is based on theories and research which finds it origin in social and behavioral sciences. These theories are relayed to chatbot with research in the field of CMC. This however does not make the theoretical framework unquestionable and completely fitting to this research. This approach, due to the lack of actual chatbot theories, form a limitation of this thesis.

6.6 Future Research

This thesis has already argued that more research is necessary to give a more profound review of chatbot interaction. Every scientific research into the impact of various interaction elements in chatbots will therefore be seen as relevant follow-up research. This thesis brings forward four specific research areas interesting for future research.

Firstly, other sectors. This thesis focusses on customer care chatbot interaction, but more applications of chatbots exist. What is the impact of various chatbot introductions in these sectors? A replication research can be held with a different conversation topic to show effects in other areas.

Secondly, the justification for experiment results. As said, this thesis has conducted an experimental research to discover the relationship between the constructs, but it cannot declare why this relationship exists. An exploratory research into this would be interesting angle for future research, as it both builds forward on results from this thesis, and as well creates new theory for chatbot research.

Thirdly, an interesting next research topic can be where users base the perceived identity of their conversational partner on. Right now, this research has examined the effect of the introduction of a chatbot. But this thesis's research is building on previous research which has all examined a specific factor in chatbot interaction. An all-encompassing research into the

6. Conclusion, Limitations & Future Research

effect size of all these factors is an interesting step forward. This would first of all put this research into context, and determine the effect of this thesis's result. This thesis examined what the reaction is from people to various introductions, but such a research would also examine the relative effect of this reaction in comparison to, for example, use of language or delay time. It would provide a comparing overview of the current state of research and the mutual relationships between them.

As final future research possibility, this thesis advises the following. As this thesis anticipates on technology which can perfectly imitate a human being, it also acknowledges current technologies are not that advanced yet. At this moment it could be perfectly well possible that users have a conversation with a neutral introducing chatbot, but later on in the conversation realize they are actually talking to a chatbot, because, for example the chatbot does not understand figurative way of speech completely. What would the user reaction be in such a case? This research would examine more the ethical aspect of chatbot interaction, and especially focus on the deception element.

6. Conclusion, Limitations & Future Research

7 BIBLIOGRAPHY

- Anderson, S., Pearo, L. K., & Widener, S. K. (2008). Drivers of service satisfaction: Linking customer satisfaction to the service concept and customer characteristics. *Journal of Service Research, 10*(4), 365–381. <https://doi.org/10.1177/1094670508314575>
- Appel, J., von der Pütten, A., Krämer, N. C., & Gratch, J. (2012). Does Humanity Matter? Analyzing the Importance of Social Cues and Perceived Agency of a Computer System for the Emergence of Social Reactions during Human-Computer Interaction. *Advances in Human-Computer Interaction, 2012*, 1–10. <https://doi.org/10.1155/2012/324694>
- Aronson, E., & Carlsmith, J. M. (1968). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (2nd Editio, Vol. 2, pp. 1–79). Addison-Wesley Publishing Company.
- Aspect. (2016). 2016 Aspect Consumer Experience Index. *Aspect Consumer Experience Index*, 1–17. Retrieved from https://www.aspect.com/globalassets/2016-aspect-consumer-experience-index-survey_index-results-final.pdf
- Atzmüller, C., & Steiner, P. M. (2010). Experimental vignette studies in survey research. *Methodology, 6*(3), 128–138. <https://doi.org/10.1027/1614-2241/a000014>
- Barger, P. B., Grandey, A. A., Barger, P. B., & Grandey, A. A. (2006). Service with a Smile and Encounter Satisfaction : Emotional Contagion and Appraisal Mechanisms. *Academy of Management Journal, 49*(6), 1229–1238.
- Berger, C. R., Roloff, M. E., & Walther, J. B. (2016). The International Encyclopedia of Interpersonal Communication, First Edition. Edited Social Information Processing Theory (CMC), (Cmc), 1–13. <https://doi.org/10.1002/9781118540190.wbeic0192>
- Bickmore, T. W., Caruso, L., & Clough-Gorr, K. (2005). Acceptance and usability of a relational agent interface by urban older adults, 1212. <https://doi.org/10.1145/1056808.1056879>
- Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *Bmj, 314*(7080), 572. <https://doi.org/10.1136/bmj.314.7080.572>

7. Bibliography

- Brandtzæg, P. B., & Følstad, A. (2016). Chatbots and the New World of HCI, *24*(4), 38–42. <https://doi.org/10.1145/3085558.Chatbots>
- Cahn, J. (2017). CHATBOT: Architecture, Design, and Development, 46. Retrieved from https://static1.squarespace.com/static/569293741c1210fdda37b429/t/59160b6bff7c50104e601a85/1494616940469/CHATBOT_thesis_final.pdf
- Candello, H., Pinhanez, C., & Figueiredo, F. (2017). Typefaces and the Perception of Humanness in Natural Language Chatbots, (October), 3476–3487. <https://doi.org/10.1145/3025453.3025919>
- Cassell, J. (2000). *Embodied conversational agents*. Cambridge: MIT Press. Retrieved from https://books.google.nl/books?id=tHiKZGh9t7sC&printsec=frontcover&hl=nl&source=gbs_ViewAPI&redir_esc=y#v=onepage&q&f=false
- Christ, T. J. (2007). Experimental control and threats to internal validity of concurrent and nonconcurrent multiple baseline designs. *Psychology in the Schools*, *44*(5), 451–459. <https://doi.org/10.1002/pits.20237>
- Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, *92*, 539–548. <https://doi.org/10.1016/j.future.2018.01.055>
- Cooper, S. B., & Leeuwen, J. Van. (2013). Computing Machinery and Intelligence. *Alan Turing: His Work and Impact*, *59*(236), 551–621. <https://doi.org/10.1016/b978-0-12-386980-7.50023-x>
- Daft, R. L., & Lengel, R. H. (1983). Information Richness: A New Approach to Managerial Behavior and Organization Design. *Organizations As Information Processing Systems: Office of Naval Research Technical Report Series*, 73.
- Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, *22*(5), 811–817. <https://doi.org/10.1017/s1351324916000243>
- Davenport, T. H. (2015). Lessons from the cognitive front lines: Early adopters of IBM’s Watson. *The Wall Street Journal*, 1. Retrieved from <http://blogs.wsj.com/cio/2015/12/03/lessons-from-the-cognitive-front-lines-early->

7. Bibliography

adopters-of-ibms-watson/

- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On Seeing Human: A Three-Factor Theory of Anthropomorphism. *Psychological Review*, *114*(4), 864–886. <https://doi.org/10.1037/0033-295X.114.4.864>
- Erdfelder, E., & Buchner, A. (2003). G*Power a general power analysis program for Macintosh OS and MS-DOS. *Computational Statistics & Data Analysis*, *30*(4), 473. [https://doi.org/10.1016/s0167-9473\(99\)90014-2](https://doi.org/10.1016/s0167-9473(99)90014-2)
- Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the New World of HCI. *Interactions*, *24*(4), 38–42.
- Foschi, M. (2014). *Hypotheses, Operationalizations, and Manipulation Checks. Laboratory Experiments in the Social Sciences: Second Edition* (Second Edi). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-404681-8.00011-X>
- Franklin, S., & Graesser, A. (2005). Is It an agent, or just a program?: A taxonomy for autonomous agents. In *International Workshop on Agent Theories, Architectures, and Languages* (pp. 21–35). <https://doi.org/10.1007/bfb0013570>
- Gaines, B. J., Kuklinski, J. H., & Quirk, P. J. (2007). The Logic of the Survey Experiment Reexamined. *Political Analysis*, *15*(1), 1–20. <https://doi.org/10.1093/pan/mp008>
- Gartner. (2011). Gartner Customer 360 Summit 2011. Retrieved from https://www.gartner.com/imagesrv/summits/docs/na/customer-360/C360_2011_brochure_FINAL.pdf?utm_source=blog&utm_campaign=rc_blogpost
- Gartner. (2017). Top Trends in the Gartner Hype Cycle for Emerging Technologies. Retrieved February 28, 2019, from <https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/>
- Gliem, J. A., & Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. *Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education*. <https://doi.org/10.1016/B978-0-444-88933-1.50023-4>
- Gnewuch, U., Morana, S., Adam, M. T. P., & Maedche, A. (2018). Faster Is Not Always Better:

7. Bibliography

- Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction. *Proceedings of the European Conference on Information Systems (ECIS)*.
- Gourov, M. (2019). The contingent effect of sentiment and personalization of a chatbot design on the User Experience. *Thesis*, (January).
- Griffin, E. A. (2012). *A First Look At Communciation Theory*. (G. G. Sparks & A. M. Ledbetter, Eds.) (8th Editio). McGraw Hill.
- Gunawardena, C. N. (1995). Social Presence Theory and Implications for Interaction and Collaborative Learning in Computer Conferences. *International Jl. of Educational Telecommunications*, 1, 147–166.
- Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2006). Studying the acceptance of a robotic agent by elderly users. *International Journal of Assistive Robotics and Mechatronics*, 7(3), 33–43. <https://doi.org/urn:nbn:nl:ui:29-221784>
- Hill, J., Randolph Ford, W., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. *Computers in Human Behavior*, 49, 245–250. <https://doi.org/10.1016/j.chb.2015.02.026>
- Hoewe, J. (2017). Manipulation Check. *The International Encyclopedia of Communication Research Methods*, 1–5. <https://doi.org/10.1002/9781118901731.iecrm0135>
- Holtgraves, T., & Han, T. L. (2007). A procedure for studying online conversational processing using a chat bot. *Behavior Research Methods*, 39(1), 156–163. <https://doi.org/10.3758/BF03192855>
- Horton, H. (2016). Microsoft deletes “teen girl” AI after it became a Hitler-loving sex robot within 24 hours. *The Telegraph*, (March). Retrieved from <http://www.telegraph.co.uk/technology/2016/03/24/microsofts-teen-girl-ai-turns-into-a-hitler-loving-sex-robot-wit/>
- Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., ... Akkiraju, R. (2018a). Touch Your Heart: A Tone-aware Chatbot for Customer Care on Social Media. <https://doi.org/10.1145/3173574.3173989>

7. Bibliography

- Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., ... Akkiraju, R. (2018b). Touch Your Heart: A Tone-aware Chatbot for Customer Care on Social Media. *CHI*. <https://doi.org/10.1145/3173574.3173989>
- Iacobelli, F., & Cassell, J. (2007). *Ethnic Identity and Engagement in Embodied Conversational Agents*. *Intelligent Virtual Agents* (Vol. 4722). https://doi.org/10.1007/978-3-540-74997-4_6
- Isbister, K., & Doyle, P. (2002). Design and Evaluation of Embodied Conversational Agents: A Proposed Taxonomy. *The First International Joint Conference on Autonomous Agents & Multi-Agent Systems*.
- Kim, G. J. (2015). Human – Computer Interaction Fundamentals and Practice. *Human Computer Interaction: Fundamentals and Practice*, 1–12. Retrieved from <http://www.ittoday.info/Excerpts/HCI.pdf>
- KPMG. (2019). KPMG bouwt platform voor chatbotmanagement. Retrieved from <https://www.accountancyvanmorgen.nl/2019/02/04/kpmg-bouwt-platform-voor-chatbotmanagement/>
- Lowenthal, P. R. (2011). Social Presence. In *Social Computing* (pp. 129–136). IGI Global. <https://doi.org/10.4018/978-1-60566-984-7.ch011>
- McTear, M., Callejas, Z., & Griol, D. (2016). *The conversational interface: Talking to smart devices*. *The Conversational Interface: Talking to Smart Devices*. <https://doi.org/10.1007/978-3-319-32967-3>
- Mind Browser. (2017). Chatbot Survey 2017 a Research Study By Mindbrowser in Association With “Chatbots Journal” Current State of Chatbots and Their Outlook in 2017 Introduction To the the Survey. *Chatbots Journal*.
- Moore, R. J., Arar, R., Ren, G.-J., & Szymanski, M. H. (2017). Conversational ux design. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 492–497).
- Mujeeb, S., Hafeez, M., & Arshad, T. (2017). Aquabot: A Diagnostic Chatbot for Achluophobia and Autism. *International Journal of Advanced Computer Science and*

7. Bibliography

- Applications*, 8(9), 209–216. <https://doi.org/10.14569/ijacsa.2017.080930>
- Murgia, A., Janssens, D., Demeyer, S., & Vasilescu, B. (2016). Among the Machines: Human-Bot Interaction on Social Q&A Websites. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1272–1279). <https://doi.org/10.1145/2851581.2892311>
- Nass, C., & Moon, Y. (2000). Mindfulness Theory and Social Issues - Machines and Mindlessness - Social Responses to Computers. *Journal of Social Issues : A Journal of the Society for the Psychological Studies of Social Issues*, Jg. 2000(1), 81–104.
- Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Human Factors in Computing Systems*, 122–129. <https://doi.org/10.1109/VSMM.2014.7136659>
- Nimavat, K., & Champaneria, T. (2017). Chatbots: An Overview Types, Architecture, Tools and Future Possibilities. *International Journal of Scientific Research and Development*, 5(7), 1019–1024. Retrieved from http://ijsrd.com/Article.php?manuscript=IJSRDV5I70501%0Ahttp://ijsrd.com/C_Article.php?manuscript=NCICTP001
- Pearson, J., Hu, J., Branigan, H. P., Pickering, M. J., & Nass, C. I. (2006). Adaptive language behavior in HCI: how expectations and beliefs about a system affect users' word choice. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1177–1180. Retrieved from <http://dl.acm.org/citation.cfm?id=1124948>
- Pinar Saygin, A., Cicekli, I., & Akman, V. (2001). Turing Test: 50 Years Later, 463–518. Retrieved from <https://link.springer.com.ezproxy.avans.nl/content/pdf/10.1023%2FA%3A1011288000451.pdf>
- Price, R. (2016). Microsoft is deleting its AI chatbot's incredibly racist tweets. *Business Insider Deutschland*, 2016. Retrieved from <https://img.sauf.ca/pictures/2016-03-24/d360716e3199095063ebd4749b78fc4c.pdf>
- PwC Digital Services. (2017). Bot.Me: A revolutionary partnership. <https://doi.org/10.3791/2722>

7. Bibliography

- Qiu, L., & Benbasat, I. (2009). Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems. *Journal of Management Information Systems*, 25(4), 145–182. <https://doi.org/10.2753/mis0742-1222250405>
- Robinson, M., Gray, J., Cowley, A., & Tan, R. (2017). Adopting the power of conversational UX Chatbots. *Deloitte*, 1–24. Retrieved from <https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/financial-services/deloitte-nl-fsi-chatbots-adopting-the-power-of-conversational-ux.pdf>
- Rushton, A. M., & Carson, D. J. (1989). The marketing of services: managing the intangibles. *European Journal of Marketing*, 23(8), 23–44.
- Salesforce. (2018). The 2018 state of chatbots report. Retrieved from <https://www.salesforce.com/blog/2018/01/why-consumers-prefer-chatbots.html>
- Scheepers, R., Lacity, M. C., & Willcocks, L. P. (2018). Cognitive Automation as Part of Deakin University's Digital Strategy. *MIS Quarterly Executive*, 17(2), 89–107.
- Schroeder, J., & Schroeder, M. (2018). Trusting in Machines: How Mode of Interaction Affects Willingness to Share Personal Information with Machines. *Proceedings of the 51st Hawaii International Conference on System Sciences*, 9. <https://doi.org/10.24251/hicss.2018.061>
- Schuetzler, R. M., Grimes, G. M., Giboney, J. S., & Buckman, J. (2014). Facilitating Natural Conversational Agent Interactions: Lessons from a Deception Experiment. *Proceedings of the International Conference on Information Systems (ICIS)*, 1–16. <https://doi.org/doi:10.2514/6.2004-6725>
- Senthil Mahesh, P. C., Rodrigues, P., & Campbell, D. T. (2013). *Experimental and quasi-experimental designs for research*. *Journal of Chemical Information and Modeling* (Vol. 53). Ravenio Books. <https://doi.org/10.1017/CBO9781107415324.004>
- Seo, D., & Rietsema, A. (2010). A Way to become Enterprise 2.0: Beyond Web 2.0 Tools. In *Proc. Intl. Conf. on Information Systems (ICIS)* (p. Paper 140). Retrieved from http://aisel.aisnet.org/icis2010_submissions/140

7. Bibliography

- Short, J., Williams, E., & Christie, B. (1976). *The Social Psychology of Telecommunications*. John Wiley.
- Slack, M. K., & Draugalis, J. R. (2001). Establishing the internal and external validity of experimental studies. *American Journal of Health-System Pharmacy*, 2182–2184.
- Spears, R., & Lea, M. (1992). *Social influence and the influence of the “social” in computer-mediated communication*. *Contexts of computer-mediated communication*. Harvester Wheatsheaf.
- Srinivasan, K., Nguyen, C., & Tanguturi, P. (2018). Chatbots are here to stay. *Accenture Digital*. Retrieved from https://www.accenture.com/t20180509T102140Z__w__us-en/_acnmedia/PDF-77/Accenture-Research-Conversational-AI-Platforms.pdf
- Summers, J., & Perdue, C. (1986). Checking the Success of Manipulations in Marketing Experiments. *Journal of Marketing Research*, 23(4), 317–326.
- Tanis, M., & Postmes, T. (2003). Social cues and impression formation in CMC. *Journal of Communication*, 53(4), 676–693. <https://doi.org/10.1093/joc/53.4.676>
- Tourangeau, R., Couper, M. P., & Steiger, D. M. (2003). Humanizing self-administered surveys: experiments on social presence in web and IVR surveys. *Computers in Human Behavior*, 19(1), 1–24. [https://doi.org/10.1016/s0747-5632\(02\)00032-8](https://doi.org/10.1016/s0747-5632(02)00032-8)
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433–460.
- University of Twente. (2019). Alphabetic list of theories. Retrieved from <https://www.utwente.nl/en/bms/communication-theories/abc/>
- Verhagen, T., van Nes, J., Feldberg, F., & van Dolen, W. (2014). Virtual customer service agents: Using social presence and personalization to shape online service encounters. *Journal of Computer-Mediated Communication*, 19(3), 529–545. <https://doi.org/10.1111/jcc4.12066>
- Vincent, J. (2016). Twitter taught Microsoft’s AI chatbot to be a racist asshole in less than a day. *The Verge*.
- Vlek, C. (2014). Geslaagd voor de Turing test! Maar wat betekent dat eigenlijk? *De Volkskrant*.

7. Bibliography

Retrieved from <https://www.volkskrant.nl/nieuws-achtergrond/-geslaagd-voor-de-turing-test-maar-wat-betekent-dat-eigenlijk-~b6c9e624/>

Von Der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S. H. (2010). “It doesn’t matter what you are!” Explaining social effects of agents and avatars. *Computers in Human Behavior*, 26(6), 1641–1650. <https://doi.org/10.1016/j.chb.2010.06.012>

Walker, J. (1995). Service encounter satisfaction: constructualized. *Journal of Services Marketing*, 9(1), 5–14.

Wassertheil, S., & Cohen, J. (2006). Statistical Power Analysis for the Behavioral Sciences. *Biometrics*, 26(3), 588. <https://doi.org/10.2307/2529115>

Wolf, M. J., Miller, K., & Grodzinsky, F. S. (2017). Why we should have seen that coming: Comments on Microsoft’s Tay “Experiment,” and Wider Implications. *ACM SIGCAS Computers and Society*, 47(3), 54–64. <https://doi.org/10.1145/3144592.3144598>

7. *Bibliography*

8 APPENDIX

8.1 Questionnaire



Hello,

Thank you very much for participating in this research.

This survey is part of a master thesis research project. Answering this survey should not take more than 10 minutes, and all answers will be handled anonymous and strictly confidential. The data collected via this survey will only be used for the purpose of this study. This survey will be completely in English. This survey works best in a Google Chrome browser.

The survey will first ask a few general demographical questions. Hereafter you will see a short video and you are asked to fill in some questions about this video. The video contains no audio, so a speaker or headphones are not necessary to complete this survey.

Press the arrow to start this survey.





What is your current age?

What is your highest completed education?

- No education (geen opleiding afgerond)
- Primary School (Basisschool afgerond)
- High school graduate (Middelbare school afgerond)
- Secondary vocational education (MBO opleiding afgerond)
- Applied sciences (HBO opleiding afgerond)
- University Bachelor (WO Bachelor opleiding afgerond)
- University Master (WO Master opleiding afgerond)
- Other, please specify

What is your gender?

- Male
- Female
- Other

I would rate my proficiency in the English language as:

- Very good
- Somewhat good
- Neither good nor bad
- Somewhat bad
- Very bad



8. Appendix



In the next part you are going to watch a movie. This movie is an excerpt of a conversation held between a company's customer care department, and a customer. Although this conversation is pre-recorded, please watch it as if you were having the conversation yourself. Imagine that you are the person interacting in this conversation.

The messages in grey squares and shown to the left are sent to you. The messages in blue squares and shown to the right are the imaginative messages you send yourself. The message flow is from top to bottom of the screen. Messages will appear in the order as they are sent. Please refer to the picture below for a graphical explanation.

After showing this movie, a few questions will be asked about this conversation. You are allowed to watch the movie as many times as you want, and when you are ready to answer the questions, press the arrow to continue.

The movies do not contain audio, so a speaker or headphones are not necessary to complete this survey.



Please watch the movie carefully, and imagine as if you were the person having this conversation. Click the arrow to continue.



Before watching the movie, imagine that you have bought a product recently at an online store. You have the product for some weeks now, but realise that it doesn't meet your requirements. So, you'd like to return the product and get a refund.

You visit the website of the store, and while searching for the product return policy. You decide to have a chat conversation to find the answer you need. Imagine the movie on the next page to be this conversation.

Press the arrow to view the conversation.



8. Appendix



Please watch the video below. It is best to be watched on full-screen. You can watch the movie as many times as you like.



Please indicate whether you were able to watch the movie correctly.

- Yes, I could see the movie correctly.
- No, I could not see the movie correctly.



8. Appendix



Next, some questions will follow on the contents of the conversation.

Please note that there are no good or bad answers, and that your honest and instinctive first responses are most important for this research.



The term Conversational Partner is used to address the other party in the conversation. The messages sent by the conversational partner were shown in grey squares.

Please rate your conversational partner on the following scales.

| | Strongly disagree | Disagree | Somewhat disagree | Neither agree nor disagree | Somewhat agree | Agree | Strongly agree |
|--|-----------------------|-----------------------|-----------------------|----------------------------|-----------------------|-----------------------|-----------------------|
| I felt a sense of human contact with the conversational partner. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| I felt a sense of personhood with the conversational partner. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| I felt a sense of sociability with the conversational partner. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| I felt a sense of human warmth with the conversational partner. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| I felt a sense of human sensitivity with the conversational partner. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |



8. Appendix



The term Conversational Partner is used to address the other party in the conversation. The messages send by the conversational partner were shown in grey squares.

Please rate your conversational partner on the following scales.

I found my conversational partner...

| | | | | | | | |
|------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|----------------------|
| extremely inhuman-like | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | extremely human-like |
| extremely unskilled | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | extremely skilled |
| extremely unthoughtful | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | extremely thoughtful |
| extremely impolite | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | extremely polite |
| extremely unresponsive | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | extremely responsive |
| extremely unengaging | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | extremely engaging |



The term Conversational Partner is used to address the other party in the conversation. The messages send by the conversational partner were shown in grey squares.

Please rate your conversational partners on the following statements.

| | Extremely dissatisfied | Moderately dissatisfied | Slightly dissatisfied | Neither satisfied nor dissatisfied | Slightly satisfied | Moderately satisfied | Extrem satisf |
|---|---------------------------|----------------------------|--------------------------|---|-----------------------|-------------------------|-----------------------|
| How satisfied are you with the conversational agent's advice? | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| How satisfied are you with the way the conversational agent treated you? | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| How satisfied are you with the overall interaction with the conversational agent? | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |



8. Appendix



The term Conversational Partner is used to address the other party in the conversation. The messages send by the conversational partner were shown in grey squares.

Please rate how you experienced your conversational partner.

My conversational partner...

| | | |
|--------------------------------|---|----------------------------------|
| Was definitely a chatbot. | <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> | Was definitely a human-being. |
| Had a very clear introduction. | <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> | Had a very unclear introduction. |



We thank you for your time spent taking this survey.
Your response has been recorded.

8.2 Messages from the video

8.2.1 Group 1: Chatbot Introduction

Setting up your conversation with a customer care chatbot.
Please wait while we set up the connection...
Conversation is set up with 'Customer Chatbot'

Hello, you are talking with a customer care chatbot. How can I help you?

I would like some help with returning a product.

You can return a product as long as it is within 30 days of purchase. When did you purchase this product?

It was two weeks ago.

Then you can return your product. Would you like to get a refund, or would you like to exchange it for another?

I'd like to get my money back.

If you open your account on our website, you find the instructions for returning the package.

Okay. I will go there and follow the instructions.

Do you have any other questions?

No. Thank you for the help.

Have a great day!

8.2.2 Group 1: Undefined Introduction

Setting up your conversation with customer care.
Please wait while we set up the connection...
Conversation is set up.

Hello, how can I help you?

I would like some help with returning a product.

You can return a product as long as it is within 30 days of purchase. When did you purchase this product?

It was two weeks ago.

Then you can return your product. Would you like to get a refund, or would you like to exchange it for another?

I'd like to get my money back.

If you open your account on our website, you find the instructions for returning the package.

Okay. I will go there and follow the instructions.

Do you have any other questions?

No. Thank you for the help.

Have a great day!

8.2.3 Group 3: Control Group – Human Being

Setting up your conversation with a customer care employee.
Please wait until an expert is available to speak to you...
Conversation is set up with 'Peter'

Hello, you are talking with a customer care employee. How can I help you?

I would like some help with returning a product.

You can return a product as long as it is within 30 days of purchase. When did you purchase this product?

It was two weeks ago.

Then you can return your product. Would you like to get a refund, or would you like to exchange it for another?

I'd like to get my money back.

If you open your account on our website, you find the instructions for returning the package.

Okay. I will go there and follow the instructions.

Do you have any other questions?

No. Thank you for the help.

Have a great day!

8.3 Message Delay Calculation

| Message Content | Amount of Words | Delay (in seconds) |
|---|-----------------|--------------------|
| I would like some help with returning a product. | 9 | 5,15 |
| You can return a product as long as it is within 30 days of purchase. When did you purchase this product? | 21 | 9,35 |
| It was two weeks ago. | 5 | 3,75 |
| Then you can return your product. Would you like to get a refund, or would you like to exchange it for another? | 22 | 9,7 |
| I'd like to get my money back. | 7 | 4,45 |
| If you open your account on our website, you find the instructions for returning the package. | 16 | 7,6 |
| Okay. I will go there and follow the instructions. | 9 | 5,15 |
| Do you have any other questions? | 6 | 4,1 |
| No. Thank you for the help. | 6 | 4,1 |
| Have a great day! | 4 | 3,4 |

8.4 Extended Statistical Results

8.4.1 Overview table of the respondents' education level.

| | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|-----------|---------|---------------|--------------------|
| High school graduate (Middelbare school afgerond) | 19 | 11,9 | 11,9 | 11,9 |
| Secondary vocational education (MBO opleiding afgerond) | 3 | 1,9 | 1,9 | 13,8 |
| Applied sciences (HBO opleiding afgerond) | 14 | 8,8 | 8,8 | 22,6 |
| University Bachelor (WO Bachelor opleiding afgerond) | 36 | 22,6 | 22,6 | 45,3 |
| Other, please specify | 9 | 5,7 | 5,7 | 50,9 |
| University Master (WO Master opleiding afgerond) | 78 | 49,1 | 49,1 | 100,0 |
| Total | 159 | 100,0 | 100,0 | |

8.4.2 Sample age distribution overview

| Age | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-----------|---------|---------------|--------------------|
| 20,00 | 3 | 1,9 | 1,9 | 1,9 |
| 21,00 | 9 | 5,7 | 5,7 | 7,5 |
| 22,00 | 12 | 7,5 | 7,5 | 15,1 |
| 23,00 | 17 | 10,7 | 10,7 | 25,8 |
| 24,00 | 16 | 10,1 | 10,1 | 35,8 |
| 25,00 | 18 | 11,3 | 11,3 | 47,2 |
| 26,00 | 17 | 10,7 | 10,7 | 57,9 |
| 27,00 | 13 | 8,2 | 8,2 | 66,0 |
| 28,00 | 7 | 4,4 | 4,4 | 70,4 |
| 29,00 | 11 | 6,9 | 6,9 | 77,4 |
| 30,00 | 4 | 2,5 | 2,5 | 79,9 |
| 31,00 | 3 | 1,9 | 1,9 | 81,8 |
| 32,00 | 4 | 2,5 | 2,5 | 84,3 |
| 33,00 | 2 | 1,3 | 1,3 | 85,5 |
| 34,00 | 2 | 1,3 | 1,3 | 86,8 |
| 35,00 | 2 | 1,3 | 1,3 | 88,1 |
| 36,00 | 4 | 2,5 | 2,5 | 90,6 |
| 37,00 | 1 | ,6 | ,6 | 91,2 |
| 38,00 | 1 | ,6 | ,6 | 91,8 |
| 42,00 | 1 | ,6 | ,6 | 92,5 |
| 43,00 | 1 | ,6 | ,6 | 93,1 |
| 44,00 | 2 | 1,3 | 1,3 | 94,3 |
| 45,00 | 1 | ,6 | ,6 | 95,0 |
| 47,00 | 1 | ,6 | ,6 | 95,6 |
| 52,00 | 1 | ,6 | ,6 | 96,2 |
| 53,00 | 2 | 1,3 | 1,3 | 97,5 |
| 54,00 | 1 | ,6 | ,6 | 98,1 |
| 56,00 | 1 | ,6 | ,6 | 98,7 |
| 58,00 | 1 | ,6 | ,6 | 99,4 |
| 70,00 | 1 | ,6 | ,6 | 100,0 |
| Total | 159 | 100,0 | 100,0 | |

8.4.3 SPSS Syntax

```
* Encoding: UTF-8.
DATASET ACTIVATE DataSet1.

FREQUENCIES VARIABLES=Q10
/ORDER=ANALYSIS.

ONEWAY Q24_1 BY Group
/MISSING ANALYSIS.

ONEWAY Q24_2 BY Group
/MISSING ANALYSIS.

MEANS TABLES=Q24_1 BY Group
/CELLS=MEAN COUNT STDDEV.

RELIABILITY
/VARIABLES=Q17_1 Q17_2 Q17_3 Q17_4 Q17_5
/SCALE('Social_Presence') ALL
/MODEL=ALPHA.

RELIABILITY
/VARIABLES=Q18_1 Q18_2 Q18_3 Q18_4 Q18_5 Q18_6
/SCALE('Perceived_Humanness') ALL
/MODEL=ALPHA.

RELIABILITY
/VARIABLES=Q19_1 Q19_2 Q19_3
/SCALE('Service_Encounter_Satisfaction') ALL
/MODEL=ALPHA.

COMPUTE SocPrTot=((Q17_1 + Q17_2 + Q17_3 + Q17_4 + Q17_5) / 5)-20.
EXECUTE.

COMPUTE PerHumTot=(Q18_1 + Q18_2 + Q18_3 + Q18_4 + Q18_5 + Q18_6) / 6.
EXECUTE.

COMPUTE EncSatTot=((Q19_3 + Q19_2 + Q19_1) / 3) - 20).
EXECUTE.

ONEWAY SocPrTot BY Group
/MISSING ANALYSIS.

ONEWAY PerHumTot BY Group
/MISSING ANALYSIS.

ONEWAY EncSatTot BY Group
/MISSING ANALYSIS.
```

8. *Appendix*

```
MEANS TABLES=SocPrTot BY Group  
/CELLS=MEAN COUNT STDDEV.
```

```
MEANS TABLES=PerHumTot BY Group  
/CELLS=MEAN COUNT STDDEV.
```

```
MEANS TABLES=EncSatTot BY Group  
/CELLS=MEAN COUNT STDDEV.
```

