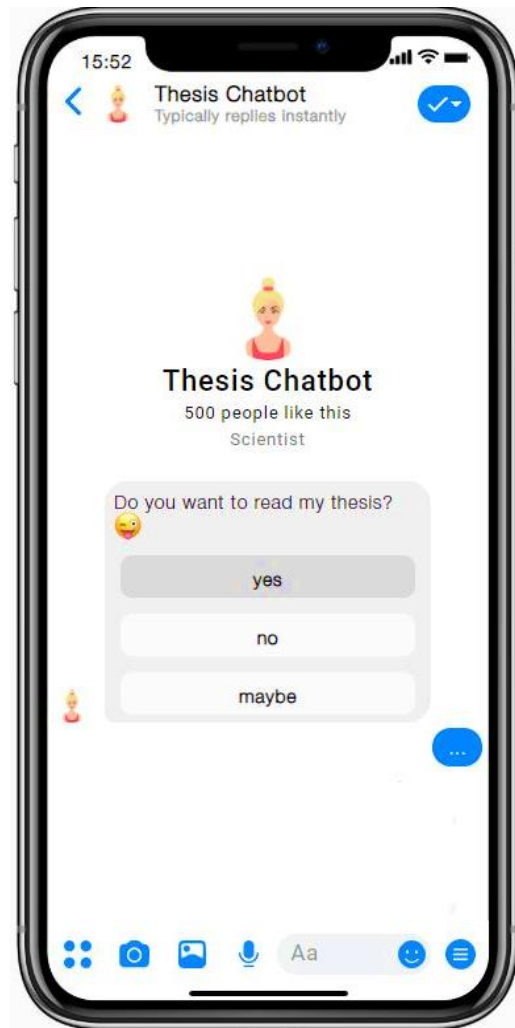


Exploring Users' Perception of Chatbots in a Mobile Commerce Environment

Creating a Better User Experience by Implementing Anthropomorphic Visual and Linguistic Chatbot Features



Lena Marie Assink

Bachelor Thesis in Communication Science (BSc)
Faculty of Behavioural, Management and Social Sciences (BMS)

UNIVERSITY OF TWENTE.

Exploring Users' Perception of Chatbots in a Mobile Commerce Environment

Creating a Better User Experience by Implementing Anthropomorphic Visual and Linguistic Chatbot Features

UNIVERSITY OF TWENTE.

Author: Lena M. Assink
University of Twente
7522 NB Enschede
The Netherlands

Name: Lena Marie Assink
Student number: 1807722
Email: l.m.assink@student.utwente.nl
Study: Communication Science
Thesis for the degree: Bachelor of Science
Supervisor: Dr. A. D. Beldad
Date: 28.06.2019

Total number of words: 13412

Abstract

In our modernized world, artificial intelligence (AI) is growing rapidly and organizations implement smart technologies continuously. AI and chatbots are edging their way into numerous industries and are changing the way customers communicate with organizations. Chatbots have huge potential as customer service, however, there have been few empirical investigations into the impact chatbots have on their users. Thus, this research investigates the implementation of anthropomorphic visual and linguistic features in chatbot applications by using an experimental 2x3 research design with m-commerce videos. The videos display either a human, an animated person or a logo with either human or robotic language. Using these methods, this study explores the extent to which chatbot appearance and language can potentially influence the perception of trust, satisfaction, and purchase intention. Due to researched literature, it was expected that respondents perceive a higher trust, satisfaction, and purchase intention when they are confronted with a chatbot that displays anthropomorphic visual and linguistic features. The data was collected through an online survey among 265 respondents. The respondents were chosen due to selection criteria and were 18-62 years old and mostly from Germany and the Netherlands. The corresponding survey was distributed through online channels, SONA, and on the campus of the University of Twente. Results of this research show that the implementation of anthropomorphism features lead to a higher satisfaction among users. Nonetheless, the results for trust and purchase intentions were insignificant. However, this study had limitations due to the oversimplification of the experimental design. More research on this topic needs to be carried out before the association between chatbot appearance and language is more clearly understood. Nonetheless, the findings of this research suggest that organizations, marketers and chatbot designers should strive for the implementation of anthropomorphic visual and linguistic cues within the development of chatbots. By doing this, organizations create a better user experience for customers who interact with intelligent agents.

Keywords: Chatbots, Human–computer interaction, m-commerce, anthropomorphism cues, conversational UI, user experience

Table of content

1. Introduction.....	6
2. Theoretical Framework.....	9
2.1. Research on Chatbots	9
2.1.1. Chatbot Appearance.....	10
2.1.2. Chatbot Language	11
2.3. Mediating Role of Trust.....	13
2.4. Satisfaction.....	14
2.5. Purchase Intention.....	15
2.6. Research Model.....	16
3. Methods.....	17
3.1. Methodology and Experiment Design.....	17
3.2. Materials.....	17
3.2.1. Design Materials.....	18
3.3. Pre-test.....	19
3.4. Final Stimuli.....	20
3.5. Manipulation Check.....	21
3.5.1. Manipulation for Chatbot Appearance.....	21
3.5.2. Manipulation for Chatbot Language.....	22
3.6. Respondents.....	22
3.7. Procedure.....	23
3.8. Measurement Instruments.....	23
3.9. Construct Validity and Reliability	25
3.9.1. Validity.....	25
3.9.2. Reliability	27
4. Results.....	28
4.1. Multivariate Analysis of Variance	28
4.2. Main Effects of Chatbot Appearance and Language.....	29
4.2.1. Effects on Satisfaction.....	29
4.2.2. Effects on Purchase Intention.....	30
4.3. Interaction Effect of Chatbot Appearance and Language	30
4.4. Overview of the Hypotheses	32

5. Discussion.....	34
5.1. Discussion of Results.....	34
5.1.1. Discussion of Main Effects.....	34
5.1.2. Discussion of Moderating Effect.....	36
5.1.3. Discussion of Mediation Effect	37
5.2. Implications.....	37
5.2.1. Practial Implications.....	37
5.2.2 Theoretical Implications	38
5.3. Limitations and Recommendations for Future Research	39
5.4. Conclusion	40
5.5. Acknowledgements	41
6. Literature.....	42
7. Appendix	48
Appendix A - Final Questionnaire	48
Appendix B - Overview of Chatbot Videos	51
Appendix C - Literature Log	52

1. Introduction

Nowadays, technology shifts and changes our world constantly. Especially, Artificial Intelligence (AI) is seen as a critical element in the digital transformation and has the power to reshape and transform businesses and organizations entirely (Ransbotham, Gerbert, Reeves, Kiron, & Spira, 2018). AI was long seen as a future imagination or theoretical construct (Buchanan, 2005). Yet, AI development is highly speeding up and is transforming the nature of almost everything that is connected to human life. In fact, robots or autonomous systems are progressively born and have already started to replace human labour increasingly (Tyagi, 2016). This future scenario of intelligent machines and bots that work and react like humans existed formerly only as a theoretical possibility. Nowadays, the development of chatbots is more than present in many industries and is seen as a popular trend (Razaque & Yang, 2018).

The evolution of AI has led to various innovators wanting to incorporate the emerging technology opportunities into their respective fields. One of these fields where the system has been used is in the development of chatbot applications. Chatbots are described as an impression of interacting with humans online, while actually just querying a database, put to life by natural language input (Radziwill & Benton, 2017). Wong (2016) defines chatbots as an application of an artificial intelligence computer program which imitates conversations with users. The fields of applications of chatbots are manifold. However, one of the most popular fields are chatbots for purchasing tickets and searching or buying products online (Zumstein & Hundertmark, 2017). Many companies have already recognized this trend and followed the movement of implementing chatbots into their services. To be precise, in 2017, more than 34.000 chatbots were already available and active in the Facebook messenger app (Statista, 2017). Big brands in the Netherlands started to offer chatbots as a customer service, such as NOS, KLM or Eneco so as to be available 24/7 to their customers (Schurer, 2017). Likewise, large German brands implemented chatbot customer services similarly, such as Lufthansa or Klarmobil (Mehner, 2018). However, the chatbot trend is not only limited to the German or Dutch market, since it is expected to grow on a global scale (Suthar, 2019).

One of the biggest advantages of chatbots for businesses is their availability, since they can be used 24/7 (Hald, 2018). The use of chatbots as a customer service on apps like Facebook is highly profitable and attractive to companies. To be precise, the potential global annual revenue generated by chatbot transactions is estimated up to 32 billion US Dollars (Business Insider 2017). In addition, not only profits are desirable, but also enables companies to establish and maintain a more direct relationship with their customers (van Bruggen, Antia, Jap, Reinartz, & Pallas, 2010). Most businesses can interact with their customers through one-to-one communication on a personal device due to chatbots. The AI agents enable companies to create new and direct customer contact points and offer automation of communication. Moreover, the automated interaction with customers is not only used to reduce costs, but also to increase customer satisfaction (Radziwill & Benton, 2017). Additionally, customers have the opportunity to communicate 24 hours a day, 7 days a week, independent of working or opening hours

of the business (Zumstein & Hundertmark, 2017). Thus, companies can save on personnel costs but can still offer customer services. Using chatbots give consumers the opportunity to get customer support, get personalized recommendations, and click to purchase within messaging apps (Shopify, 2016). To conclude, outstanding benefits for companies are the opportunity to offer 24/7 customer services, reduce costs, create direct customer contact, and to increase customer satisfaction and purchase intention.

As the aforementioned section describes, it is highly desirable for companies to obtain customer satisfaction and purchase intention. Chatbots can have huge advantages to companies, however, they must be implemented in the right manner. Zumstein and Hundertmark (2017) state that the users often experience mistrust with intelligent technologies, such as chatbots. Moreover, from other technology areas, it is well researched that trust is a critical factor in the user's uptake of interactive systems (Corritore, Kracher & Wiedenbeck, 2003). In addition, other research shows that users' trust in chatbots for customer service was found to be affected by factors concerning the specific chatbot appearance, specifically the quality of their human-likeness (Følstad, Nordheim & Bjørkli, 2018). Another key difficulty to the adoption and use of chatbots is that the interaction with them often does not feel natural or human-like (Schuetzler, Grimes, Giboney, & Buckman, 2014). Thus, the main desire of a user is to experience a natural conversation with a chatbot, that feels human-like (Garcia, 2018).

Hence, it is desirable to create a chatbot as human as possible. Social or human-like cues, such as style of language can influence and increase the perception of anthropomorphism (Araujo, 2018). Similarly, Higashinaka, Minami, Dohsaka and Meguro (2010) state that the dialogue quality of a virtual assistant can lead to an improvement in customer satisfaction. Toma (2010) argues that not only the style of language, referred to as textual information, elicits trustworthiness online, but also visual cues play a crucial role. Visual cues, such as chatbot appearance, are design features for making chatbot interactions appear more natural and human-like (Appel, Pütten, Krämer, & Grach, 2012). Amdocs (2017) suggests that consumers even prefer the female gender in chatbot appearances, although most brands and companies do not use human pictures or animations at all, and instead create logos for their online services.

In fact, most websites of e-services frequently lack in human appearances, which may hinder the purchase intention of potential customers as well as the development of trust, since online interactions with social presence is believed to be crucial in the creation of customer trust (Gefen & Straub, 2003). The theory of social presence argues that through the interaction with human-like cues, an anthropomorphism feeling can be created without the actual human contact (Gefen & Straub, 2004). Previous research has shown that social cues, such as a human-like appearance through pictures or animation, can create perceptions of social presence (Qui & Benbasat, 2009). Moreover, using human pictures or animations increases the perception of social presence, which positively influences satisfaction and purchase intention (Hassaein & Head, 2007).

Hence, in this research the effects of chatbot appearance, namely a human, an animated person and an organizational logo of Eventim within a chatbot setting to find a fitting concert ticket will be

examined. Eventim is an events and ticket agent and is Europe's largest ticket retailer (Miller, 2010). Since the key factor which determines the adoption and use of chatbots is the perceived anthropomorphism, two crucial factors will be examined. Next to the chatbot appearance, the perceived humanness of the chatbots language will be explored. Hence, this research additionally examines the chatbots language, which can either be robotic or humanlike.

The current development of mobile messenger chatbots is still in its infancy and due to its novelty, there is currently little to no research on the anthropomorphic visual and linguistic chatbot features in the mobile messenger interface environment. Many businesses only use their company's logo in the chatbot design in order to engage with their customers. However, theories such as the social presence theory suggest that the simple presence of a human picture can have a better impact on the user's engagement with the chatbot. Thus, implementing anthropomorphic cues in terms of chatbot appearance and language can have a better impact on the perception of the user in comparison to only a logo. Hence, it is therefore valuable to know whether users trust different type of chatbots, depending on the degree of anthropomorphic features of appearance or language. Moreover, implementing the right chatbot design in businesses can not only lead to user's trust, but also increase the satisfaction and purchase intention which is crucial to a business's success. Consequently, the aim of this research is formulated in the following research questions:

1: To what extent does chatbot appearance influence trust, satisfaction and purchase intention?

2: To what extent does robotic/human language influence trust, satisfaction and purchase intention?

3: To what extent are the effects of a chatbot appearance on trust, customer satisfaction, and purchase intention dependent on the robotic/human language used for the interaction?

4: To what extent are the effects of chatbot appearance and robotic/human language on trust, satisfaction and purchase intention mediated by trust?

This research is divided into multiple sections. Firstly, chapter two depicts a theoretical framework with the dependent (trust, satisfaction, purchase intention) and independent (chatbot appearance, language) variables of this research. Hypotheses are concluded out of the framework and is followed by the research model of this study. Secondly, the research methods and designs are elaborated in chapter three. Further, the results of this research are presented in chapter four, followed by chapter five in which a discussion of the results is depicted. Lastly, limitations of this research and a conclusion is given.

2. Theoretical Framework

2.1. Research on Chatbots

Artificial Intelligence has gained popularity by many scholars and therefore, numerous definitions of chatbots have emerged recently. Wong (2016) simply defined chatbots as a computer program that imitates conversations with users, applying artificial intelligence. Other scholars described chatbots as the impression of human communication online by using natural language (Radziwill & Benton, 2017). Not only the importance of using natural language should be highlighted, but also the ability of chatbots to interaction over text or voice in real-time should be highlighted (Razaque & Yang, 2018). However, definitions of chatbot applications have been around for a longer period of time.

In fact, Eliza was one of the first chatbots in the 1960s, which is one of the earliest Natural Language Applications (NLP) by using simple pattern matching and a template-based response mechanism in order to match the conversational style of a psychotherapist (Weizenbaum, 1983). The ability to conduct a conversation via textual methods by machines are designed to convincingly simulate human behaviour and responses as a conversational partner. Thus, the interests and opportunities in creating human-like chatbots increased as well as the futuristic prospect of a chatbot being indistinguishable from a human that could pass the Turing Test one day (Turing, 1950).

Creating human-like chatbots or machines that can fool actual humans into thinking that they interact with another human are the main characteristics of the Turing Test and opened up new areas of competition in artificial intelligence. The Loebner Prize Competition is an annual competition for conversational agents, such as chatbots, where they are being tested via the Turing Test method (Bradeško & Mladenović, 2012). The most desirable aspects to implement are anthropomorphic features of the chatbot, to make them look and feel more human-like (Araujo, 2018). In addition, various developers and scientists have improved artificial agents over the last years by studying and modulating specific aspects which are essential in human interactions, such as the physical appearance of a chatbot (Giard & Guitton, 2010).

The development of chatbots is growing rapidly and organizations implement the smart agents continuously since they offer many benefits to companies. By creating and implementing the human-like chatbots into customer service, many companies see a huge potential in the conversational agents. Chatbots in customer service are available 24/7 (Hald, 2018), can lead to a direct purchase transaction (Business Insider 2017), establish a direct relationship with customers (van Bruggen, Antia, Jap, Reinartz, & Pallas, 2010), reduce costs, increase customer satisfaction (Radziwill & Benton, 2017) and trust (Følstad, Nordheim & Bjørkli, 2018).

2.1.1 Chatbot Appearance

In order to translate the benefits of chatbots into business practice, it is crucial for companies to choose the right design features of a chatbot. Focusing on the importance of anthropomorphic features in visual cues of chatbots, Appel, Pütten, Krämer and Gratch (2012) highlight the importance of choosing the right design of an AI agent since the appearance influences the user's interaction and perception of it.

The theory of social presence stresses that through the implementation of human-like cues, such as visual chatbot features, an anthropomorphism feeling can be created without the actual human contact (Gefen & Straub, 2004). Research has shown that social cues, such as the human-like appearance represented through a pictures or animation, can create a feeling of social presence (Qui & Benbasat, 2009). By using human pictures or animations, not only the perception of social presence increases but it also positively influences the user satisfaction and purchase intention (Hassaein & Head, 2007). In addition, online human interaction with social presence is believed to be crucial in the creation of customer trust and increases the overall customer experience (Gefen and Straub, 2003). Additionally, the majority of users prefer a female human appearance within the chatbot design, although most brands and companies use their own logos for their online services (Amdocs, 2017). Likewise, this preference of chatbot appearance is supported by Gustavsson (2005) who stresses that users prefer a female human-like appearance in the online environment.

Chatbot designers implement human-looking pictures as an approach to compensate for the lack of social presence in the online environment. An international study explored this aspect a bit further and revealed that 46 percent of clients stated that they would like a chatbot with a human appearance, while only 20 percent would want to see them as an animated picture (Singh, 2017). In addition, a newer and more recent study by CapGemini found out that consumers want chatbots to feel human, however, they do not necessarily want them to look human. In fact, one in two consumers say they are not comfortable with human physical features in chatbots; however, 64 percent of consumers want AI and chatbots to feel more human-like (Garcia, 2018). The current development of mobile messenger chatbots is still in its infancy and due to its novelty, there is only little to no research on the correct design choices for anthropomorphic visual chatbot features. Nonetheless, regardless of which chatbot appearance consumers prefer, different chatbot avatars images lead to different results (Tinwell, 2009).

Only a few empirical studies have been executed about the human appearance of a chatbot and their anthropomorphic effect. For this reason, this research explores the aspects of anthropomorphic visual features closer and uses a human picture, a human animated person and a logo in a m-commerce setting. Based on the findings from previous sections, the following hypothesis has been formulated:

H1: The perception of (a) trust, (b) satisfaction, and (c) purchase intention is higher when people are confronted with a chatbot using a human picture compared to people using a chatbot with an animated picture or organizational logo.

2.1.2. Chatbot Language

Over the last decade, new human-computer interfaces have emerged, which combine numerous human language technologies that enable humans to interact and communicate with computers using spoken or written dialogue for information access, creation, and processing (Zue & Glass, 2000). These platforms that mimic a conversation with a real human are called conversational interfaces (CUI). CUI's give users the opportunity to communicate with a computer (or chatbot) in their natural language or in other words human language, instead of in a syntax specific command (Brownlee, 2016). Historically, this was the only way for a user to interact with computers since they relied on graphical user interfaces (GUI) and the user's interaction by pressing syntax specific commands (e.g. "close" or "next") which were then translated into actions that the computers could understand (Myers, 1998). As soon as the usage of personal machines grew, likewise did the desire grow to communicate with machines in the same way as with other humans by using natural language (Atwell & Shawar, 2007). Human interaction with computers (or chatbots) via natural language is a topic that is widely researched for many years (Zadrozny, Budzikowska, Chai, Kambhatla, Levesque & Nicolov, 2000), however, it is still highly complex to grasp. Especially chatbot conversations with human interaction lack in empirical research.

As defined, chatbots are conversational software agents activated by natural language input, which can be in the form of text or voice (Razaque & Yang, 2018). They provided conversational output in responses, which can either feel natural or in other words, human-like. Although chatbots have improved enormously over the last years, they are still clearly distinguishable from human conversations. In fact, Hill, Ford, and Farreras (2015) found out key differences between a human and a chatbot conversation. Differences were found between words per message, words per conversation, word uniqueness, and use of profanity, shorthand, close questions and emoticons. To be precise, people interacted with chatbots longer but with shorter messages than they would with another human. In addition, chatbot conversations lacked in richness of vocabulary and did not exhibit profanity. As a result, there is overall a notable difference in the content and quality of chatbot-human and human-human conversations.

Moreover, the effect of natural language that results from a chatbot conversation has effects on human perception. The main desire of humans is to have a natural experience with chatbots, that feel human-like (Garcia, 2018). This can be provided by a high level of natural, human-like language. If a chatbot conversation provides responses with good and efficient natural language, it results in higher user satisfaction (Deshpande, Shahane, Gadre, Deshpande, & Joshi, 2017). Similarly, Higashinaka, Minami, Dohsaka, and Meguro (2010) state that the dialogue quality of a virtual human assistant can lead to an improvement in customer satisfaction. Moreover, conversations, such as those with conversational agent interface, are closely linked to trustworthiness (Cassell, Bickmore, Billinghurst, Campbell, Chang, Vilhjálmsón, & Yan, 1999). For instance, human or natural language conversations have higher ratings of trust in comparison to machine-like speech (Muralidharan, 2014). Further,

Johnson, Patron, and Lane (2007) found out that interactions feel less natural and trustworthy when the structure of the language lacks familiarity, which mostly results from a chatbot conversation, rather than a human-human one. Another aspect of textual conversations within the online environment, and especially when customers are in contact with customer service (regardless of human or chatbot), is the development of the intent to purchase (Gupta, Varshney, Ijhamtani, Kedia & Karwa, 2014). Concludingly, having a good, anthropomorphic conversation can result in satisfaction, purchase intention and trust.

Moreover, scholars define chatbots continuously in combination with natural language and highlight the importance of their synergy (Razaque & Yang, 2018). Similarly, the chatbot language in terms of tone of voice was found to have a moderating effect on consumer responses in social media, such as the Facebook messenger (Keyzer, Dens, & Pelsmacker, 2017). Another study outside of the online environment highlighted the importance of the interaction between the tone of voice and the human face (Zuckerman, Amidon, Bishop, & Pomerantz, 1982). In addition, Dessalegn and Landau (2013) explored that language has a moderating effect on our most important non-linguistic system – vision. Although there is no empirical evidence between visual and linguistic anthropomorphism features, the closely related fields might be hints for an interaction effect between chatbot appearance and language.

Thus, natural language is not only a dependent variable within this research. It is assumed that chatbot language affects the direction or strength of the relationship between chatbot appearance on the independent variables trust, satisfaction, and purchase intention. Within this research, high and low natural language is displayed, based on the key characteristics of Hill, Ford, and Farreras (2015) to clearly distinguish the human and chatbot conversations. These two conditions are named ‘robotic’ for low natural language and ‘human’ for high natural language performance. Based on the findings from the previous sections, the following hypotheses have been formulated:

H2: The perception of (a) trust, (b) satisfaction, and (c) purchase intention is higher when people are confronted with a chatbot using human language compared to people using a chatbot with a robotic language.

H3: The perception of (a) trust, (b) satisfaction, and (c) purchase intention is higher when people are confronted with a chatbot using a human picture and human language compared to people using a chatbot using a logo or animated picture with human-like language.

H4: The perception of (a) trust, (b) satisfaction, and (c) purchase intention is higher when people are confronted with a chatbot using robotic language with either an animated picture or logo compared to people using a chatbot with robotic language and human-like language.

H7: Language moderates the impact of chatbot appearance on (a) trust, (b) satisfaction, and (c) purchase intention.

2.3. The Mediating Role of Trust

Various research has shown that trust is a crucial element in the online environment. In fact, trust can explain the relationship and link between chatbot appearance, natural language, and satisfaction as well as purchase intention. However, it is crucial to comprehend the definition of trust as well as the application to the online environment. A variety of definitions of the term trust have been suggested, such as the concept of trust which can be seen as (Grazioli & Jarvenpaa, 2000) ‘a state of perceived vulnerability or risk that is derived from individual’s uncertainty regarding the motives, intentions, and prospective actions of others on whom they depend’ (p. 571). Moreover, Mayer, Davis, and Schoorman (1995) describe trust as ‘the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party’ (p. 712). Concluding out of the definitions, there is a relationship suggested between trustor and trustee. In the online context, the provider of online services is seen as the trustee, with the user assuming to be the role of the trustor.

Moreover, trust has a strong relation to the independent variables of this research, namely, chatbot appearance and natural language. Emerging chatbot application areas, such as online customer service is just in the early stages of development. However, from other technology areas, it is well researched and proven that trust is a critical factor in user’s uptake of interactive systems (Corritore, Kracher & Wiedenbeck, 2003). Further, users' trust in chatbots for customer service was found to be affected by factors concerning the specific chatbot appearance, specifically the quality of its human-likeness, which both can be embodied by different types of chatbot appearance and natural language (Følstad, Nordheim & Bjørkli, 2018). Toma (2010) takes on the same approach and argues that visual (chatbot appearance) and textual information (conversational interfaces and natural language) online elicits trustworthiness. In addition to the importance of the right chatbot appearance, Gefen and Straub (2003) argue that social presence, such as from a human chatbot, is believed to be crucial in the creation of customer trust. Further, Muralidharan (2014) points out the development of trust is depended on the interaction, for instance with a chatbot, with natural language instead of machine-like language. Thus, the correct choice and combination of chatbot appearance as well as the conversational interface with natural language build the groundwork for the perceived trust of a user in the online environment.

Focusing on role of trust, it is closely linked to the outcome of purchase intention as well as satisfaction. For instance, there are numerous research initiatives that have elaborated on the factors affecting online purchase intention such as trustworthiness (Adam, Aderet & Sadeh, 2008), perceived risk and consumer trust (Kim, Ferrin, & Rao, 2008). In addition, Hsin Chang and Wen Chen (2008) found out that the quality of a website, which for instance can rely on the right choice of chatbot appearance and conversational interface, affect consumers’ trust and in return consumers purchase intention. Likewise, customer satisfaction and trust are closely linked to each other since research proposes that trust precedes satisfaction, which means that first customers have to trust the

organization's services which then results in satisfaction (Gul, 2014). Moreover, a crucial determinant of satisfaction is the factor of customer trust (Bejou, Ennew, & Palmer, 1998).

While analysing literature about the variables of this thesis, namely, chatbot appearance and natural language as independent variables, and purchase intention as well as satisfaction as the dependent variables, studies have indicated that trust plays a vital role in relation to the dependent and independent variables. Hence, trust is seen as the mediating variable of this research. Based on the findings from the previous section, the following hypotheses have been formulated:

H5: The effects of chatbot appearance on (b) satisfaction, and (c) purchase intention are mediated by trust.

H6: The effects of natural language on (b) satisfaction, and (c) purchase intention are mediated by trust.

2.4. Satisfaction

One of the most crucial goals of any organization is to create a high level of customer satisfaction. Customer satisfaction is described as a vital component for each organization since it results in competitive advantages. Explicitly, customer satisfaction results in customers being less sensitive to price changes, higher profit, return on investment, positive word-of-mouth and customer loyalty (Thusyanthy & Tharanikaran, 2017). Further, Rust and Oliver (1994) describe customer satisfaction as the extent to which a person believes that a certain experience created positive feelings. Thus, companies that provide a high level of customer satisfaction will profit from it in the future in terms of customer loyalty (Anderson & Sullivan, 1993). However, in order to obtain the benefits that result from customer satisfaction, companies have to provide certain qualities to their customers. Gul (2014) proposes that trust precedes satisfaction, which means that first customers have to trust the organization's services which then result in satisfaction. This moderating role of trust is similarly explored by Madjid (2013) who explored customer trust as relationship mediation for customer satisfaction.

However, trust is not the only component that has to be achieved by a company. Creating user satisfaction is seen as an indicator for the success of technological applications, such as chatbots (Mahmood, Burn, Gemoets, & Jacquez, 2000). Hence, creating fitting visual and textual information for the chatbot interface is key for a better user experience and satisfaction among users (Toma, 2010). This can be provided by means of a fitting chatbot appearance design and conversational interfaces in order to increase satisfaction (Higashinaka, Minami, Dohsaka, & Meguro, 2010). In addition, central to creating satisfaction are anthropomorphic features of the chatbot design (Tinwell, 2009). Thus, factors such as chatbot appearance, the type of chatbot language, as well as perceived trust determine the degree of satisfaction for this research. Consequently, satisfaction is used as a dependent variable for this experiment.

2.5. Purchase Intention

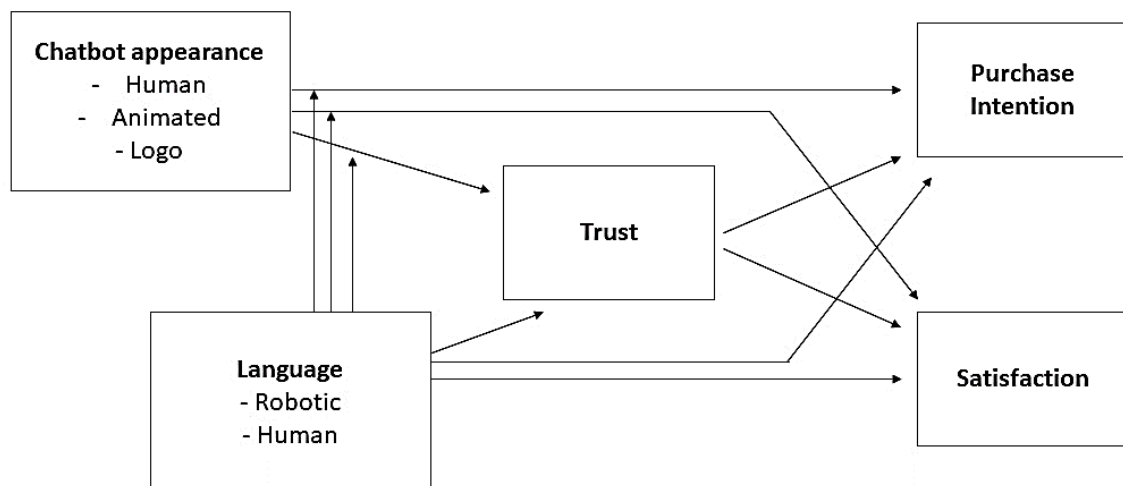
Grewal, Monroe, and Krishnan (1998) defined purchase intention as the probability of the customers' likelihood to purchase a particular product. Likewise, Mirabi, Akbariyeh, and Tahmasebifard (2015) describe purchase intention as a state in which consumers tend to buy a certain product. In addition, Morwitz and Schmittlein (1992) argue that purchase intention is an important factor to increase sales as well as purchase intention being a useful tool for sales forecasts. From an organizational point of view, it is motivating to predict purchases since it can be beneficial to marketing decisions in order to identify the demand of a product and to create fitting promotional strategies (Tsotsou, 2006). According to the Theory of Planned Behaviour (TPB) an individual's performance of a certain behaviour, such as purchasing a product, is determined by his or her intent to perform that behaviour (Georg, 2014). This intent is strongly linked to trust since it is the most direct influential factor in predicting online purchase intention (Sam, Fazli, & Tahir, 2009). Further, numerous research initiatives have elaborated on the factors affecting online purchase intention, which is, inter alia, trustworthiness (Adam, Aderet & Sadeh, 2008). Moreover, Toma (2010) argues that online visuals, such as the chatbot appearance, as well as textual information, which relates to natural language, are factors that elicit trustworthiness, which then in return leads to a higher customer purchase intention. The effect of human chatbots appearance on purchase intention is similarly explained by Reeves and Nass (1996) since they found out that static human images, photographs, and speech can help to attract users and persuade them to buy goods. Likewise, human-like visual cues such as animations of people or avatars have a similar effect (Hassanein & Head, 2007). Not only the chatbot appearance has a direct effect on purchase intention, similarly, natural language is directly linked to it. As an example, Gupta, Varshney, Jhamtani, Kedia, and Karwa (2014) identified different linguistic features within the online environment and the direct impact on the customers intend to purchase.

2.6. Research Model

Based on the reviewed literature and previous studies, a research model was designed, which is depicted in figure 1. It aims to explore the effects of chatbot appearance and language on trust, purchase intention and satisfaction.

Figure 1.

3x2 Research Model



3. Methods

3.1. Methodology and Experiment Design

In order to investigate the effect of chatbot appearance as well as language, this research carried out a 2x3 design. The three different conditions range from chatbot appearance, namely human, animated and organizational logo. Those are combined with the two factors of robotic or natural language. This three by two experimental design is depicted in table 1.

Table 1.

2x3 Experimental design with 6 conditions

		Chatbot appearance		
		Human	Animated	Logo
Language	Natural Language	Conversation 1	Conversation 2	Conversation 3
	Robotic Language	Conversation 4	Conversation 5	Conversation 6

3.2 Materials

The stimuli were six different Facebook chatbot interface conversations. Six different conversational interfaces were created for this experiment. The conversational interface depicted a Facebook messenger interface with three different chatbot appearances and two different language types. The first chatbot type was created with a (1) human appearance, the second was (2) an animated human picture and the third simply used the Eventim (3) logo. Each of the three chatbots showed a conversational interface in either natural (4) or robotic (5) language. These six different conditions were displayed as a video which was around 30 seconds long. One of the six conditions were randomly assigned to the participants. Before the interaction with the chatbot, the participant had to read through a scenario. This scenario described the option to use the Eventim chatbot on Facebook in order to find a fitting gift for one's friend. The participant had to watch one of the six videos in which the conversation with one of the chatbots was displayed. Since the participants were not able to directly interact with the chatbots themselves, it was asked to imagine to be the person who interacted with the chatbot in the given video.

3.2.1 Design Materials

Three different pictures were shown within the conversation of the chatbot. As Amdocs (2017) found out, users prefer the female gender in chatbot appearances, although most brands animations or logos online. Hence, the different chatbot appearances were depicted as a (1) human, the second picture was (2) an animated human and the third picture simply used the Eventim (3) logo. The used chatbot pictures are illustrated in figure 2.

Figure 2.

Images for the chatbot appearance



1) Human picture



2) Animated human

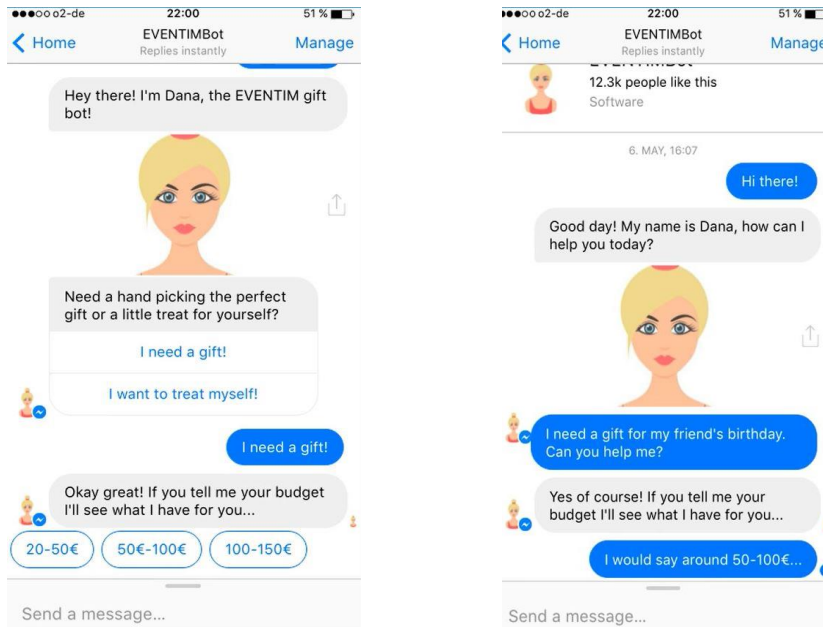


3) Eventim logo

Each of the three chatbots showed a conversational interface in either natural (4) or robotic (5) language. Since Hill, Ford and Farreras (2015) found out the differences between a human and a chatbot conversation lay between words per message, words per conversation, richness in vocabulary, close questions and word uniqueness. In addition, people interacted with chatbots longer but with shorter messages than they would with another human. These elements were implemented in the two different conversational types. An example of these two conditions are given in figure 3, with the chatbot appearance of an animated human.

Figure 3.

Human and robotic language conversation with animated picture



After participants watched one of the six conditions, an online questionnaire followed in order to measure the effects. The questionnaire of the pre-test can be found in Appendix A.

3.3. Pre-test

In order to decrease possible side effects, a pre-test was performed. The pre-test had the aim to pinpoint problem areas, uncertainties, reduce measurement errors and to determine whether or not respondents were interpreting the survey questions correctly, and ensure that the order of questions was not influencing the way the respondent answered.

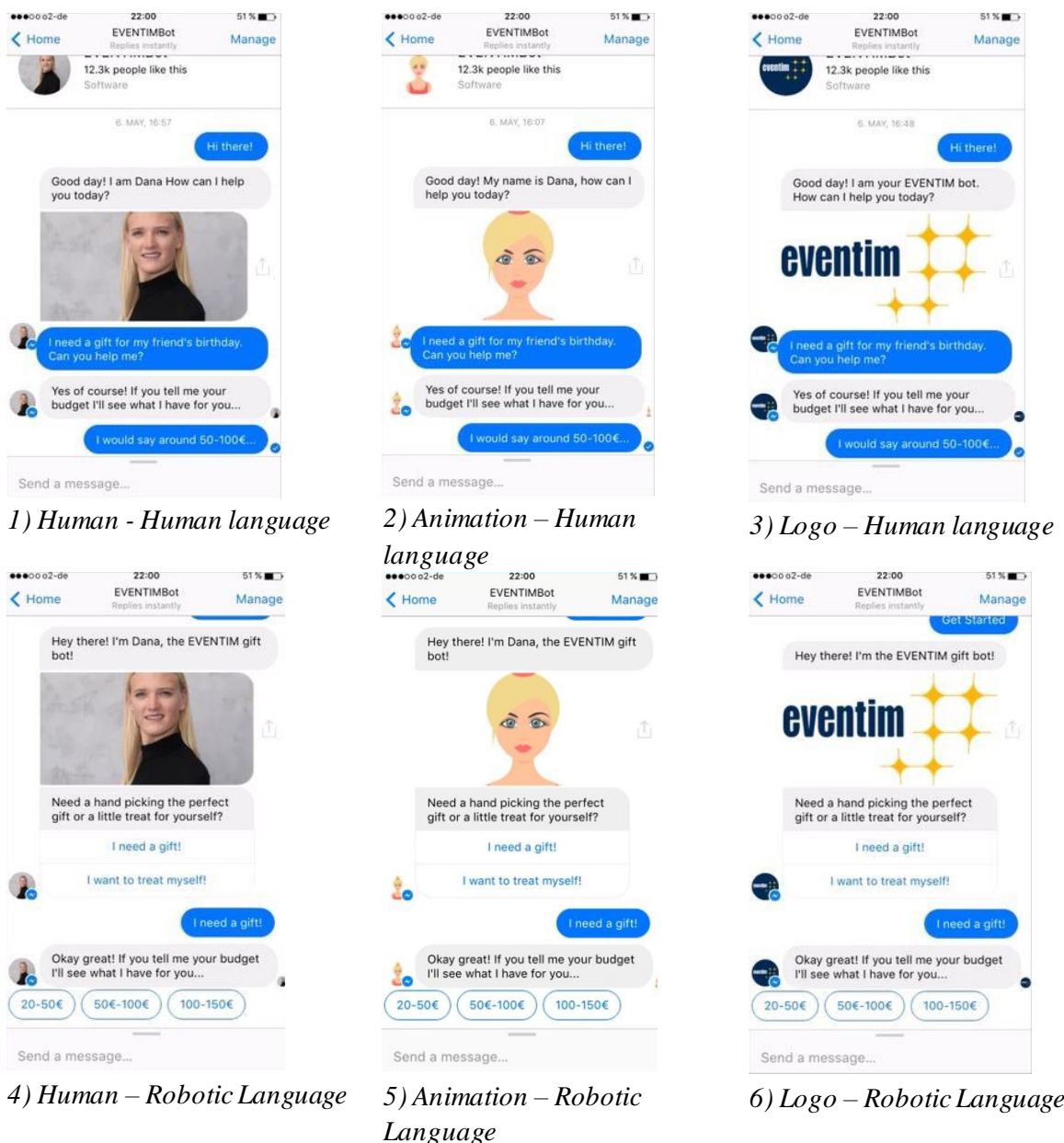
A non-probability sample of 16 people participated in the pre-test. Each responded was exposed to one of the six conditions. This test was conducted in order to examine if the human picture, animated picture, and logo are correctly perceived as either human, animation or logo. Moreover, each question of the survey was pre-tested in order to reduce ambiguity and errors. Further, each participant performed the questionnaire and was asked for feedback to identify mistakes and unclarity. Amendments, such as decelerating the speed of the videos (to 40-50 seconds), rephrasing questions and highlight different wordings, were made based on the feedback from the participants to ensure that all items are apparent and correct.

3.4. Final Stimuli

After adjusting the feedback into the survey and videos, the main study took place. In the main study, a total of 265 respondents completed the questionnaire. Each participant was randomly assigned to one of the six conditions and was exposed to either a human, animated or logo picture which either displayed robotic or human language. After watching the chatbot interaction video, a questionnaire was used to measure the variables. The main study tests if the dependent variables are influenced by the two independent variables and their conditions. Figure 4 depicts screenshots of the 3x2 videos from the main study. Appendix B gives a full overview and access to the designed chatbot interfaces and videos.

Figure 4.

Screenshots of the six different chatbot videos.



3.5. Manipulation Check

For this study, a manipulation check was performed as an indicator of the internal validity of this experiment. The manipulation check was conducted in order to investigate if the manipulation of the chatbot appearance and language. Firstly, a manipulation check with a one-way ANOVA and Post Hoc test was performed for chatbot appearance, followed by a t-test for language.

3.5.1. Manipulation for Chatbot Appearance

Within the survey of this experiment, the participants had to answer 7 items on a 5-point Likert scale about chatbot appearance. The semantic scale ranged from 1 'strongly agree' to 5 'strongly disagree'. Due to that measurement, the lower the mean value, the higher the perceived anthropomorphism of the chatbot appearance. In order to determine if there are any statistically significant differences between the means of the three groups of the chatbot appearance, namely human, animation and logo, a one-way ANOVA was performed. Firstly, the seven chatbot appearance items were combined with their means as a new variable in SPSS. Thereafter, a one-way ANOVA was performed to check if the three groups have significant differences between the means. Further, a Post Hoc test was executed to explore where the differences occurred between the groups.

Looking at the results of the ANOVA test, it can be stated that there was a significant effect for three conditions. To be precise, the values show that there are significant differences $M = 1.91$, with $F(2, 221) = 13.13$, $p < 0.001$ between the three groups. Further, a Post Hoc test was conducted to confirm where the differences occurred between the groups. The results of the test indicate that there is a significant difference between the human and animated group, $p < 0.001$, and the human and logo group $p < 0.001$. However, the Post Hoc test also revealed that there is no significant difference between the animation and logo group ($p = 0.954$). The results confirm the assumption that the chatbot using a human picture is perceived as more human than the animated picture and the Eventim logo.

3.5.2. Manipulation for Chatbot Language

Next to the groups of chatbot appearance, this research created two different language groups (human and robotic) for the independent variable 'language'. Participants of the survey had to answer 7 items on a 5-point Likert scale about their perceived anthropomorphism of the chatbot language. Likewise, the semantic scale ranged from 1 'strongly agree' to 5 'strongly disagree' and shows that a low mean value can be translated into a high perception of anthropomorphism. An independent samples t-test was conducted to confirm that there were significant differences between the means of the two language groups. The results of the t-test showed that there were significant differences between the robotic language ($M = 2.44$, $SD = 0.91$) and the human language ($M = 1.84$, $SD = 0.91$), with $t(222) = 5.52$, $p < 0.001$. These statistically significant results suggest that respondents recognized the two different

language styles within the study. However, the difference between the two mean values of the language groups are not as big as expected.

3.6. Respondents

For this experiment, a total of 267 participants have filled in the questionnaire. However, 43 questionnaires were deleted due to incomplete answers or participants who did not fit into the criteria of this research. Thus, the used data set from this study is from 224 respondents. Since chatbots are increasingly implemented by big brands in Germany (Mehner, 2018), the Netherlands (Schurer, 2017) and on a global scale (Suthar, 2019), the participants of this study were mostly, but not limited to, German and Dutch citizens. They were males and females with a minimum age of 18 years and a maximum of 62. According to social media statistics, it is stated that the main audience of Facebook and the messenger lays between 18-64-year-old (West, 2019). Further, a Ticketmaster study revealed that mostly millennials and boomers attend concerts, who lay in the same age range as the Facebook users (Peoples, 2015). For that reason, participants who were older than 64 or younger than 18 years, had to be eliminated from the dataset. The mean of the participants' age scored $M = 24.5$ years, $SD = 7.5$. Further, most respondents, in fact, 74 of them, hold a high school degree, followed by 61 people who have a bachelor's degree. In addition, 51 of the respondents have some college credit, but no degree and 19 people hold a master degree. Lastly, 12 respondents have an associate's degree, 2 people of this study have less than a high school degree and 2 hold a PhD degree. Moreover, 83% use the Facebook messenger and have interacted with a chatbot before. The respondents were equally divided into the six conditions of this research. There are no significant differences between the participants in the six different conditions. Hence, the participants' data can be used for further analyses and evaluations.

3.7. Procedure

Prior to commencing the study, ethical approval was sought from the ethical committee of the University of Twente. The survey for the pre-test and main study was designed in English, in order to not only limit the respondents to German and Dutch citizens. The online survey was created with the tool Qualtrics and the chatbot interaction videos were designed with the online tool botpreview. The survey of the main study was distributed through online channels, such as email, social media (Facebook, LinkedIn, Reddit, WhatsApp) and the university platform SONA. Moreover, students of the University of Twente were asked in person to fill out the survey on the campus of the University of Twente.

Before filling out the survey, respondents had to read an introduction about the study, their data protection, the right to stop the study at any point and lastly, had to give their consent to participate voluntarily. If a respondent did not agree, the questionnaire was closed automatically. After obtaining the consent of the participants, a questionnaire regarding the demographics was depicted.

Thereafter, participants had to read through a scenario in which they had to imagine to be in a situation in which the participants would like to purchase concert tickets as a present for a friend's birthday. The participants were asked to imagine to be the person who interacts with the chatbot and to watch the video carefully. After that, one of the six videos of the chatbot interactions were randomly assigned to the participant. During the video, the chatbot appeared with either a human, animated or logo picture. Next to the three different options of the chatbot appearance, the chatbot either performed the conversation with robotic or human language.

After the confrontation with the video, respondents were asked to fill out the questionnaire. The survey had questions sets regarding chatbot appearance, language, satisfaction, purchase intention, and trust (although the trust data was not used for further analyses and evaluations).

3.8. Measurement Instruments

At the start of the survey, the standard demographic set from Qualtrics was portrayed. The questionnaire for this study used a 5-point Likert scale ranging from 'strongly disagree' to 'strongly agree' and one time from 'extremely unlikely' to 'extremely likely'. Values, such as 'strongly agree' and 'extremely likely' were coded as 1, whereas 'extremely unlikely' and 'strongly disagree' were coded with 5. Hence, the lower the values are in the analysis, the higher the actual result. In total, the survey consisted of 44 questions, measuring five different variables. Moreover, only existing scales for the measurement of the variables were used. For the two independent variables, the same existing question sets were implemented into the survey to measure the perceived anthropomorphism. Further, some survey questions were slightly rephrased due to the feedback of the pre-test. The complete questionnaire can be found in Appendix A.

Chatbot appearance

Respondents were asked to rate two key concepts of human-robot interaction (HRI) after being confronted with either a human, animated or logo picture. These two concepts are anthropomorphism and animacy. To test the degree of perceived anthropomorphism, the human likeness item scale was used. This questionnaire was used since this research aims to test a human-robot interaction. These two concepts can measure the human perceptions of robots that they interact with (Bartneck, Kulic, Croft, & Zoghbi, 2009). The seven questions of the two HRI concepts were displayed with a 5-point semantic scale, ranging from 1 = 'strongly agree' to 5 = 'strongly disagree'. An example question of this question set was: 'The impression of the chatbot's picture felt alive'.

Chatbot Language

Likewise, the variable of chatbot language aimed to measure if participants perceive the chatbot language as a human or robot. Hence, the perception of anthropomorphism of the chatbots language was similarly measured with the seven-item anthropomorphism and animacy question set (Bartneck, Kulic,

Croft, & Zoghbi, 2009). Thus, the same scale applies, ranging from 1 = 'strongly agree' to 5 = 'strongly disagree'. For instance, respondents had to answer questions such as 'My impression of the chatbot's language felt humanlike' with this answer scale.

Trust

The level of trust can be measured with the propensity to trust question set. This is a scale conceived to measure a stable and unique trait of an individual, which helps to provide useful insights to predict the initial level of trust on robots (Yagoda, 2012). The trustworthy scale, strongly relates to the robot type, level of automation, animacy and perceived function (Lee & See, 2004) and can be used to measure the human-robot trust during an interaction. Four questions from that question set were depicted, using a 5-point Likert scale ranging from 1 = 'strongly agree' to 5 = 'strongly disagree'. An example item of this scale was: 'The chatbot was reliable'.

Customer satisfaction

This study only tested the overall user satisfaction with the chatbot interface interaction. In order to measure the users' trust, Anderson and Srinivasan (2003) use Oliver's (1980) multi-item scale to measure customer satisfaction in an online environment. Since this research specifically focuses on the human-chatbot interaction, only the overall user satisfaction questions of this questionnaire were asked and slightly modified, which consisted of 3 items. These satisfaction items were depicted on a 5-point Likert scale ranging from 1 = 'strongly agree' to 5 = 'strongly disagree'. One of the corresponding questions was 'I am satisfied with the chatbot interface'.

Purchase intention

The variable purchase intention is supposed to measure to what extent respondents were willing to buy a concert ticket from Eventim after the chatbot interface interaction. The measures aim to identify the online visitors' behavioural intentions in the near future and in six months, represented as three-items on a 5-point Likert scale. The respondents were asked if they are willing to buy tickets from Eventim either now, in three or six months, ranging from 1 = 'very likely' to 5 = 'very unlikely' (Gefen & Straub, 2004).

3.9. Construct Validity and Reliability

In order to investigate how the items of this research performed in relation to other variables, a construct validity test was conducted by means of a validity factor analysis, explained variance, the eigenvalues and a calculation of Cronbach's alphas to investigate the reliability.

3.9.1. Validity

To prove the validity of the study, a factor analysis was performed. In total, 24 items, separated by 5 factors, which are two dependent and three independent variables, were analysed. The aim was to find out whether or not the variables measure, what they were supposed to measure and if the five factors would be distributed in the expected five constructs.

Table 2 shows the conducted SPSS factor analysis, wherein the dependent and independent variables are depicted. The items from the variables chatbot appearance, language, trust, satisfaction, and purchase intention were portrayed. The items of the variables chatbot appearance, language, and purchase intention ended up in one factor column, which means the items measured what they were supposed to measure. However, the items of the variable satisfaction and trust were found in the same column and hence, the two variables measured the same factor instead of two separate ones. Trust and satisfaction strongly correlated within their measurements; however, it is not conceptually possible to merge them into one construct. It was stated in the hypotheses that trust had an expected mediating effect, which has to be rejected due to the elimination of the variable. Furthermore, all hypotheses containing the effects of trust have to be rejected.

Moreover, the explained variance of all variables scores 68,33%. Higher percentages of explained variance indicate a stronger strength of association. Hence, the explained variance of this research scored relatively high and is therefore acceptable. Although the explained variance did not score over 70%, it can still be significant, since it is indicating that the regression model has statistically significant explanatory power.

In addition, the eigenvalues show the strength of a transformation in a particular direction. Each eigenvalue for every factor of this study is over and above 1, which indicates that the items of this research are valid.

Table 2.

Validity factor analysis

Item	Factor			
	1	2	3	4
Factor 1: Chatbot appearance ($\alpha = .914$)				
The impression of the chatbot's picture felt alive.		.81		
The impression of the chatbot's picture felt lively.		.78		
The impression of the chatbot's picture felt natural.		.79		
The impression of the chatbot's picture felt interactive.		.71		
My impression of the chatbot's picture felt natural.		.81		
My impression of the chatbot's picture felt humanlike.		.79		
My impression of the chatbot's picture felt lifelike		.83		
Factor 2: Chatbot language ($\alpha = .933$)				
My impression of the chatbot's language felt natural.	.85			
My impression of the chatbot's language felt humanlike.	.85			
My impression of the chatbot's language felt lifelike	.85			
My impression of the chatbot's language felt alive.	.82			
My impression of the chatbot's language felt lively.	.76			
My impression of the chatbot's language felt natural.	.80			
My impression of the chatbot's language felt interactive.	.69			
Factor 3: Trust ($\alpha = .840$)				
The chatbot was reliable.			.77	
The chatbot was dependable.			.68	
The chatbot was competent.			.79	
The chatbot was able.			.79	
Factor 4: Satisfaction ($\alpha = .809$)				
I am satisfied with the chatbot interface.			.57	
My choice to ask the chatbot questions for a gift was a wise one.			.63	
I am satisfied with the way the chatbot helped me.			.70	
Factor 5: Purchase Intention ($\alpha = .884$)				
I am very likely to buy a ticket from Eventim.				.76
I intend to buy a ticket within 3 months from Eventim.				.91
I intend to buy a ticket within 6 months from Eventim.				.90

3.9.2. Reliability

Furthermore, the reliability of this research was tested. The Cronbach's Alpha from the variables were calculated to find out more about the internal consistency, that is, how closely related a set of items are as a group. An overview of the Cronbach Alpha's of each variable is depicted in table 2 of the validity factor analysis. Each Cronbach Alpha from the four variables, chatbot appearance, language, satisfaction and purchase intention score over and above 0.7 which suggests that the items have relatively high internal consistency and therefore confirm an acceptable value.

4. Results

In this study, six different conditions were designed in order to examine the influence of the different conditions on the dependent variable satisfaction, purchase intention and trust (although this variable has been eliminated and will not be used from this point anymore). Firstly, a multivariate analysis of variance was performed, with the aim to analyse the survey data that involved more than one dependent variable at a time. Moreover, the MANOVA explored the hypotheses regarding the effects of the two independent variables (chatbot appearance and language) on the two dependent variables (satisfaction and purchase intention). Further, the main effects of language and chatbot appearance on satisfaction and purchase intention are elaborated. In the following section the not existing interaction effect is explained further. Lastly, an overview of the supported or rejected hypotheses is portrayed.

4.1. Multivariate Analysis of Variance

This study used a MANOVA analysis to examine the different effects of the independent variables (chatbot appearance and language) on the dependent variables (trust, satisfaction and purchase intention). The results are presented separately for each dependent variable.

Before analysing the MANOVA results, a Wilks' Lambda was performed to check the overall effect between the two independent variables. Table 3 depicts the descriptive statistics of the independent variables, chatbot appearance and language. Looking at the first main effect of this study, chatbot language, it can be stated that there is a main effect of chatbot language, with $\Lambda = 0.952$, $F(3, 216) = 3.625$, $p = 0.014$. Moreover, the second main effect, chatbot appearance, has the significance value of $\Lambda = 0.945$, $F(6, 432) = 2.084$, $p = 0.054$, which means that there is no main effect of chatbot appearance. However, the p-value is almost significant ($p = 0.054$) and hence, it will be used further within this analysis and will not be eliminated, although it is an insignificant result. Furthermore, looking at the interaction of the two main effects, language * chatbot appearance, it indicates that there is no significant interaction effect between the two groups, with $\Lambda = 0.961$, $F(6, 432) = 1.443$, $p = 0.197$.

The hypotheses of this research expected that the variable language moderates the impact of chatbot appearance. However, table 3 shows that there is no significant interaction effect between the two independent variables, and therefore there is no moderation effect as it was expected. For this reason, H7 has to be rejected.

Table 3.

Multivariate Test; Descriptive statistics of the independent variables

Effect		Value	F	Sig.	Partial Eta Squared
Language_Group	Wilks' Lambda	.952	3.625	.014	.048
Chatbotbot_Group	Wilks' Lambda	.945	2.084	.054	.028
Language_Group *	Wilks' Lambda	.961	1.443	.197	.020
Chatbot_Group					

4.2. Main Effects of Chatbot Appearance and Language

The main effects of the independent variables, language and chatbot appearance, on the dependent variables (satisfaction, purchase intention) were explored further in the following sections.

4.2.1. Effects on Satisfaction

As depicted in table 4, the effects between the independent variables (chatbot appearance and language) and dependent variables (satisfaction, purchase intention) were measured. The results indicate that there is a main effect between language and satisfaction, with $p = 0.033$ ($F = 4.612$), and between chatbot appearance and satisfaction, with $p = 0.016$ ($F = 4.192$).

The satisfaction items of this study were measured with a 5-point Likert scale ranging from 1 = 'strongly agree' to 5 'strongly disagree'. The Bonferroni results of the MANOVA reveal that there is a difference between the three chatbot groups (human, animation, logo). Respondents of the survey were the most satisfied with a human chatbot appearance ($M = 1.84$, $SD = 0.085$), followed by the animated picture ($M = 1.95$, $SD = 0.081$) and lastly the logo ($M = 2.95$, $SD = 0.082$), with $F(2, 218) = 4.192$, $p = 0.016$.

It was hypothesized that participants have a higher perception of satisfaction when they are confronted with a chatbot which displays a human picture compared to people interacting with a chatbot which displays an animated picture or organizational logo. Hence, the hypothesis (H1b) that the perception of satisfaction is higher when people are confronted with a chatbot using a human picture compared to people using a chatbot with an animated picture or organizational logo, can be confirmed due to the results.

Moreover, respondents of this study were the most satisfied with the human language of a chatbot ($M = 1.87$, $SD = 0.067$) in comparison to the robotic language ($M = 2.08$, $SD = 0.067$), with $F(1, 218) = 4.612$, $p = 0.033$.

It was hypothesized that participants have a higher perception of satisfaction when they are confronted with a chatbot using human language, compared to people interacting with a chatbot which

uses robotic language. Hence, the hypothesis of H2b can be confirmed according to the investigated results.

4.2.2. Effects on Purchase Intention

In table 4, the results for purchase intention show that there is no main effect between language and purchase intention, with $p = 0.137$ ($F = 2.223$). Likewise, there is no main effect between chatbot appearance and purchase intention, with $p = 0.262$ ($F = 1.346$).

It was hypothesized that participants have a higher intent to purchase when they are confronted with a chatbot which displays a human picture compared to people interacting with a chatbot which displays an animated picture or organizational logo. Moreover, it was expected that participants have a higher intent to purchase when they are confronted with a chatbot using natural language, compared to people interacting with a chatbot which uses robotic language.

However, since there is no main effect of language or chatbot appearance on purchase intention the hypothesis H1c and H2c have to be rejected.

Table 4.

Test of Between-Subjects Effects

Source	Dependent Variable	df	F	Sig.
Language_Group	Satisfaction	1	4.612	.033
	Purchase Intention	1	2.246	.137
Chatbotbot_Group	Satisfaction	2	4.192	.016
	Purchase Intention	2	1.346	.262

4.3. Interaction Effect of Chatbot Appearance and Language

Although it was already stated that there is no interaction effect between chatbot appearance and language, a MANOVA was performed to measure the interaction effects of chatbot appearance and language on the dependent variables (satisfaction, purchase intention). Looking at the interaction between the two main effects (language and chatbot appearance) in table 4, it indicates that there is no significant interaction effect between chatbot appearance and language, with $\Lambda = 0.961$, $F(6, 432) = 1.443$, $p = 0.197$. Moreover, table 5 shows that there is no interaction effect of chatbot appearance * language on the two different dependent variables with $p = 0.175$ ($F = 1.756$) for satisfaction and $p = 0.288$ ($F = 1.260$).

It was expected from two hypotheses that certain combinations of the three chatbot appearances and language types would lead to different effects of the dependent variable satisfaction and purchase

intention. However, since the effect of chatbot appearance * language on the dependent variables are insignificant, H3 and H4 have to be rejected.

Table 5.

Test of Between-Subjects Effects

Source	Dependent Variable	df	F	Sig.
Language_Group *	Satisfaction	2	1.756	.175
Chatbotbot_Group	Purchase Intention	2	1.260	.288

4.4. Overview of the Hypotheses

Table 6 gives an overview of the 7 hypotheses and whether or not they are supported or rejected by the results of this research.

Table 6.

Overview of hypotheses

Hypotheses	Supported?
H1a The perception of trust is higher when people are confronted with a chatbot using a human picture compared to people using a chatbot with an animated picture or organizational logo.	No
H1b The perception satisfaction is higher when people are confronted with a chatbot using a human picture compared to people using a chatbot with an animated picture or organizational logo.	Yes
H1c The perception of purchase intention is higher when people are confronted with a chatbot using a human picture compared to people using a chatbot with an animated picture or organizational logo.	No
H2a The perception of trust is higher when people are confronted with a chatbot using natural language compared to people using a chatbot with machine-like language.	No
H2b The perception of satisfaction is higher when people are confronted with a chatbot using natural language compared to people using a chatbot with machine-like language.	Yes
H2c The perception of purchase intention is higher when people are confronted with a chatbot using natural language compared to people using a chatbot with machine-like language.	No
H3a The perception of trust is higher when people are confronted with a chatbot using a human picture and natural language compared to people using a chatbot using a logo or animated picture with human-like language.	No
H3b The perception of satisfaction is higher when people are confronted with a chatbot using a human picture and natural language compared to people using a chatbot using a logo or animated picture with human-like language.	No

H3c	The perception of purchase intention is higher when people are confronted with a chatbot using a human picture and natural language compared to people using a chatbot using a logo or animated picture with human-like language.	No
H4a	The perception of trust is higher when people are confronted with a chatbot using robotic language with either an animated picture or logo compared to people using a chatbot with robotic language and human-like language.	No
H4b	The perception of satisfaction is higher when people are confronted with a chatbot using robotic language with either an animated picture or logo compared to people using a chatbot with robotic language and human-like language.	No
H4c	The perception of purchase intention is higher when people are confronted with a chatbot using robotic language with either an animated picture or logo compared to people using a chatbot with robotic language and human-like language.	No
H5a	The effects of chatbot appearance on satisfaction are mediated by trust.	No
H5b	The effects of chatbot appearance on purchase intention are mediated by trust.	No
H6a	The effects of natural language on satisfaction are mediated by trust.	No
H6b	The effects of natural language on purchase intention are mediated by trust.	No
H7a	Language moderates the impact of chatbot appearance on trust	No
H7b	Language moderates the impact of chatbot appearance on satisfaction	No
H7c	Language moderates the impact of chatbot appearance on purchase intention.	No

5. Discussion

The aim of this study was to gain a better insight into the effects of using anthropomorphic characteristics in the chatbot appearance and language by either implementing visual or verbal human cues. Additionally, the effect of trust, satisfaction and purchase intention in regards to online messenger chatbot interactions were explored. Six different conditions were created to test the effect of anthropomorphism for chatbot appearances (human, animation, and logo) and language (human, robotic) on perceived trust, satisfaction, and purchase intention. Moreover, language as a moderator and trust as a mediator were examined. To research these objectives, the following four research questions have been proposed (1) ‘To what extent does chatbot appearance influence trust, satisfaction, and purchase intention?’; (2) ‘To what extent does robotic/human language influence trust, satisfaction, and purchase intention?’; (3) ‘To what extent are the effects of a chatbot appearance on trust, customer satisfaction, and purchase intention dependent on the robotic/human language used for the interaction?’; (4) ‘To what extent are the effects of chatbot appearance and robotic/human language on trust, satisfaction and purchase intention mediated by trust?’. These research questions were answered with an experimental design and seven main hypotheses.

5.1. Discussion of Results

5.1.1. Discussion of Main Effects

The most important relevant findings of this research were the two confirmed hypotheses of a human picture and human language leading to higher satisfaction levels among users. Researched literature showed that human-like cues, such as the style of language or visual appearance can influence and increase the perception of anthropomorphism (Araujo, 2018) and therefore lead to an improvement in customer satisfaction (Higashinaka, Minami, Dohsaka & Meguro, 2010). Observed studies suggest implementing anthropomorphic visual cues by using a human picture increases the perception of social presence, which positively influences satisfaction (Hassaein & Head, 2007). Due to this literature finding, it was expected that participants of this research are more satisfied with a chatbot that displays a human picture, instead of an animation or logo. Moreover, it was expected that respondents are more satisfied with the anthropomorphic linguistic features of the human language group, in comparison to the robotic language group. The results of this study did show that respondents of this research were the most satisfied with anthropomorphic visual and linguistic chatbot features. Thus, these findings further support the idea of implementing social presence in the online environment by means of anthropomorphic features to create higher user satisfaction. In addition, this finding has important implications for developing and designing chatbots in order to create a better user experience among

users. Implementing anthropomorphic cues in chatbots creates a better user experience, and establishes happier customers.

However, these findings have to be interpreted with caution because the design of this research was overly simplified. It is possible to hypothesise that these results are less likely to occur in a more realistic setting. Participants did not actually interact with the chatbot themselves, since they only watched an interaction video and filled out the corresponding, instead of interacting with chatbots themselves. Thus, the result of this research might be consistent with the theory of social presence, however, results might be different in a more realistic setting. Contrary to the expectations of the social presence theory, the uncanny valley theory argues that photo-realistic human agents can trigger an uncanny, unsettling feeling (Tinwell, 2009). The theory describes that human-like cues, such as those anthropomorphic cues in chatbots, are seen as a chance to increase satisfaction among users until they become so human that users find their robotic flaws disturbing and upsetting (Mori, 1970).

Moreover, observed literature indicated that online visuals, such as the chatbot appearance, as well as textual information, are factors that can lead to a higher purchase intention (Toma, 2010). Therefore, it was expected that human pictures in the variable chatbot appearance would lead to a higher purchase intention of Eventim. In addition, this research expected that the human language group would score higher in the intent to purchase in comparison to the robotic group.

However, the results of this research have to reject this assumption since there is no main effect of language or chatbot appearance on purchase intention. According to the Theory of Planned Behaviour (TPB) an individual's performance of a certain behaviour, such as purchasing a product or a ticket from Eventim, is determined by his or her intent to perform that behaviour (Georg, 2014). Although this research elected the age range of respondents who fit into the criteria for people who are going to concerts, this experiment did not measure the actual interest to purchase concert tickets. Hence, the outcome of the insignificant results might have been biased, since the video of the chatbot interaction displayed a scenario in which the participants had the opportunity to buy Hip Hop concert tickets. Conclusively, this research did not take into account whether or not the participants had an overall interest in purchasing concert tickets. In addition, it was not measured if respondents of this survey enjoy Hip Hop music since the suggested concert tickets were from Drake. Thus, these limitations might explain the insignificant outcome of the research.

Further, research has shown that users' trust in chatbots was found to be affected by factors concerning the specific chatbot appearance, specifically the quality of its human-likeness, which both can be embodied by different types of chatbot appearance and natural language (Følstad, Nordheim & Bjørkli, 2018). Consequently, it was expected that the anthropomorphic cues would lead to higher trust perceptions. However, the factor analysis showed that the trust results were not statistically significant and correlated with satisfaction. It was conceptually not possible to merge them and therefore had to be eliminated. The trust data was not used anymore for further analyses and evaluations. A reason for the insignificant results might be that this research only used four trust items to get an insight to predict the

initial level of trust on chatbots from the participant. This scale was oversimplified to measure online trust. Other items and factors should have been taken into account, such as propensity to trust, the knowledge-based trust, such as familiarity or reputation of the brand Eventim. Moreover, chatbots collect sensitive data and hence perceived security and privacy might have influenced the perception of trust towards the chatbot as well. For this reason, the trust variable could not be explored further.

Concludingly, anthropomorphic visual and linguistic features have resulted in higher satisfaction among users within this study. This finding supports the theory of social presence and shows that the application of this theory applies similarly to the chatbot m-commerce environment. Moreover, it helps chatbot designers as well as companies to develop m-commerce chatbots that satisfy customers. However, this experiment should be executed in a more realistic and improved setting in the future, in terms of research design and measurements. Especially the measurement instruments for purchase intention and trust have to be improved. In addition, having a real interaction between humans and chatbot would test if the findings would be similar to this research results since the theory of the uncanny valley argues that too many anthropomorphic cues can lead to the opposite results of this study.

5.1.2. Discussion of Moderating Effect

The visual appearance of chatbots and the importance of the displayed language is seen as a crucial synergy (Razaque & Yang, 2018). Moreover, in previous studies the tone of voice was found to have a moderating effect on consumer responses in social media (Keyzer, Dens, & Pelsmacker, 2017) as well as having an interaction between the tone of voice and the human facial features (Zuckerman, Amidon, Bishop, & Pomerantz, 1982). Thus, the impact of the chatbot appearance was expected to be dependent on the chatbot language. The insignificant results of this research, however, have to reject this assumption. An implication of this is the possibility to design simpler chatbot language features that do not have to be aligned with the appearance of the chatbot since chatbot language does not have a moderating effect.

Nonetheless, it is important to bear in mind the possible bias in these responses since they did not actually interact with the chatbots themselves. Moreover, a possible explanation for these insignificant results may be the lack of adequate research measurements. Respondents had the same measurement scale for chatbot appearance and chatbot language which might not have led to accurate results. Future research should focus on more linguistic elements such as tone of voice, personality or empathy. More research on this topic needs to be undertaken before the association between chatbot appearance and chatbot language cues are more clearly understood.

5.1.3. Discussion of Mediation Effect

There are several research initiatives that have elaborated on the factors affecting online purchase intention and satisfaction such as trustworthiness (Adam, Aderet & Sadeh, 2008), perceived risk and consumer trust (Kim, Ferrin, & Rao, 2008). Moreover, the determinants of satisfaction or purchase intention are factors such as customer trust from the customer's perspective (Bejou, Ennew, & Palmer, 1998). In addition, observed studies have indicated that trust plays a vital role in relation to the chatbot appearance and language in relation to purchase intention and satisfaction. Hence, trust was expected to have a mediation effect within this research. However, the factor analysis showed that these results were not statistically significant and the entire variable trust had to be taken out of this experiment. The question if the effects of chatbot appearance or natural language on satisfaction and purchase intention are mediated by trust remain unanswered at present and thus, need future research.

5.2. Implications

5.2.1. Practical Implications

The aim of this research was to give organizations, chatbot developers and marketing strategists a helpful insight into the right design features of chatbots. This study resulted in a better understanding of anthropomorphic visual and linguistic cues for conversational agents in a mobile commerce setting. The results of this study can be used for marketing implementations, chatbot designers or developers in order to create a better user experience among users. This research found out that implementing visual and linguistic anthropomorphism cues, by means of a human picture or a more human language, leads to higher satisfaction and user experience.

This finding has important implications for developing and designing chatbots. Developers are encouraged to test and implement anthropomorphic features into their chatbot design. It can be stated that users do prefer human cues over simple organizational logos according to the findings of this research. Logos are predominately used by organizations, which scored the lowest in satisfaction within this research. Chatbot developers and designers can use this insight to experiment with different levels of anthropomorphism in the appearance of a chatbot as well as in the conversational tone to improve the user experience. Further, it is possible that chatbots interactions improve to a degree in which costumers choose the conversational agent over human interactions and human to human interaction might no longer be required in customer service support. This gives companies the opportunity to offer 24/7 customer service and reduces costs.

However, since there is a thin line between increasing satisfaction and experiencing the uncanny valley effect, it is recommended to test the chatbots anthropomorphism features before publishing them.

Moreover, it is advised to implement a statement or greeting in which the chatbot states that the user is interacting with an intelligent agent to avoid the uncanny valley effect.

5.2.2 Theoretical Implications

Only some research has been carried out on chatbots, but there have been only a few empirical investigations into chatbot appearances and linguistic features for online commerce. This study focused on exploring anthropomorphism cues for chatbot appearance and language features in a Facebook messenger setting. Hence, it adds to the field of research about chatbot design implications of appearance and language on trust, satisfaction and purchase intention. The findings of this research can be used as a foundation to examine the effects of a human chatbot appearance and language features within an m-commerce setting. This combination of findings provides some support for the conceptual premise that implementing anthropomorphism cues in visual and linguistic appearance leads to higher satisfaction and user experience.

Moreover, findings of this research show that the theory of social presence is applicable for chatbots and within the mobile commerce environment. Thus, it supports previous research from Hassaein and Head (2007) who suggest implementing anthropomorphic visual cues by using a human picture to increase satisfaction. The findings of this research may help to understand which anthropomorphism cues can lead to an increase in satisfaction and a better user experience. Taking the social presence theory a step further, this research might help to build the groundwork for the application of the social presence theory for linguistic anthropomorphism cues. Future research should focus on more linguistic elements such as tone of voice, personality or empathy. More research on this topic needs to be undertaken before the association between chatbot appearance and chatbot language cues are more clearly understood.

Further, next to the expectations of the social presence theory, this research shows that anthropomorphism cues with photo-realistic human agents do not lead to an unsettling feeling in a m-commerce chatbot environment, unlike the uncanny valley theory suggest. The theory describes that human-like cues, such as those anthropomorphic cues in chatbots, are seen as a chance to increase satisfaction among users until they become so human that users find their robotic flaws disturbing and upsetting (Mori, 1970). The findings of this research suggest that human-like cues within the chatbot environment can be implemented without leading to the users' feeling of being disturbed or upset.

5.3. Limitations and Recommendations for Future Research

Finally, a number of important limitations need to be considered. Firstly, this study had a small size for pre-test which only included family members and friends of the researcher. Moreover, some of the participants knew the person who was depicted in the chatbot picture. This limitation might have led to different outcomes and to invalid results of this study.

Secondly, the literature of Amdocs (2017) suggests that consumers prefer the female gender in chatbot appearances. Hence, this research only used a female chatbot picture and animation. However, the gender of the chatbot might also influence the outcome of the test. In further research, a pre-test should be executed to see if the gender of the chatbot has a different effect on the dependent variables.

Thirdly, the study is limited by the lack of information on demographics. Respondents were not asked about their gender, which might have led to limitations or effects that cannot be explained. For instance, it might be possible that women trust chatbots significantly more than men do or vice versa. Moreover, this research elected the age range of respondents who fit into the criteria of people who are interested in purchasing concert tickets. However, this experiment did not measure the actual interest in concerts or the genre of hip hop music, which might have influenced the insignificant outcome of the variable purchase intention. Future research has to obtain more demographic information from the respondents to test the effects of gender on chatbot variables. Moreover, the respondents of future studies have to be more critically selected and asked about their interest in the product (or ticket) that is advertised in a chatbot conversation. For that reason, actual chatbot-human interaction with participants would be interesting to investigate further. In this type of conversation, participants might have an actual interest in the product that they are researching with the chatbot since they can steer the conversation and express their interest. Especially the results of purchase intention might change through this change in the research design.

Fourthly, the conversation was displayed as a Facebook messenger interface conversation, represented as a video to the participants. Hence, participants did not interact with the chatbot themselves but had to imagine to be the conversational partner of the chatbot whilst watching the interaction as a video. Participants only judged the language that was displayed in the videos instead of judging a real conversation with a chatbot. Thus, the results of this study are not as reliable in comparison to an actual conversation between a participant and a chatbot. Further, it is possible that participants were not paying attention during the video. Concludingly, it would be interesting to try out this experiment but with a real interaction between the bot and the participant. However, future research about interaction effects between anthropomorphic characteristics in the linguistic and visual features of chatbot interactions are very scarce and need to be further explored in the future.

Fifthly, the scope of this study was limited in terms of eliminating the trust variable. Since trust correlated with satisfaction, it had to be taken out of this research. The scale was oversimplified to measure the trust in the displayed chatbot. Other items and factors should have been taken into account,

such as propensity to trust, the knowledge-based trust, such as familiarity or reputation of the brand Eventim. Moreover, chatbots collect sensitive data and hence perceived security and privacy might have influenced the perception of trust towards the chatbot as well. Thus, future research should implement more elements of trust and should not oversimplify the survey questions.

5.4. Conclusion

The present study was designed to gain a better insight into the effects of using either human or robotic characteristics in chatbot appearances and language within the online environment. Additionally, the effects of chatbot appearance and language on trust, satisfaction, and purchase in m-commerce were explored. The most noticeable finding was that the respondents were the most satisfied with a chatbot, using a human picture and human language. However, implementing these cues does not increase purchase intention according to the results of this research. Nonetheless, organizations and chatbot designers should strive for anthropomorphic cues within the development and implementation of chatbots to create a better user experience for users who interact with the intelligent agents. Thus, organizations, developers, designers, and marketers should focus on choosing a human-like appearance as well as anthropomorphic language cues in order to increase satisfaction and an overall better user experience.

5.5. Acknowledgments

I would like to thank my supervisor Dr. Ardion Beldad, for his patient guidance, feedback, and advice he has provided throughout my research. In addition, I would like to thank every single respondent who filled out my survey - this thesis would not have been done without you. Lastly, I would like to thank my family and friends for the support and encouragement throughout the last three years of studying at the University of Twente.

6. Literature

- Adam, A. M., Aderet, A., & Sadeh, A. (2008). Do ethics matter to e-consumers? *Journal of Internet Commerce*, 6(2), 19–34.
- Amdocs. (2017). Consumers want female and funny – but not youthful – chatbots. Retrieved from <https://www.amdocs.com/media-room/consumers-want-femaleand-Funny-not-youthful-chatbots>.
- Anderson, R. E., & Srinivasan, S. S., (2003). E-satisfaction and e-loyalty: A contingency framework. *Psychology and Marketing*, 20(2), 123–138.
- Anderson, E. W., & Sullivan, M. W. (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing science*, 12(2), 125-143.
- 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 interactions. *Advances in Human-Computer Interaction*, 2012, 13.
- Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, 85, 183-189.
- Atwell, E & Shawar, B. A. (2007). Chatbots: are they really useful?. *Ldv forum* (Vol. 22, No. 1, pp. 29-49).
- Bartneck, C., Kulic, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal Social Robotics*, 1, 71-81.
- Bejou, D., Ennew, C. T., & Palmer, A. (1998). Trust, ethics and relationship satisfaction. *International Journal of Bank Marketing*, 16(4), 170-175.
- Bickmore, T., & Cassell, J. (2001). Relational agents: a model and implementation of building user trust. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 396-403). ACM.
- Bradeško, L., & Mladenčić, D. (2012). A survey of chatbot systems through a Loebner Prize competition. In *Proceedings of Slovenian Language Technologies Society Eighth Conference of Language Technologies* (pp. 34-37).
- Brownlee, J. (2016). Conversational Interfaces, Explained.
- Buchanan, B. G. (2005). A (very) brief history of artificial intelligence. *AI Magazine*, 26(4), 53-53.

- Business Insider (2017). *Here's how chatbot metrics differ from traditional apps*, BI Intelligence. Retrieved from www.businessinsider.com/heres-how-chatbot-metrics-differ-from-traditional-apps-2016-10
- Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: concepts, evolving themes, a model. *International journal of human-computer studies*, 58(6), 737-758.
- Deshpande, A., Shahane, A., Gadre, D., Deshpande, M., & Joshi, P. M. (2017). A Survey of Various Chatbot Implementation Techniques. *International Journal of Computer Engineering and Applications*, 11.
- Dessalegn, B., & Landau, B. (2013). Interaction between language and vision: It's momentary, abstract, and it develops. *Cognition*, 127(3), 331-344. doi:10.1016/j.cognition.2013.02.003
- Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018, October). What makes users trust a chatbot for customer service? An exploratory interview study. In *International Conference on Internet Science* (pp. 194-208). Springer, Cham.
- Garcia, M. (2018). Consumers want chatbots to feel human, not look human. Retrieved from <https://www.phocuswire.com/Consumers-want-chatbots-to-feel-human-not-look-human>.
- George, J.F. (2004). The theory of planned behavior and internet purchasing. *Internet Research*, 14(3), 198-212.
- Gefen, D., & Straub, D. W. (2003). Managing user trust in B2C e-services. *e-Service Journal*, 2(2), 7-24.
- Gefen, D., & Straub, D. W. (2004). Consumer trust in B2C e-commerce and the importance of social presence: experiments in e-products and e-services. *Omega*, 32, 407-424.
- Giard, F., & Guitton, M. J. (2010). Beauty or realism: The dimensions of skin from cognitive sciences to computer graphics. *Computers in Human Behavior*, 26(6), 1748-1752.
- Giovanis, A. N., Tomaras, P., & Zondiros, D. (2013). Suppliers logistics service quality performance and its effect on retailers' behavioral intentions. *Procedia-Social and Behavioral Sciences*, 73, 302-309.
- Grazioli, S., & Jarvenpaa, S. L. (2000). Perils of internet fraud: An empirical investigation of deception and trust with experienced internet consumers. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 30(4), 395-410.
- Grewal, D., Monroe, K. B., & Krishnan, R. (1998). The effects of price-comparison advertising on buyers' perceptions of acquisition value, transaction value, and behavioral intentions. *Journal of Marketing*, 62, 46-59.
- Gul, R. (2014). The relationship between reputation, customer satisfaction, trust, and loyalty, *Journal of Public Administration and Governance*, 4(3).

- Gunawardena, C. N., & Zittle, F. J. (1997). Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American Journal of Distance Education*, 11(3), 8-26. doi:10.1080/08923649709526970.
- Gupta, V., Varshney, D., Jhamtani, H., Kedia, D., & Karwa, S. (2014, May). Identifying purchase intent from social posts. In *Eighth International AAAI Conference on Weblogs and Social Media*.
- Gustavsson, E. (2005). Virtual Servants: Stereotyping Female Front-Office Employees on the Internet. *Gender, Work and Organization*, 12(5), 400-419. doi:10.1111/j.1468-0432.2005.00281.x.
- Hald, G. (2018, February 15). 7 Benefits of using chatbots to drive your business goals. Retrieved from <https://medium.com/botsupply/7-benefits-of-using-chatbots-to-drive-your-business-goals-5a3a5e809951>.
- Hassanein, K., & Head, M. (2007). Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. *International Journal of Human-Computer Studies*, 65, 689-708.
- Higashinaka, R., Minami, Y., Dohsaka, K., & Meguro, T. (2010, October). Issues in predicting user satisfaction transitions in dialogues: Individual differences, evaluation criteria, and prediction models. In *International Workshop on Spoken Dialogue Systems Technology* (pp. 48-60). Springer, Berlin, Heidelberg.
- Hill, J., Ford, W. R., & 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.
- Hsin Chang, H., & Wen Chen, S. (2008). The impact of online store environment cues on purchase intention: Trust and perceived risk as a mediator. *Online information review*, 32(6), 818-841.
- Johnson, N., Patron, P., & Lane, D. (2007, June). The importance of trust between operator and auv: Crossing the human/computer language barrier. In *OCEANS 2007-Europe* (pp. 1-6). IEEE.
- Keyzer, F. D., Dens, N., & Pelsmacker, P. D. (2017). Don't be so emotional! How tone of voice and service type affects the relationship between message valence and consumer responses to WOM in social media. *Online Information Review*, 41(7), 905-920. doi:10.1108/oir-08-2016-0219
- Kim, D. J., Ferrin, D. L., & Rao, H. R. (2008). A trust-based consumer decision making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision Support Systems*, 44(2), 544-564.
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1), 50-80.

- Madjid, R. (2013). Customer trust as relationship mediation between customer satisfaction and loyalty at Bank Rakyat Indonesia (BRI) Southeast Sulawesi. *The International Journal of Engineering and Science (IJES)*, 2(5), 48-60.
- Mahmood, M. A., Burn, J. M., Gemoets, L. A., & Jacquez, C. (2000). Variables affecting information technology end-user satisfaction: a meta-analysis of the empirical literature. *International Journal of Human-Computer Studies*, 52(4), 751-771.
- Mayer, R., Davis, J., & Schoorman, F. (1995). An integrative model of organization trust. *Academy of Management Review*, 20(3), 709-734.
- Mehner, M. (2018, February 05). Mit diesen Bots quatschen wir gerne. Retrieved from <https://www.lead-digital.de/mit-diesen-chatbots-quatschen-wir-gerne/>
- Miller, E. (2010, January 11). Eventim enters UK ticketing market. Retrieved from <http://www.musicweek.com/news/read/eventim-enters-uk-ticketing-market/041562>
- Mirabi, V., Akbariyeh, H., & Tahmasebifard, H. (2015). A study of factors affecting on customers purchase intention. *Journal of Multidisciplinary Engineering Science and Technology (JMEST)*, 2(1).
- Mori, M. (1970). The uncanny valley. *Energy*, 7(4), 33-35. Retrieved from <http://www.movingimages.info/digitalmedia/wp-content/uploads/2010/06/MorUnc.pdf>
- Morwitz, V. G., & Schmittlein, D. (1992). Using segmentation to improve sales forecasts based on purchase intent: Which “intenders” actually buy?. *Journal of marketing research*, 29(4), 391-405.
- Muralidharan, L., de Visser, E. J., & Parasuraman, R. (2014, April). The effects of pitch contour and flanging on trust in speaking cognitive agents. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems* (pp. 2167-2172). ACM.
- Mullins, J. (2005). Whatever happened to machines that think?. *New Scientist*, 186(2496), 32-37.
- Myers, B. A. (1998). A brief history of human computer interaction technology. *Interactions*, 5(2), 44-54.
- Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 17(4), 460-469.
- Peoples, G. (2015, August 24). Millennials and Boomers Love Experiences, But Who Attends More Concerts? Retrieved from <https://www.billboard.com/articles/business/6671232/millennials-and-boomers-love-experiences-but-who-attends-more-concerts>
- 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.

- Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent Conversational agents. *arXiv preprint arXiv:1704.04579*.
- Ransbotham, S., Gerbert, P., Reeves, M., Kiron, D., & Spira, M. (2018). Artificial Intelligence In Business Gets Real. *MIT Sloan Management Review*, September, 17.
- Razaque, A., & Yang, G. (2018). Chatbot: Efficient and Utility-Based Platform. In *Intelligent Computing: Proceedings of the 2018 Computing Conference* (Vol. 1, p. 109). Springer.
- Reeves B. and Nass C. 1996. *The media equation: How people treat computers, Television and New Media Like Real People and Places*, Cambridge University Press, New York.
- Rust, R. T., & Oliver, R. L. (1994). Service quality: insights and managerial implications from the frontier. *Service quality: New directions in theory and practice*, 1-19.
- Sam, M., Fazli, M., & Tahir, M. N. H. (2009). Website quality and consumer online purchase intention of air ticket. *International Journal of Basic & Applied Sciences*, 9(10).
- Schuetzler, R. M., Grimes, G. M., Giboney, J. S., & Buckman, J. (2014). Facilitating natural conversational interactions: Lessons from a deception experiment. *Proceedings of the 35th International Conference on Information Systems*, Auckland, NZ, pp. 1-16.
- Schurer, F. (2017, May 15). Dit is de stand van zaken van chatbots in Nederland. Retrieved from <https://www.marketingfacts.nl/berichten/dit-is-de-stand-van-zaken-van-chatbots-in-nederland>
- Singh, N. (2017). Why consumers want your chatbot to be more human? Retrieved from April, 2019, from: <https://www.entrepreneur.com/article/300412>
- Statista (2017). Statista, VentureBeat, 11/2016. *The number of chatbots for the Facebook Messenger*. Retrieved from <https://de.statista.com/statistik/daten/studie/662144/umfrage/anzahl-der-verfuegbaren-chatbots-fuer-den-facebookmessenger/>, accessed 20th of May 2019.
- Suthar, S. (2019, April 22). Top 11 Chatbots Trends to Keep an Eye on in 2019. Retrieved from <https://acquire.io/blog/chatbots-trends/>
- Suzuki, N., Takeuchi, Y., Ishii, K., & Okada, M. (2003). Effects of echoic mimicry using hummed sounds on human–computer interaction. *Speech Communication*, 40(4), 559-573.
- Thusyanthy, V., & Tharanikaran, V. (2017). Antecedents and outcomes of customer satisfaction: A comprehensive review. *International Journal of Business and Management*, 12(4), 144. doi:10.5539/ijbm.v12n4p144
- Tinwell, A. (2009, July). Uncanny as usability obstacle. In *International Conference on Online Communities and Social Computing* (pp. 622-631). Springer, Berlin, Heidelberg.

- Toma, C. L. (2010). Perceptions of trustworthiness online: the role of visual and textual information. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 13-22). ACM
- Tsiotsou, R. (2006). The role of perceived product quality and overall satisfaction on purchase intentions. *International Journal of Consumer Studies*, 30(2), 207-217.
- Turing, A. M. (1950). Computing Machinery And Intelligence. *Mind*, LIX(236), 433-460. doi:10.1093/mind/lix.236.433
- Tyagi, A. (2016). Artificial Intelligence: Boon or Bane?. Available at SSRN 2836438.
- Van Bruggen, G. H., Antia, K. D., Jap, S. D., Reinartz, W. J., & Pallas, F. (2010). Managing Marketing Channel Multiplicity. *Journal of Service Rese*
- West, C. (2019, June 05). Social media demographics to drive your brand's online presence. Retrieved from <https://sproutsocial.com/insights/new-social-media-demographics/#Facebook>
- Weizenbaum, J. (1983). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 26(1), 23-28. doi:10.1145/357980.357991
- Wong, J. (2016). What is a chatbot, and should I be using one. Retrieved from <https://www.theguardian.com/technology/2016/apr/06/what-is-chat-bot-kik-bot-shopmessaging-platform>
- Yagoda, R. E., & Gillan, D. J. (2012). You want me to trust a ROBOT? The development of a human–robot interaction trust scale. *International Journal of Social Robotics*, 4(3), 235-248.
- Zadrozny, W., Budzikowska, M., Chai, J., Kambhatla, N., Levesque, S., & Nicolov, N. (2000). Natural language dialogue for personalized interaction. *Communications of the ACM*, 43(8), 116-120.
- Zuckerman, M., Amidon, M. D., Bishop, S. E., & Pomerantz, S. D. (1982). Face and tone of voice in the communication of deception. *Journal of Personality and Social Psychology*, 43(2), 347-357. doi:10.1037//0022-3514.43.2.347
- Zue, V. W., & Glass, J. R. (2000). Conversational interfaces: Advances and challenges. *Proceedings of the IEEE*, 88(8), 1166-1180.
- Zumstein, D., & Hundertmark, S. (2017). Communicating and transacting with chatbots: insights from public transport. In *16th International Conference WWW/Internet 2017, Algarve, Portugal, October 18-20, 2017* (pp. 55-62). IADIS.

7. APPENDIX

Appendix A – Final Questionnaire

Introduction

This research is part of my bachelor thesis at the University of Twente. You will read through a scenario text in which you have to engage with a chatbot video. Please take your time to read through the text carefully. Likewise, pay attention to the presented chatbot video.

It will take approximately 10 minutes of your time.

Your participation in this study is completely voluntary and you can stop any time. Further, your personal data as well as your given answers are treated anonymously and confidentially. Your data will only be used for the purpose of this research. Only the researcher and supervisor will have access to the information obtained by this study. If you have any questions or would like to have more information about this study, don't hesitate to reach out to the researcher Lena Assink

(l.m.assink@student.utwente.nl)

By clicking on the arrow below you

- Have read the information above
- Participate in this study voluntary
- Are 18 years or older

Demographics

What is your age?

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

- ☐ Less than high school
- ☐ High school graduate
- ☐ Some college credit, no degree
- ☐ Associate degree
- ☐ Bachelor degree
- ☐ Master degree
- ☐ Doctorate

Where are you from?

- ☐ Netherlands
- ☐ Germany
- ☐ Other

Text block

Imagine you would like to buy concert tickets as a present for your friend's birthday. However, you are not sure yet which tickets you would like to purchase. A friend of yours told you about the new chatbot from Eventim that can help you to find a fitting gift for your friend. The video you will see soon is a recorded conversation with the Eventim chatbot. Imagine being the person who interacts with the chatbot.

Assigned video

One of the six conditions were assigned (Can be found in Appendix B)

Please watch this video carefully.

Additional question block for manipulation

The chatbot picture presents a real person. Yes/No

The chatbot picture presents an animated person. Yes/No

The chatbot picture presents a logo. Yes/No

Chatbot appearance

Strongly agree - Strongly disagree

The Impression of the chatbots picture felt alive	1	2	3	4	5
The impression of the chatbot's picture felt lively.	1	2	3	4	5
The impression of the chatbot's picture felt natural.	1	2	3	4	5
The impression of the chatbot's picture felt interactive.	1	2	3	4	5
My impression of the chatbot's picture felt natural.	1	2	3	4	5
My impression of the chatbot's picture felt humanlike.	1	2	3	4	5
My impression of the chatbot's picture felt lifelike	1	2	3	4	5

Chatbot language

Strongly agree - Strongly disagree

The Impression of the chatbots language felt alive	1	2	3	4	5
The impression of the chatbot's language felt lively.	1	2	3	4	5
The impression of the chatbot's language felt natural.	1	2	3	4	5
The impression of the chatbot's language felt interactive.	1	2	3	4	5
My impression of the chatbot's language felt natural.	1	2	3	4	5
My impression of the chatbot's language felt humanlike.	1	2	3	4	5
My impression of the chatbot's language felt lifelike	1	2	3	4	5

Trust

Strongly agree - Strongly disagree

The chatbot was reliable	1	2	3	4	5
The chatbot was dependable.	1	2	3	4	5
The chatbot was competent.	1	2	3	4	5
The chatbot was able.	1	2	3	4	5

Satisfaction

Strongly agree - Strongly disagree

I am satisfied with the chatbot interface.	1	2	3	4	5
My choice to ask the chatbot questions for a gift					

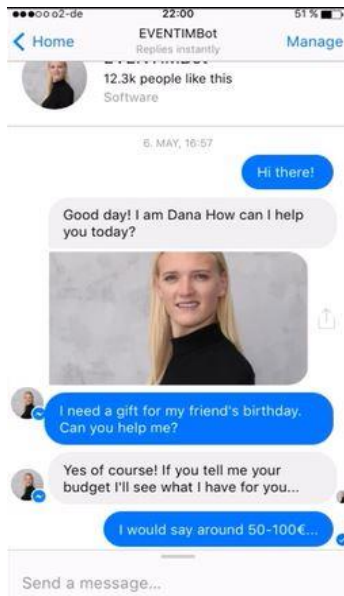
was a wise one.	1	2	3	4	5
I am satisfied with the way the chatbot helped me.	1	2	3	4	5

Purchase Intention

Extremely likely- Extremely unlikely

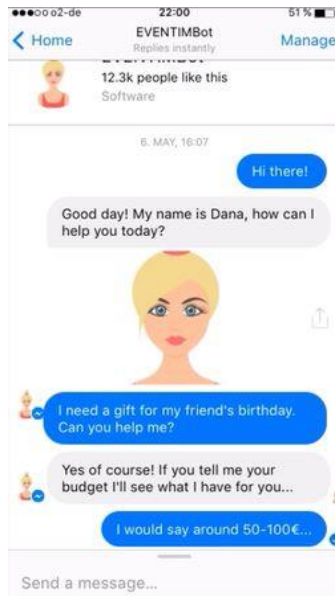
I am very likely to buy a ticket from Eventim.	1	2	3	4	5
I intend to buy a ticket within 3 months from Eventim.	1	2	3	4	5
I intend to buy a ticket within 6 months from Eventim.	1	2	3	4	5

Appendix B – Overview of Chatbot Videos



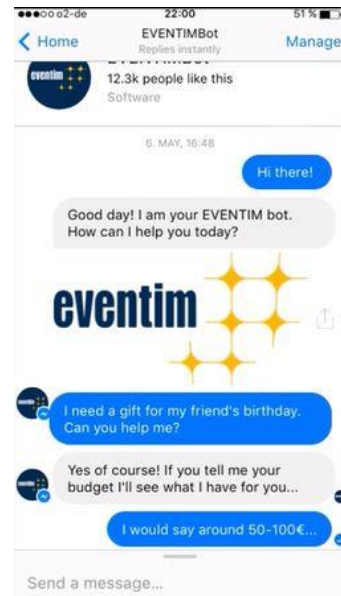
1) Human - Human language
Youtube Link:

<https://www.youtube.com/watch?v=6k9TMp73Nh8>



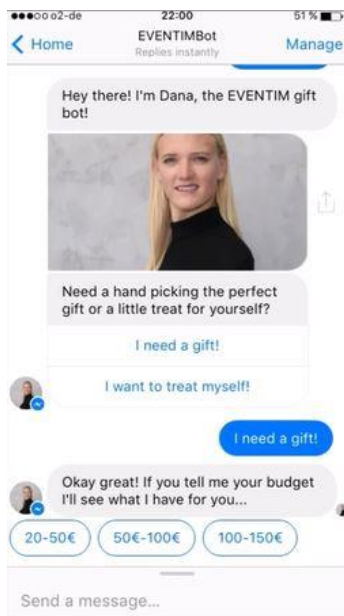
2) Animation – Human language
Youtube Link:

<https://www.youtube.com/watch?v=Akv2YK1RGSM>



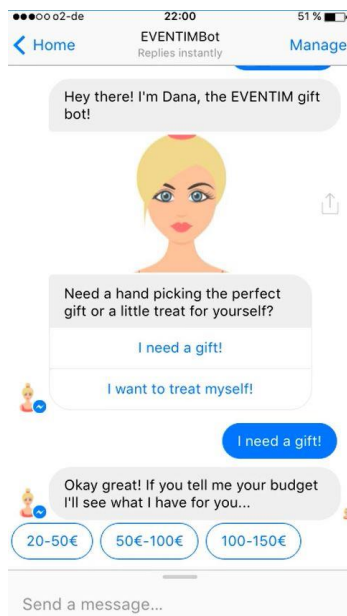
3) Logo – Human language
Youtube Link:

<https://www.youtube.com/watch?v=JWU1vJpgC4I>



4) Human – Robotic
Language Youtube Link:

<https://www.youtube.com/watch?v=vYtyAPlvLU>



5) Animation – Robotic
Language Youtube Link:

https://www.youtube.com/watch?v=qWBwFk005_o



6) Logo – Robotic
Language Youtube Link:

<https://www.youtube.com/watch?v=Vgrp5Ag7oZA>

Appendix C – Literature Study Log

Research questions literature study

Formulate research questions regarding your literature study. These question(s) can differ from the research question(s) that are used in your research proposal

Does chatbot appearance influence trust, satisfaction and purchase intention?

Does chatbot language influence trust, satisfaction and purchase intention?

Are the effects of a chatbot appearance on trust, customer satisfaction, and purchase intention dependent on the robotic/human language used for the interaction?

Does trust mediate the effects of chatbot appearance and language on trust, satisfaction and purchase intention?

Concepts in research questions (most important terms in the research questions)

- Chatbot appearance
- Chatbot design
- Chatbot development
- Social presence
- Social presence chatbots
- Anthropomorphic visual cues
- Natural Language
- Tone of Voice chatbots
- Anthropomorphic language cues
- Conversational UI
- Trust eCommerce
- Trust mCommerce
- Trust online
- Customer satisfaction
- Online satisfaction
- User satisfaction
- User experience
- Purchase intention
- Purchase intention mobile commerce
- Purchase intention eCommerce

Criteria preferred materials

Formulate the criteria to use in selecting materials.

Since the topic of chatbots and their appearance and language are still in its infancies, materials are quite limited. However, many chatbot blogs are online which were used for non-scientific sources. However, most literature was chosen from scientific databases, such as scopus or google scholar.

Selected Databases

Discuss why these are the most appropriate databases to use.

Scopus and Google Scholar were used the most for this research because it is easy to search across many disciplines and sources: articles, theses, books, abstracts and court opinions, from academic publishers, professional societies, online repositories, universities and other web sites. Google Scholar helps me to find relevant work across the world of scholarly research

Relevant terms**Concepts Related terms Smaller terms Broader terms**

Concepts	Related terms	Smaller terms	Broader terms
Concept 1: chatbot appearance	<ul style="list-style-type: none">- Chatbot design features- Social presence- Social presence chatbots- Anthropomorphic visual cues	Chatbot design Chatbot development	Artificial intelligence
Concept 2: chatbot language	<ul style="list-style-type: none">- Natural language- Anthropomorphic language cues- Tone of voice- Conversational UI	Natural language	Conversational Interface
Concept 3: trust	<ul style="list-style-type: none">- Online trust- Trust in ecommerce- Trust in mcommerce- Chatbot trust- Anthropomorphic- Uncanny valley	Safety Reliance Belief Vulnerability	Confidence
Concept 4: satisfaction	<ul style="list-style-type: none">- Customer satisfaction- User satisfaction- User experience- Satisfaction in chatbots- Uncanny valley	User Experience	Performance
Concept 5: purchase intention	<ul style="list-style-type: none">- Purchase intention- Purchase intention mobile commerce- Purchase intention eCommerce- Theory of planned behaviour	Money Buying Intention	Sales

Search actions

	Date	Database	Search action	Total Hits
1	13.03.2019	Google Scholar	Chatbot appearance	1960
2	21.04.2019	Google Scholar	Anthropomorphic design cues	19800
3	14.03.2019	Google Scholar	Social presence chatbot	3090
4	13.03.2019	Google Scholar	Chatbot natural language	7430
5	21.04.2019	Google Scholar	Anthropomorphic linguistic cues	11100
6	14.03.2019	Google Scholar	mobile commerce trust	638000
7	14.03.2019	Google Scholar	Chatbot trust	2700
8	14.03.2019	Google Scholar	Chatbot user experience	5790
9	14.03.2019	Google Scholar	Chatbot satisfaction	2390
10	14.03.2019	Google Scholar	Purchase Intention	1919000
11	14.03.2019	Google Scholar	Chatbot purchase intention	2180

Found references in APA style

- Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, 85, 183-189.
- Chung, H., Iorga, M., Voas, J., & Lee, S. (2017). Alexa, can I trust you?. *Computer*, 50(9), 100-104.
- 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.
- Ghose, S., & Barua, J. J. (2013, May). Toward the implementation of a topic specific dialogue based natural language chatbot as an undergraduate advisor. In *2013 International Conference on Informatics, Electronics and Vision (ICIEV)* (pp. 1-5). IEEE.
- Robinson, R., & Smith, C. (2002). Psychosocial and demographic variables associated with consumer intention to purchase sustainably produced foods as defined by the Midwest Food Alliance. *Journal of nutrition education and behavior*, 34(6), 316-325.
- Sah, Y. J., & Peng, W. (2015). Effects of visual and linguistic anthropomorphic cues on social perception, self-awareness, and information disclosure in a health website. *Computers in Human Behavior*, 45, 392-401.
- Shawar, B. A., & Atwell, E. (2007, April). Different measurements metrics to evaluate a chatbot system. In *Proceedings of the workshop on bridging the gap: Academic and industrial research in dialog technologies* (pp. 89-96). Association for Computational Linguistics.
- Siau, K., & Shen, Z. (2003). Building customer trust in mobile commerce. *Communications of the ACM*, 46(4), 91-94.
- Tatai, G., Csordás, A., Kiss, Á., Szaló, A., & Laufer, L. (2003, September). Happy chatbot, happy user. In *International Workshop on Intelligent Virtual Agents* (pp. 5-12). Springer, Berlin, Heidelberg.
- Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017, May). A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 3506-3510). ACM.

Reflection

Reflect on the following issues:

What important choices have you made in your search process to get to the qualitative good information?

Since chatbots are a very new topic, it was sometimes difficult to find the right empirical literature to find information I was looking for. However, it was easier to see vital papers through the citations or references to sources from chatbot magazines or blogs.

How did you orientate yourself on the subject?

I researched about the chatbot topic and obtained a lot information and inspiration from my supervisor Dr. A. Beldad. Moreover, I watched tedtalks for inspiration next to literature research. Further, I became part of LinkedIn chatbot groups to obtain more information form professionals in this field

Which (combinations of) terms were important.

One of the most important terms was to implement chatbot before one of the dependent variables. Moreover, Anthropomorphic seemed to be a useful term for my literature search.

Which databases?

Mostly Google Scholar and Scopus.

Which searching techniques?

Trying out combinations of word terms until I found fitting literature.

To what extent did they deviate from the terms which you have initially used?

To a moderate extent. Chatbots are barely researched, hence the literature was quite limited and did not lead to huge changes in search results

How did you assess the relevance and quality of found articles, books or other materials?

Mostly cited and by relevant journals

What would you do differently in a next search operation?

Not too much. Maybe look into more different literature databases.