NTNU
Norwegian University of
Science and Technology

Personality Matters! Improving The User Experience of Chatbot Interfaces

Personality provides a stable pattern to guide the design and behaviour of conversational agents

Tuva Lunde Smestad

Master in Interaction Design
Submission date: June 2018
Supervisor: Frode Volden, ID
Co-supervisor: Anders-Petter Andersson, ID

Norwegian University of Science and Technology
Department of Design

Preface

This master thesis is the final part of my Master in Interaction Design degree at the department of Design at the Norwegian University of Science and Technology (NTNU). The project planning, preliminary studies and literature review was conducted during the autumn of 2017. The work presented in this thesis was conducted and written during the spring of 2018 and the workload corresponds to 30 ECTS.

This project was undertaken after I was challenged with building my first chatbot, and was surprised to find that much work is still to be done regarding how we can improve the user experience of conversational interfaces. After having tried, failed, improved, and completed the design process of this chatbot, I was left with a lot of experiences and insights that I wished to investigate further. And having read the recent reviews and overall reception of chatbots, I was concerned whether this could lead to abandonment of chatbots as service. I therefore wanted to use my final year of my masters, to try and prove at least one assumption regarding conversational agents: can personality improve the user experience of chatbots?

I aim with this master thesis to add to the field of human-computer and human-robot interaction by providing evidence to explain how personality impacts the user experience of conversational interfaces, and how personality affects how humans perceive chatbots. In addition to this I also wanted to create a framework that can be adopted by others to build their own user-centred chatbot personalities. This framework was implemented to answer whether personality matters to the user experience of conversational agents.

01-06-2018

Acknowledgment

I would like to thank the following persons for their great help during the course of this thesis project:

My supervisor Frode Volden, for your excellent guidance and support during this process. Thank you for keeping my research aim on track, reminding me to not forget about the overall aim of the thesis, and not become too concerned about the design and performance of the chatbot prototype. I am grateful for all your help and your sound knowledge in SPSS, which I would not have made it without!

My co-supervisor Anders-Petter Andersson for providing challenging questions that made me stop to reassess and question my choices throughout the process. Pointing me in the direction of valuable resources and articles, designs and research to help me get a thorough understanding of the field I now have explored.

My partner Christopher Solem for providing emotional support, extraordinary listening skills, helpful commentary and for proof reading not only my thesis, but the literature review and project planning report. I would not have survived my masters degree without you.

Mom and dad for your long standing support, both emotional and material, and and extra thank you for lending me your car so I could get around to all who participated in my experiment.

The participants for taking time out of their busy schedule to help me with my thesis project.

T.L.S.

Abstract

Recent advances in machine learning has contributed to the rebirth of the chatbot. Lately we have seen a rise in chatbot technology being made available on the web and on our mobile devices, and recent reports states that 57 % of companies have implemented or are planning to implement a chatbot in the near future. Chatbots are therefore a big part of an AI powered future, however recent reviews find chatbots to be perceived as unintelligent and non-conversational. Such findings have not slowed down the rapid implementation of chatbots online, and the same mistakes seems to be repeated over and over again. Chatbot services have been found to save companies an estimate of $8 billion by 2022, and extends to customer service tasks, product purchasing, shopping assistants, recommender systems, service or product support. This explains why so many are eager to implement their own chatbots, but the reviews make one wonder whether we now are forcing users to adopt technology which they find frustrating and useless. Chatbots are becoming an extension of the services companies provide, therefore ensuring a great user experience is important not only for a company's brand image, but also for the users of their services. Existing literature regarding how humans perceive conversational agents have found that personality can offer a stable pattern to how the chatbot is perceived, and add consistency to the user experience. This master thesis project investigated how we could improve the user experience of chatbots through personality. The thesis is twofold, the first part of the thesis consists of a detailed description of the personality framework developed and implemented to build a chatbot prototype. The framework combines techniques from user-centred design, branding, and personality theory, to build user-centred chatbots through a design process with a basis in personality. The thesis also offers a method to test and evaluate the modelled personality in regards to whether it is perceived as intended by the designer. The second part of the thesis consists of an experiment to investigate whether personality has an improved effect on the user experience of chatbot interfaces. The experiment found that the chatbot personality built using the personality framework had a significant improved effect on the user experience.

Contents

List of Figures

List of Tables

Acronyms

UCD User Centred Design
UX User Experience
CUI Conversational User Interface
VUI Voice User Interface
NUI Natural User Interface
AI Artificial Intelligence
NLP Natural Language Processing
API Application Programming Interface
UI User Interface
CA Conversational Agent
VA Virtual Agent
ECA Embodied Conversational Agent
SDS Spoken Dialogue System
PM Project Manager
OCEAN Openness, Conscientiousness, Extroversion, Agreeableness, Neuroticism
HEXACO Honesty-humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness, Openness
TOV Tone of Voice
PQ Pragmatic Quality
HQ-I Hedonic Quality - Identity
HQ-S Hedonic Quality - Stimulation
ATT Attractiveness
IoT Internet of Things

1 Introduction

The chatbots of 2018 have not truly improved since the first chatbot ELIZA created back in 1966. While recent advances in machine learning has contributed to fast improvements in Artificial Intelligence (AI) and Natural Language Processing (NLP) of conversational interfaces, chatbots are still not perceived as intelligent conversational actors. There has been an explosion of available chatbot agents online the past year, and they are now representing companies by directly providing services to their customers. Ever since ELIZA, the goal of chatbot systems have been to pass the Turing Test, and convince humans that they are conversing with a human, not a machine (McTear et al. 2016*a*). However, available chatbots does not appear to possess human conversational skills, rather they perform as machines that responds to user commands. The recent advances has contributed to chatbot systems becoming more flexible in regards to it being able to understand vast variations of the same command, and the implementation of cloud-based systems has allowed for an explosive growth of devices connected through the internet, also known as the Internet of Things (IoT). Therefore access to AI has become widespread, and through application programming interfaces (APIs), chatbots have access to vast amounts of information and knowledge through thousands of databases online. All this sounds promising, and explains in large part why chatbots have seen a rebirth recently, but all this does not matter if chatbots cannot live to the expectations of users. Predictions find chatbots to be a big part of an AI powered future, but recent reviews have found them to be unintelligent and non-conversational (Stokke 2017, Orf 2017, Piltch 2017, Vincent 2017, Boutin 2017). Piltch (2017) states that we should not be carried away by the positive outlook researchers presents in regards to the possibilities of advances in AI for chatbot technology, as the reality is that most chatbots are falling flat. The fast implementation of chatbots have resulted in flawed interfaces that fails to predict the simplest of questions. Despite cautions and recent negative reviews, Forrester (2017) found that 57 % of companies have implemented or are planning to implement a chatbot as part of the services they provide in the near future. JuniperResearch (2017) released a report in which they have found that chatbots will save companies $8 billion in costs by 2022. Therefore, as the trends predict many benefits for companies implementing chatbots, are we forgetting to assess whether the same systems are beneficial for its users? It would appear that we are forcing users to adopt technology which they find frustrating and useless. If the reviews find chatbot interactions as unintelligent, pointless, and not more effective than conducting a Google search or contacting a human customer service agent, what effects might this have on the user experience and the future of conversational interfaces? This thesis project explores how we can build chatbots that offer a better user experience and are perceived as more intelligent social actors through focusing on the design of chatbot personalities. The thesis will use personality as a the stable pattern to guide the design process of a chatbot prototype, and use this personality

to evaluate its effects on the user experience. This personality framework will be based on a deep understanding of users and their needs, the brand it represents, and personality theory to build an appropriate chatbot personality.

1.1 Justification, Motivation and Benefits

There has always been a pattern of new technologies failing to be adopted by users, because they have been released before we know how to ensure that they create more value to its users than existing solutions. Chatbots promises a lot, but fail to live up to its promises. Understanding the ways in which we can improve the user experience of chatbots can help turn this around, as we can meet user expectations and create value for users, even though there is still much to be done before chatbots truly can fulfil their potential. The one thing we can improve is how they are perceived. Understanding why chatbots are failing to be perceived as more than a computer, can have a great impact on improving the user experience. AI and NLP helps chatbots understand language, but we must take into account many other factors of human interaction beyond language understanding only, if we want to improve how they are perceived by human users. Humans have bodies, feelings, emotions, different personalities and behaviours that all influence how we communicate, behave, and interact. Chatbots needs to be able to simulate these skills for it to be perceived as more than a computer; they have to become skilled social actors. If chatbots are to become this, we must understand the social cues that make up human interaction, and the personalities that drives them. As trends show that companies are rapidly implementing chatbots as part of their services, consumers should not be forced to adopt solutions that does not improve the effectiveness, efficiency or satisfaction compared to traditional systems.

This Master's thesis aims to benefit designers of chatbots, brands that wishes to add chatbots to their services, and the users of chatbot systems. The thesis will provide this by developing, implementing, and evaluating a personality framework to help design user-centred chatbot personalities. The aim of this framework is to improve the user experience of chatbot interfaces by focusing on the personality of the chatbot, ensuring that it meets users' needs and is consistent with the brand image it represents.

1.2 Research Questions

The research questions and sub-questions to be addressed in the master thesis project are:

1. How can we design chatbot personalities to guide the design process of a chatbot interface?
    a) Can personality be used as a stable pattern to guide the design process of chatbot interfaces?
    b) Which elements must be considered to inform a chatbot personality?
    c) What components needs to be in place for a chatbot personality to meet user needs and expectations?

2. Will chatbots with a defined personality improve the user experience of chatbot interfaces?

1.3 Planned Contributions

Through this thesis the researcher hopes to contribute to the understanding of how designers can create improved chatbot interfaces, and investigate whether personality has an improved effect on the user experience. This investigation is twofold, as the first part involves developing and implementing a personality framework to build user centred chatbot personalities. This will be the first contribution that will allow other designers, developers, or researchers to use the framework and improve it. The second part of the thesis uses the modelled personality to understand whether the personality is perceived as intended, and investigate the effects personality has on the user experience of chatbots. The thesis offers insights into how thinking early about personality provides a stable pattern to the design process of chatbot interfaces. The experiment contributes to the knowledge of the effects personality has on the user experience, and also offers two methods to evaluate the personality. The first evaluation method helps determine whether users perceive the personality as intended, while the other method evaluates the personality in regards to the user experience. Both evaluation methods collects quantitative data to evaluate the chatbot personality.

1.4 Thesis Outline

The thesis consists of 6 chapters.

Chapter 1 introduces the problem being addressed by the thesis, the justification, motivation and benefits of writing this thesis as well as the planned contributions this thesis offers to the research community and beyond.

Chapter 2 provides the reader with the theory, background and existing literature regarding chatbots and conversational agents, a summary of a preliminary literature review that reviews factors that affects how humans perceive conversational agents, and personality theory to inform how to build chatbot personalities.

Chapter 3 describes the methodology. This is twofold, as the first part consists of the personality framework that was used to build the chatbot prototype, while the second part includes the experiment methodology used to evaluate the modelled personality and investigate the effects the personality has on the user experience.

Chapter 4 presents the results from the statistical analysis conducted to interpret the results from the evaluation of the personality and the results from the user experience evaluation.

Chapter 5 includes the discussion of the implementation of the personality framework, and a discussion and interpretation of the results from the experiments assessing the personality and user experience. Possible confounders, contributions and limitations are also discussed in this chapter.

Chapter 6 offers a summary and the final conclusions taken from the implementation of the framework and evaluation of the personality. The research questions, and hypotheses, will be answered in this chapter. Recommendations for future research is also offered.

2 Theory, Background, Existing Literature

To understand the scope of research related to the interaction between chatbots and humans, and to understand what important factors to consider to design chatbot personalities, the researcher conducted a literature review as part of the course *IMT4215 Specialisation Project*. This review investigated the scope of research as it related to how humans perceive computers that talk, by looking at research related to the effect anthropomorphism, humanness, and personality have on the user experience of conversational interfaces. The review was concerned with theories in human-computer interaction research, while also exploring human-robot interaction, and theories within human factors and psychology. Through this review the researcher found that emotional intelligence, anthropomorphism, humanness, politeness, human etiquette, humour and gender are important variables to consider when designing a CA, as these are important to how users perceive and experience a CA. In addition, the review also found that researchers (Callejas et al. 2011, Xiao et al. 2005, McTear et al. 2016*b*) believe that personality can be used as a stable pattern to form the behaviour and characteristics of a CA and manage how humans perceive the system. Findings from this review will be summarised and discussed in this chapter in addition to the theory and background regarding chatbots, and how personality theory and insights from branding can inform the design of chatbot personalities.

2.1 Chatbots and Conversational Agents

Chatbots are considered to be a form of "weak AI" (De Angeli & Brahnam 2008), this means that they do not exceed human intelligence and are often used to complete tasks and for analysing and processing information. Chatbots follow scripted rules, and respond from a set of stored, pre-defined responses, with the goal of simulating human language and conversation. This approach, called a "stimulus-response approach" (McTear et al. 2016*a*, : 57), was first realised with the ELIZA chatbot back in 1966. ELIZA was a conversational agent, written by Weizenbaum (1966), that played the role as a Rogerian psychotherapist very convincingly. The stimulus-response approach functioned by ELIZA matching the user input against a large set of stored patterns, which informed its response output. This means that ELIZA makes a prediction in regards to which of its responses matches to the user's query. This approach proved to be very successful for ELIZA's role as a psychotherapist, and made it appear that she was convincingly conversing with users. The same approach is still used today, and there are two different models: retrieval-based models and generative models (Kothari et al. 2016). The first model uses predefined responses, while the generative model can put together new responses based on users' input. The second model is of course much more difficult to implement, while the first model provides control of which responses the chatbot gives and is much

quicker and easier to build and implement. Most of the chatbots we see today that are provided by brands, are based on the retrieval-based models. While being more in control of the output of the chatbots in the retrieval-based model, the downside is that they often appear as less intelligent and are prone to errors if they fail to recognise the correct user intention. ELIZA was successful when she conversed with users, because her role as a psychotherapist allowed her to act like she knew nothing of the world (Weizenbaum 1966). The way in which ELIZA conversed, "tell me more about x", appeared familiar and consistent to how therapy sessions are usually run, and therefore supported users' mental model and expectations when conversing with a therapist. ELIZA inspired future generations of chatbots to simulate human natural language, and ultimately beat Alan Turing (1950) "Imitation Game": to convince humans that they are conversing with a human, not a machine. Therefore, the history and development of chatbots have always had this goal in mind. Today we see chatbots in education, customer service, e-commerce, as virtual assistants or plainly for entertainment as recent advances in machine learning allows for more advanced AI and NLP capabilities for chatbots.

Michael Mauldin (1994) coined the term Chatterbots, which today is known as chatbots, to describe the robots that humans could chat with. Chatbots are conversational agents within the broader term conversational user interfaces (CUI). There are many terms used to describe CUIs such as natural user interfaces (NUI), voice user interfaces (VUI), no user interface or invisible user interface. CUI and NUI are often used interchangeably, however a NUI often allow for other natural inputs beyond conversation such as touch and gestures. While chatbots refers to CUIs in which users interact with through a chat interface, a VUI only interacts through voice input and output (e.g. Apple's Siri) (Pearl 2017). Other types of conversational interfaces or conversational agents includes spoken dialogue systems (SDS), embodied conversational agents (ECAs) and social robots (McTear et al. 2016*a*). ECAs make use of facial expressions, animated bodies, and other gestures as well as speech to engage with users. An ECA is therefore a form of a chatbot that make use of more elements than speech to add to the conversation. With the emergence of speech recognition technology and mobile devices, it is common that chatbots also support voice input-output to allow for a more accessible interface.

Conversational agents therefore come in a variety of different forms and purposes, but they all use conversation to interact with humans. Chatbots were originally designed to be conversational partners, rather than a system to help users perform tasks (McTear et al. 2016*b*). Today this definition has changed, as the trends show an increasing rise in the popularity of chatbots as virtual assistants (VA) on the web. Companies are rapidly implementing chatbots as an extension of the services they provide to their customers, moving beyond its role as a conversational partner. As stated earlier, a report published by Forrester (2017) found that 57 % of companies either use or plan to implement a chatbot in the near future. The conversational element is used to offer a natural way to interact with brands to retrieve information, service support, product purchasing, or other uses. This trend can be explained by the commercial access to APIs which provides easy access to AI and NLP. In addition to this there is a rise in messaging platforms that supports the hosting of chatbots such as Facebook Messenger, Slack, and Skype among others. Humans have also been

introduced to CAs more in their everyday life, as most smartphones offer voice assistants to help manage tasks (e.g. Apple's Siri, Google Assistant, Microsoft Cortana).

This rapid trend has had its downside, as too many of these chatbots perform poorly, and their conversational skills do not meet user expectations (Stokke 2017, Boutin 2017). Recently the Norwegian website DinSide.no reviewed Norwegian customer service chatbots implemented in 2017 and the headline reads: "So Stupid Are the Norwegian Robots" Stokke (2017). The article concludes that Google is more efficient in answering your questions, and more accurate, than the customer service chatbots. The reason for this conclusion is that the user needs to adapt its language for the chatbot to understand their query, which can be a time-consuming process, and the conversational element is missing. Apparently, access to AI and NLP is not what makes or breaks a chatbot, and designers must explore other approaches and techniques to design better conversational interfaces.

2.2 Personality to Dictate Human Perception

Through the literature review, the researcher investigated how humans perceive conversational agents and which factors that can be used to influence human perception. The findings suggests that personality is an important factor in relation to alter and dictate the way in which humans perceive CAs. The following sections will provide a summary of the most important findings from the literature review and why personality is so important to the design of CAs.

2.2.1 Emotions and personality

Emotional intelligence is an important part of how humans perceive themselves as intelligent beings. In order to assess artificial intelligence critics have always used emotional intelligence to define something as a sentient being. Psychologists describe emotional intelligence as the ability to tailor behaviour to environment through necessary emotional processing (Callejas et al. 2011). This ability is crucial to conversation, as conversation happen through dynamic relationships between the conversational actors. Therefore, to understand how designers can improve the conversation of chatbots, we must look at these elements of human interaction. The literature review revealed that emotional intelligence is important for humans to perceive CAs as thinking beings as it is part of natural social interactions (Griol et al. 2015, 2017, Balzarotti et al. 2014, Lemon et al. 2012, Mencía et al. 2012, McTear et al. 2016*b*). Research into conversational interfaces and emotion have mostly focused on embodied conversational agents (ECA) (Lester et al. 1997, Stern 2003, Beun et al. 2003, Reeves & Nass 1996) as the focus has been on the appearance of robots and its emotional responses through body and facial gestures, and its ability to read the "mood" of its conversational partner. Chatbots can make use of the same findings in their design, as these can be used to determine the extent to which a user anthropomorphise (more in section 2.2.2) a CA.

In human interaction we make use of several social cues that dictates how we behave and how we are perceived by our conversational partners. Fogg (2002) propose that there are five primary social cues:

1. **Physical:** face, eyes, body, movement
2. **Psychological:** preferences, humour, feelings, empathy
3. **Language:** interactive language use, spoken language, language recognition
4. **Social dynamics:** turn-taking, cooperation, praise, question answering, reciprocity
5. **Social Roles:** doctor, teammate, opponent, teacher, pet, guide

Our social interactions are dynamic, in which we mirror and change our behaviour to our conversational partners. Our social role is another important factor that influence how we behave in different situations; we act differently if we take on the role as a parent than we would as a friend. One of the driving forces behind how humans behave as social actors is personality. Our personality can be used to influence our environment, emotions and cognitions as well as our motivations. Callejas et al. (2011) listed evidence from empirical research, the psychology of emotional intelligence, and the principle of similarity attraction to explain how personality impacts users perception and willingness to interact with CAs. Stern (2003) found that children interacting with emotional agents or virtual characters forms emotional relationships. There are therefore empirical evidence to support that emotional intelligent agents are more likely to form emotional relationships between the human and the virtual character/agent.

Callejas et al. (2011) and Xiao et al. (2005) believes that personality is the stable pattern that dictates the behaviour of a CA. Personality is defined as a "dynamic and organized set of characteristics possessed by a person that uniquely influences their environment, cognitions, emotions, motivations, and behaviors in various situations" (McTear et al. 2016*a*). Research have found that personality, the characteristics that dictates our behaviour, plays an important part in regards to how users perceive conversational interfaces, and can be the determining factor to whether users wish to interact with the agent again (Callejas et al. 2011). Norman (2007) wrote in his book *Emotional Design* that emotional expressions and the personality of things would increase user satisfaction and inform users of what the system is capable of. When designing a conversational agent, the personality can be used to allow for a consistent interaction with the system. As Pavlus (2016) states: "in conversational UIs, personality is the new UX". The personality provides users with a consistent interaction, as inconsistent personalities can cause users to feel that they are talking to different "people" in one interaction. When Microsoft's Virtual Assistant Cortana was released with Windows 10, she was described as: "like Siri with a human personality" (Beres 2015). The PM of Cortana, Susan Hendrich, explained that they had interviewed several celebrities' personal assistants in order to design the right personality users would expect from a personal assistant. The personality of Cortana gave her an edge, that differentiated it from similar available solutions, and by basing her personality on real personal assistants, the design team were able to understand the success criteria of personal assistants (Hendrich 2017). As chatbots are scripted systems and personality can be used to dictate the behaviour of the CA, personality can also help guide the design process of chatbots and help write the conversation flow and plan for different user scenarios.

2.2.2 Anthropomorphism

Anthropomorphism is defined as "the attribution of human personality or characteristics to something non-human, as an animal, object, etc" (Dictionary n.d.). Anthropomorphism is therefore human's ability to attribute human motivations, beliefs, and feelings to non-human entities. Researchers have found that anthropomorphism is a normal occurrence in human-computer interaction (Reeves & Nass 1996, Cohen et al. 2004, Pearl 2017, Lee 2010), and that personality can be used as a design variable to manage how users anthropomorphise computers (Xiao et al. 2005). According to Schroeder & Epley (2016) the "humanlike mind" is an essential component of anthropomorphism, as humans needs to consider the machine as a thinking being to some extent in order for them to perceive the CA as having a mind of its own. While the conversational element is the main element chatbots can make use of in order to simulate a humanlike mind, it is therefore very important that the conversation appear natural and adhere to the rules of social conduct. For ECAs the use of facial gestures, body movements, expressions are important elements in regards to how humans anthropomorphise the agent. Researchers have also found that levels of humanness also affect how humans anthropomorphise, as well as being an important factor for managing trust (Prada et al. 2003, Meyer et al. 2016, Dautenhahn et al. 2002, Terada et al. 2015, Epley et al. 2007, Lee & See 2004).

When a human anthropomorphises a computer system or other entity, the humanlike characteristics they attribute to the system is determined by how they perceive the system. Therefore, designers can control how humans attribute characteristics to the CA by designing a personality and use this to guide how it behaves, reacts and how it responds. Through a preliminary study, conducted for the course *IMT4898 Specialisation in Interaction Design*, the researcher found that users anthropomorphised the agent consistently with the predefined personality. The participants were asked to describe the chatbot after having conversed with it, and the words used to describe it matched the predefined personality in which the system was based upon. In this study, the researcher presented participants with the same chatbot personality, however half the participants were presented with a version which had high levels of humanness and the other half with a version with low levels of humanness. The findings from this preliminary study are consistent with findings from other similar experiments where participants rated chatbots according to the personality traits ascribed to it (Holtgraves et al. 2007). However, although users perceived the personality as consistent independent of the levels of humanness, the users did not perceive the two agents equally. The agent with high levels of humanness guided users to engage in natural conversation, while the bot with low levels required more prompts from the moderator to help users interact with it.

2.2.3 Humanness

Humanness is defined as "the extent to which an agent is designed to act and appear human [...] encompassing the objectively established human capabilities (having eyes, a face or the ability to respond politely)" (Meyer et al. 2016). Therefore, researchers distinguish anthropomorphism from humanness as anthropomorphism relates to the psychological attribution of humanlike features

(Epley et al. 2007, Mori 1970, Nass & Moon 2000). In simpler terms, humanness refers to the extent the agent looks human through incorporating human appearance and capabilities, while anthropomorphism can be attributed to entities that does not resemble humans in its presentation. This distinction is important, because while anthropomorphism is encouraged, different levels of humanness can have both negative and positive effects on how humans perceives the agent. Mori (1970) coined the term "the uncanny valley" when describing the effects high levels of humanness can have. He found that robots that resembles humans to a very high degree are perceived as creepy, and humans interacting with them feel uncomfortable or fearful of it. Therefore, designers must consider the level of humanness of the agent to not evoke negative emotions. However, although too much humanness can have a negative effect, higher levels of humanness have been found to increase trust.

Visser (in Meyer et al. 2016) states that the degree of humanness should be decided based on the objective. He explained that increased humanness is recommended when the objective is to increase trust, e.g. in systems where errors in automation are more likely to occur. While in systems that deals with situations where users are vulnerable, should have decreased humanness in order to appear more logical, consistent, and fair: without emotion or human judgement (Visser, in Meyer et al, 2016: 281). Terada et al. (2015) found that high levels of humanness had a positive effect on people's buying motivations when CAs were used to recommend products. This they stated might be due to increase in familiarity, as a human form is more familiar in a buying situation, but also because they appeared to be of higher intelligence than agents with low levels of humanness. Disalvo et al. (2002) states that it is important to maintain levels of "robot-ness" to make sure users do not develop false expectations in regards to the capabilities of the agent. If the agent then does not appear to be human when interacting with it, it only looks human, this can cause frustrations and a lack of trust in the system. Examples of CAs with low levels of humanness are Apple's Siri, Microsoft Cortana or Amazon's Alexa, and this is to make it completely clear that these are not human agents and does therefore not possess human characteristics. By keeping this distinction clear humans will treat these CAs as machines rather than humans, and this will manage their expectations towards these systems.

Therefore, while anthropomorphism is encouraged in order to build an emotional relationship between the human and the CA, humanness can be used to determine the extent to which we want humans to anthropomorphise the system. In addition to humanness as a variable, building a consistent personality for the chatbot has been proposed to help manage how the agent is anthropomorphised.

2.2.4 Other factors for how humans perceive CAs

Politeness

Meyer et al. (2016) proposed in their article that human politeness and etiquette can be used as a variable to design a chatbot's behaviour. Researchers have found that humans perceive polite agents

more positively than those who were less polite or machine like (Inbar & Meyer 2015, Holtgraves et al. 2007). Polite behaviour also provide consistency and meets user's expectations, as different social roles also dictates behaviour in terms of expected politeness, and this can be enough to achieve desired perceptions. While appearing human can be desired by the designers of chatbots, it is important that human users of the chatbot are aware of what or who they are conversing with. Therefore, if the context in which the chatbot interaction occurs makes it important that users are very aware that they are conversing with a machine; politeness and human etiquette can be used instead of humanness to offer a positive interaction. It is also important to consider social and cultural differences regarding rules of conduct.

Humour

Humour has been found to also have a positive effect on how humans perceive chatbots. Humour is an important part of everyday social human interaction (Dirk 2003) and can be used to foster engagement. In computer systems in which tasks might be long and boring, humour can be used to maintain long-term interactions and alleviate boredom (McTear et al. 2016*a*). In addition to this, a chatbot that is humorous might encourage more positive involvement, and increase whether humans perceive it as being emotionally intelligent (Dybala et al. 2009). Humour is therefore a great way to add emotion to a conversation, as chatbots can be trained through computational humour (Augello et al. 2011) to recognise humour expressions, user's mood and emotions, and display appropriate emotions and humour responses in return. Through the preliminary study conducted by the researcher for *IMT4898 Specialisation in Interaction Design*, participants found the use of emojis to be a great way to add humour and emotions to the conversation. They also stated that it helped communicate the chatbots personality in regards to which emojis it used, and how frequently. It is therefore also important to assess the target audience and context to understand whether the use of emojis, and/or jokes, is appropriate.

Gender

Which gender to assign to chatbots are problematic in several ways. Through research on CAs and gender, researchers have found that gender have a huge effect on how humans perceive a CA (Zimmerman et al. 2005, Brahnam & De Angeli 2012, Vala et al. 2011, Kulms et al. 2011). Female CAs were more likely to be attributed negative stereotypes, and received implicit and explicit sexual attention and swear words (Brahnam & De Angeli 2012). Female ECAs often receive more talk regarding their appearance. The same study found that disembodied female CAs received more attention regarding their appearance than male disembodied CAs, but less than ECAs. In all robotic or androgynous CAs receive much less negative, sexual or profound language than gendered CAs. It is therefore important that those who implement the chatbot also consider whether they should support or break with gender stereotypes, and how the chatbot should handle sexual attention or profound language.

2.3 Personality theory

In order to design a chatbot personality that keeps in line with the chatbot's role and the expectations users might have, designers can consider personality types. Personality, as mentioned earlier in this chapter, is defined as the combination of your behaviour, motivations, characteristics and qualities that forms an individual's character. In short, your personality describes who you are, compared to or distinct from everyone else. While no one person are exactly the same, we can have traits and characteristics in common. Personality theory is the attempt at understanding which factors personalities consists of, and how we can organise these factors into personality types based on which factors we have or have not in common. Carl Jung (1923) coined the term psychology types in which he offered a model to categorise and determine different personality types. His types have since been used to form the bases of type theory, and today there exists several different models to determine personality types, most based on self-evaluation questionnaires. In Jung (1923) theory, the types are based on attitude types: extroversion vs introversion, and function types: sensation vs intuition, thinking vs feeling. An individual often displays either more extroverted or introverted attitudes, while the functions describe four primary types of psychological functions describing ways in which humans perceive the world. Therefore, he proposed eight types:

1. Extroverted or Introverted: sensation-thinking,
2. Extroverted or Introverted: sensation-feeling
3. Extroverted or Introverted: intuition-thinking
4. Extroverted or Introverted: intuition-feeling

While Jung (1923) proposed the early conceptual theory, Myers & Myers (2010)[1980] were the first to offer a type indicator in which one could model personality based on self-assessment. The Myers-Briggs Type Indicator (MBTI) is based on Jung (1923), where they sorted the function types into four dichotomies resulting in sixteen types rather than Jung's eight. These sixteen types are then referred to by four letters. They added the dimension of judgement and perception which describes the individual's preference regarding the other two dimensions, whether they prefer the judging (thinking-feeling) or the perceiving function (sensing-intuition) (Myers & Myers 2010). Other models of personality types are the Big Five (also known as five-factor model or OCEAN) and HEXACO. The Big Five model, is based on five factors (Toegel & Barsoux 2012):

1. Openness to experience
2. Conscientiousness
3. Extroversion
4. Agreeableness
5. Neuroticism

While HEXACO added a sixth dimension (hexaco.org):

1. Honesty-Humility
2. Emotionality

3. eXtraversion
4. Agreeableness (versus Anger)
5. Conscientiousness
6. Openness to Experience

The Big Five and HEXACO are both based on lexical theories, which uses adjectives in language that describes behaviours and tendencies among individuals.

2.3.1 The Big Five

The Big Five "provides a descriptive taxonomy that organizes the myriad natural-language and scientific trait concepts into a single classifactory framework" (John & Srivastava 1999). The Big Five personality framework is the most widely known and used framework to model personality. According to Ackerman (2017) the Big Five can be applied in multiple countries and cultures, and the assessment scale has been found to be valid and reliable for measuring the five factors. Lewis Goldberg defined the five factor model in the 1960's, and the validity of his model was confirmed by McCrae & Costa (1987) which was named the "Big Five". Each of the five factors includes many traits and characteristics that are related, and organised within each factor. The factors includes terms on the dimensions from positive to negative, e.g. generosity and aggressiveness, are both included in the agreeableness factor. In the next subsections each factor will be explained, and a few of the traits and characteristics for each factor will be given in order to give a clear idea of what each factor includes. Every personality is the sum of the traits within the five factors, some have more traits belonging in one factor than the other. John & Srivastava (1999) found that the labels for each of the five factors are often misunderstood, as the labels themselves are not correctly describing the traits, they therefore sought to create short definition to avoid confusion and misunderstandings.

Openness to Experience

> "Openness to Experience (versus closed-mindedness) describes the breadth, depth, originality and complexity of an individual's *mental and experiential life*" (John & Srivastava 1999, p.121).

John & Srivastava (1999) defined openness to experience as the complexity of an individual's mental life and experiences; a person's willingness to try new things, inquiring intellect and imagination. Lebowitz (2016) states that an individual that is high in openness to experience engages in creative careers, enjoys getting to know new people, the arts and learning.

Conscientiousness

> "Conscientiousness describes *socially prescribed impulse control* that facilitates task- and goal-directed behavior, such as thinking before acting, delaying gratification, following norms and rules, and planning, organizing, and prioritizing tasks" (John & Srivastava 1999, p.121).

Conscientious individuals have the tendency to control impulses, have the will to achieve, are hard-working and rule-followers.

Extroversion

> "Extroversion implies an *energetic approach* to the social and material world and includes traits such as sociability, activity, assertiveness, and positive emotionality" (John & Srivastava 1999, p.121).

The factor Extroversion includes two ends of the spectrum: Extroversion & Introversion. Extroverted individuals draws energy from interacting with people, while introverted individuals will become tired from social interactions and draws energy from solitude.

Agreeableness

> "Agreeableness contrasts a *prosocial and communal orientation* toward others with antagonism and includes traits such as altruism, tender-mindedness, trust, and modesty" (John & Srivastava 1999, p.121).

Agreeableness can be explained by how likely you are to be liked by people around you, or how well you get along with others. Also known as social adaptability, likeability, friendly compliance, and includes traits such as: polite, humble, trusting, modest, loyal, unselfish, amiable, and cheerful.

Neuroticism

> "Neuroticism contrasts emotional stability and even-temperedness with *negative emotionality*, such as feeling anxious, nervous, sad, and tense" (John & Srivastava 1999, p.121).

Neuroticism is the factor where a high score indicates more negative traits, while the other factors where individuals have high scores indicate more positive traits. Neuroticism is a factor which explains how comfortable you are in your own skin or how confident you are.

While this has been a short and concise introduction to personality theory and the big five, it describes the most important general understanding of the framework. This understanding of the five factors, and the traits which each consists of, will be used to guide the design of the chatbot personality.

2.4 Designing chatbot personalities

In order to understand the tools to use in order to design a personality for a chatbot, researchers and designers have offered some insights, models and techniques to base the personality on. Through chatbot forums and communities online, designers suggest basing the chatbot's personality on the users who are going to use it, the brand in which the chatbot represents, and then use techniques from character development in order to write the character. In order to design a personality that is based on the people who are going to interact with it, designers must have access to relevant user information in order to understand the user group. User-centred design methods offer several techniques for designers to build a solid understanding of their users. The most common practice is the development of user personas through user research techniques such as interviews, observation, contextual inquiries or market research (Courage et al. 2015). User personas are defined as "concrete representations of the different types of people that the system or service is being designed

for" (Benyon 2014, : 55). A user persona should always be based on data collected through user research, and often there are more than one persona to represent the entirety of the user group. User personas should include aims and goals for using your system (ibid), and can be used for designers to always remember who they are designing for. The great thing about user personas is that they include, names, age, gender, background, and goals and aspirations, as well as frustrations and tensions related to existing solutions or the context in general. If designers want to design user-centred chatbot personalities, they should mirror the user personas by supporting their goals and aspirations as well as having knowledge related to their frustrations and tensions.

2.4.1 The Importance of Social Roles

Even the best persona is still a broad representation of a user group, and can often be prone to stereotypes. Therefore, as there is usually more than one particular user, the chatbot personality should be designed to not only mirror users, but to fit within its social role. As this can help maintain an appropriate dynamic, even when the user's personality is not a direct match. When humans enters a specific social role it sets expectations and goals for the interaction, as mentioned earlier in section 2.2.1, and the role and social conduct becomes more important. The social role of a chatbot is determined by its job, the task it carries out and how a person within that role is expected to behave. E.g. a customer service chatbot might bring about very specific expectations in terms of how it conducts itself and treat users. Therefore, by defining the role will provide a lot of information as well as guiding specific areas of the chatbot persona.

2.4.2 Brand Tone-of-Voice

One way that brands successfully communicates a brand personality is through Tone-of-Voice. The tone of voice of brands is used to guide how the company presents itself through all platforms of communication; whether through digital or printed communications, marketing materials, website and more (Cummings 2017). The brand tone of voice ensures consistency in how the brand is presented, and in particular well-established brands' tone of voice are familiar to consumers. Breaking with their tone of voice can have large implications on how the brand is perceived. The researcher hypothesize that the same is true for chatbot interfaces, and that a well-defined personality and tone-of-voice is important to allow for consistent user experience in which the users perceive the brand through the chatbot. The researcher believes that this will increase trust as the brand tone of voice and the chatbot's use of it will contribute to a more familiar and consistent user experience. In branding the concept of tone of voice is used to inform and design a brand image. Cummings (2017) writes: "tone of voice is not what you say, but how you say it". For chatbot design, how they say it is important to communicate the right personality and to meet user expectations. The tone of voice refers to written and spoken words: the words chosen, their order, rhythm and pace (Cummings 2017). The tone of voice informs a company's written copy, which extends to their website, social media messages, emails and packaging. "A tone of voice both embodies and expresses the brand's personality and set of values" (Cummings 2017). As the commercial implementation of

chatbots are usually as an extension of services companies provide, it would seem that treating a chatbots personality to reflect a company's tone of voice is important to its design.

Tone of voice matters for the design of conversational interfaces such as chatbots, because tone of voice refers to a linguistic message. While the concept of tone of voice has been used in advertising and marketing as a communication tool towards consumers, tone of voice is also used to convey a company's "personality" (Delin 2005, : 2). Examining the effects of a company's tone of voice relates to the link of familiarity with trust, and being consistent will increase familiarity, and this is why defining a clear tone of voice and using this consistently in all communication with consumers, will increase consumer trust. And according to Cialdini (2007) familiarity is an important tool for persuasion. Norman (2007) makes a point out of how not knowing what to expect when interacting with services, computers, products, or humans, because the behaviour is inconsistent, will make people frustrated and irritated. He therefore stresses the importance of matching personality to market segment to be consistent in interaction with people, as he argues that even if a personality is obnoxious, as long as this is consistent you will know what to expect and therefore plan for it (p. 57). Therefore, tone of voice is one of the most important variables to create a consistent communication between companies and consumers. This is an important idea to understand how personality can be used as a variable to plan for different user scenarios and maintain frustration with the system. By applying a personality, the designer will be able to plan ahead to how the chatbot should respond, and also allow users a consistent experience that could help manage expectations.

A chatbot is working 24 hours 7 days a week, and can therefore alleviate customer frustrations, worries, and stress and handle tasks when it suits the consumers; beyond opening hours. Chatbots makes companies more accessible, usable, and can increase user satisfaction. A chatbot can therefore offer great value not only to the services brands offer, but also to consumers. The researcher therefore believes that by creating a framework to design chatbot personalities that incorporates a brand's tone of voice, and offers a consistent user experience, will create more usable chatbot interfaces which benefits both companies and its consumers. This will result in designers making better, informed, user-centred design decisions when designing the chatbot user experience, the interface, and personality, that will result in more users adopting chatbots in their interaction with brands.

The discussion above shows why personality can be a powerful tool in the design of chatbot interfaces as it can contribute to a more consistent, familiar, trustworthy and satisfactory user experience. The next chapter will show how the researcher used this knowledge to form a design framework to build chatbot personalities.

3 Methodology

The methodology for this master thesis project consists of a detailed outline of the chatbot personality framework and how it was implemented to build the chatbot prototype, description of the experiment setup, hypotheses to be tested, and how data was collected.

The combination of the implementation of the researcher's personality framework and the experiment will be used to answer the research questions:

1. How can we design chatbot personalities to guide the design process of a chatbot interface?
   a) Can personality be used as a stable pattern to guide the design process of chatbot interfaces?
   b) Which elements must be considered to inform a chatbot personality?
   c) What components needs to be in place for a chatbot personality to meet user needs and expectations?
2. Will chatbots with a defined personality improve the user experience of chatbot interfaces?

The first research question and sub-questions will be answered through the researcher's experience implementing the developed personality framework. As there are no precise design methodology to build user-centred chatbots with a basis in personality, the researcher has combined techniques from user-centred design, branding, and personality theory in order to build a personality framework for chatbot interfaces. The design process will be user-centred and collect user research through both secondary research techniques and qualitative methods to gather insights into user needs and system requirements. The second research question will be answered through the experiment laid out later in this chapter. This experiment will assess the chatbot prototype against another version of the same chatbot which behaves more like a machine than human; the experiment will be used to collect quantitative data to measure whether the chatbot personality improves the user experience. This chapter will first explain how the personality framework and design process was implemented, second the experiment setup, hypotheses and data collection will be laid out in depth.

3.0.1 Brand, Domain, and Usage

To build the chatbot prototype, test the personality framework, and follow a user-centred design approach, the chatbot domain will be based on a real brand. There are no formal collaboration with this brand, therefore it will be anonymised in this thesis. It was necessary to use a real life example to base the prototype on in order to show how the chatbot personality represents the brand's tone of voice, mission and values. This also informed user personas, and suitable users for the chatbot prototype to model the personality on. In addition to this it also informed the role and job the

chatbot should have to add value to users and support the mission of the brand.

The chosen brand and domain was chosen at random and only to be used to provide as an example to apply and test the framework. The framework can however be used to design chatbot personalities for any brand and any domain. The chatbot prototype will be used to further the mission of the brand which is to increase the consumption of fruits and vegetables and reduce food waste.

3.0.2 User group

The intended user group for the chatbot prototype are between the ages of 25 to 40, aiming at young couples living together, preferably with young children. Research have found that this age group eats less fruits and vegetables and waste the most food compared to other age groups and other life situations (Haraldsen 2011). Therefore, to support the brand's mission of increased consumption of fruit and vegetables, the chatbot will focus on this target audience. This group usually have hectic days where healthy eating and activity can be difficult to maintain, and some are, or soon to be, in charge of their children's diet and activity levels as well. As learning good habits starts when we are children, parents have a major impact in regards to teaching children the right habits. This shows that this group in particular are in need of a service which can help them plan healthy meals for themselves and their families, receive assistance to find easy and economical ways to add more nutritious produce to their meals, which will contribute to less food wasted.

3.1 The Personality Framework

To build the personality framework the researcher identified four components that the chatbot personality must be based on:

1. The brand mission, goals and values
2. A deep understanding of the users and their needs
3. The role/job of the chatbot
4. An appropriate personality model

The first component must be met to ensure that the chatbot's personality and behaviour is consistent with the goals, values and tone of voice of the brand it represents, and supports the mission of the brand. The second component must be met to ensure that the personality supports the goals of the users, and to determine which personality traits that are appropriate for the user group. The third component is important as it dictates the social role of the chatbot, which again will help find appropriate traits that are compatible with its role. The final component, an appropriate personality model, is necessary to organise and map out the personality traits into a suitable framework. The following sections will explain how the personality framework was applied to build the chatbot personality that will later be the basis of the user experience experiment.

3.1.1 Design process

The design process used a user-centred design methodology; to gain insights into the needs of users and inform the requirements of the chatbot prototype. This UCD process was divided into the stages of: 1) inspiration 2) ideation 3) implementation (following the human-centred methodology as proposed by IDEO.org (2015)). According to Gould & Lewis (1985) there are three key principles of UCD: an early focus on users and tasks, empirical measurement of product usage, and iterative design. The three stages will therefore all follow the key principles of UCD, where the first stage (inspiration) will focus on user and domain research (component 1 and 2 of the personality framework), ideation will focus on designing and building the chatbot prototype and define the personality (component 3 and 4 of the personality framework), and implementation will focus on the evaluation of the final prototype, which will evaluate the personality in regards to how it's perceived and its effects on the user experience. As a UCD approach is characterised by empirical measurement, iterative design, and focus on users, each deliverable will be empirically tested through user testing techniques in iterations to inform the final prototype, while the final evaluation will be used to gather quantitative data to answer the second research question, and belonging hypotheses (stated in 3.4).

3.1.2 The Brand Mission, Goals, and Values

To design a chatbot that conforms to the brand it represents, the brand image was analysed and defined. The mission statement was identified, the company and core values was defined, and a tone-of-voice analysis was conducted.

Mission statement and core values

The brand describes itself as having a broad social responsibility. They supply fruits and vegetables, with suppliers on all continents providing products from all over the world. Freshness is the top priority and high quality. In addition to providing fresh produce, the brand recognises having a fair and sustainable business model, which extends to proper nutrition, increased activity, and societal and environmental responsibilities. Their goal is to increase the consumption of fruit and vegetables, therefore in addition to supplying fresh produce, they take an active role in increasing knowledge and engagement regarding healthy lifestyles and positive societal development.

Mission

Their mission: *Fresher and Healthier*; to contribute to increased focus on healthy diets and physical activity. A driver and pioneer for sustainable development regarding the environment and societal responsibility within the company's product area.

Brand Values

To create value for the customers, is the brand's driving force throughout the business. The conduct of employees are characterised by the values stated in the first column in Table 1 and the core values in the second column:

Internal Values	Core Values
Team spirit Courage to do the right thing: Openness, honesty, probity Compassion: mutual respect, care, and tolerance	Vital Current Fresh Daring

Table 1: Internal values and core values

Tone of Voice analysis

If the chatbot is to be perceived as an extension or continuation of the brand, it must adhere to the brands tone-of-voice. Tone-of-voice is defined as a brands personality, communicated through its written communication as well as the visual communication. Norman & Nielsen (Meyer 2016) defined a framework to determine tone-of-voice based on 4 dimension: funny vs. serious, formal vs. casual, respectful vs. irreverent, enthusiastic vs. matter-of-fact.

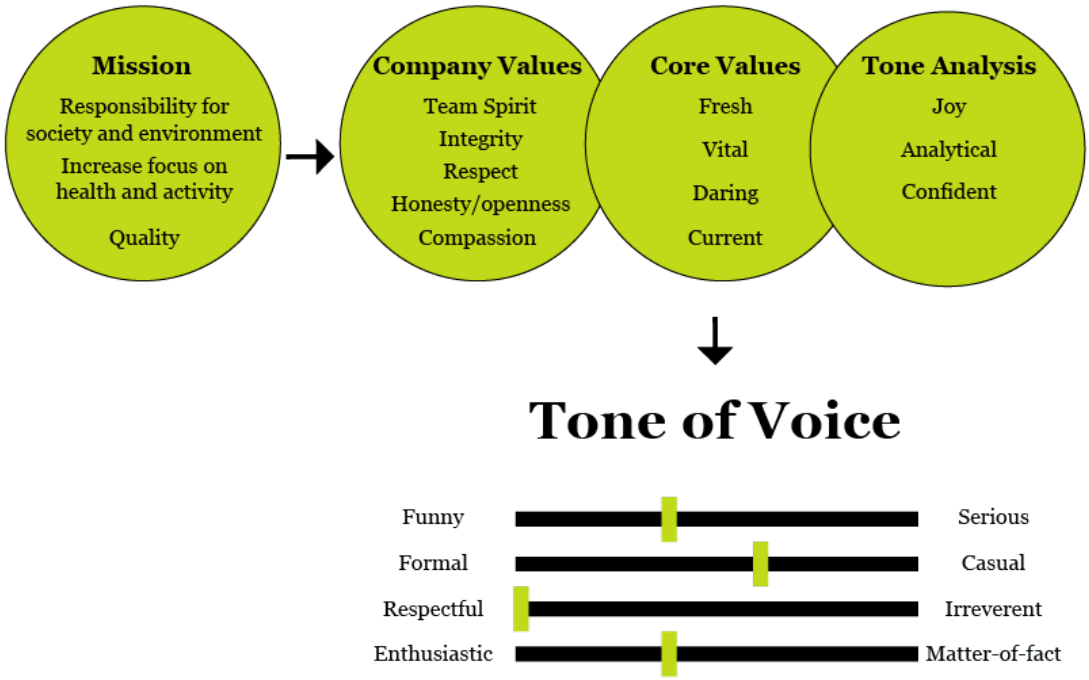


Figure 1: This figure shows how the tone-of-voice was defined by understanding the mission, the values and conducting a tone analysis of the written copy

Figure 1 shows how the tone-of-voice was determined by assessing the mission statement, the company's values inwards and outwards, and through a tone analysis. The tone analysis was conducted using the IBM tone analyser tool, the analyser assessed the written copy found on the brands website. Their tone-of-voice as reflected in their copy is more serious than funny, but not too serious as the tone is also a nice in-between of formal and casual. Their tone is very respectful, which follows their core values of tolerance and respectfulness. Lastly they are leaning towards more enthusiastic than matter-of-fact as they carry a very positive and joyful view of the future and their path to fulfil their goals.

3.1.3 Understanding User Needs

To meet the second component; to build a deep understanding of the users and their needs, both primary and secondary research was gathered in order to understand the user group, trends, causes and statistics regarding healthy eating and food waste.

User Interviews

A series of interviews was conducted in order to understand the experiences of the users; how do they view their intake of fruit and vegetables, what obstacles and challenges do they face, how they assess their own habits, and their thoughts regarding food waste etc. Eight users were recruited to participate in these interviews, or four couples, in the ages of 29 to 36, four mothers and four fathers. All four couples had two children in kindergarten and/or early elementary school age. Six of the participants works full time, while one mother was on maternal leave as the interviews were conducted. Two participants, one male and one female, were part-time workers and/or students at the time of the interviews. The interview guides prepared were semi-structured and aimed at mapping the daily routines, views, habits, pains and frustrations of the users during an average week. In particular the interviews aimed to see how parents assess their own eating and activity habits, and whether they are aware of food waste occurring and if so why. The interview guide can be found in Appendix A.1.

User Personas and User Scenarios

Based on the findings from the user interviews and information from secondary research, two user personas were developed, one male and one female, (see A.2), to summarise the user research and inform the personality of the chatbot. The user personas describes which goals the users have, as well as their frustrations and pains during an average week, what motivations that drives them and their specific needs. The user persona acts as a summary of the insights from user research, and are helpful tools to inform the requirements of the system. By understanding users' preferences, goals and pains will help determine suitable characteristics and personality traits. The personas described two parents that feels guilty because they sometimes have to make shortcuts to achieve everything they have to do in a day. E.g. sometimes buy processed foods or ready meals, throw away leftovers, and go for unhealthy alternatives. This informs us that they are aware, and feel guilty about their shortcomings, therefore a lecturing tone might not help ease this guilt, but rather increase it. The personas are therefore helpful to understand which traits that are appropriate and

supports the goals of the users. E.g. the personas also informs us that the mother is more likely to be concerned about eating healthy and implement changes, while the father wants to support the mother in these goals. Therefore the personality should support both parents in this. The personas were used alongside other important findings to write specific user scenarios in which the chatbot solves a problem or need for the user. The user scenarios followed the method laid out by Benyon (2014), which uses insights from user research to describe user stories, to conceptual scenarios, to concrete scenarios to use cases (see Figure 2). Examples of the conceptual stories can be found in A.3. The use cases formed the conversation flows, an example of these can be found in A.4. The most important tasks that added the most value to users, was to make use of the chatbots AI to make planning dinners easier. Therefore the chatbot will help its users with planning meals on a weekly basis, each recipe will provide the necessary nutritional value needed for each member of the family, and the recipes will be "smart": generate grocery lists automatically and ensure as little leftovers as possible (both in the fridge/cupboards and on the plate). This will guarantee healthy eating for all users, decrease the needs for trips to the grocery store, and prevent food waste to a degree.

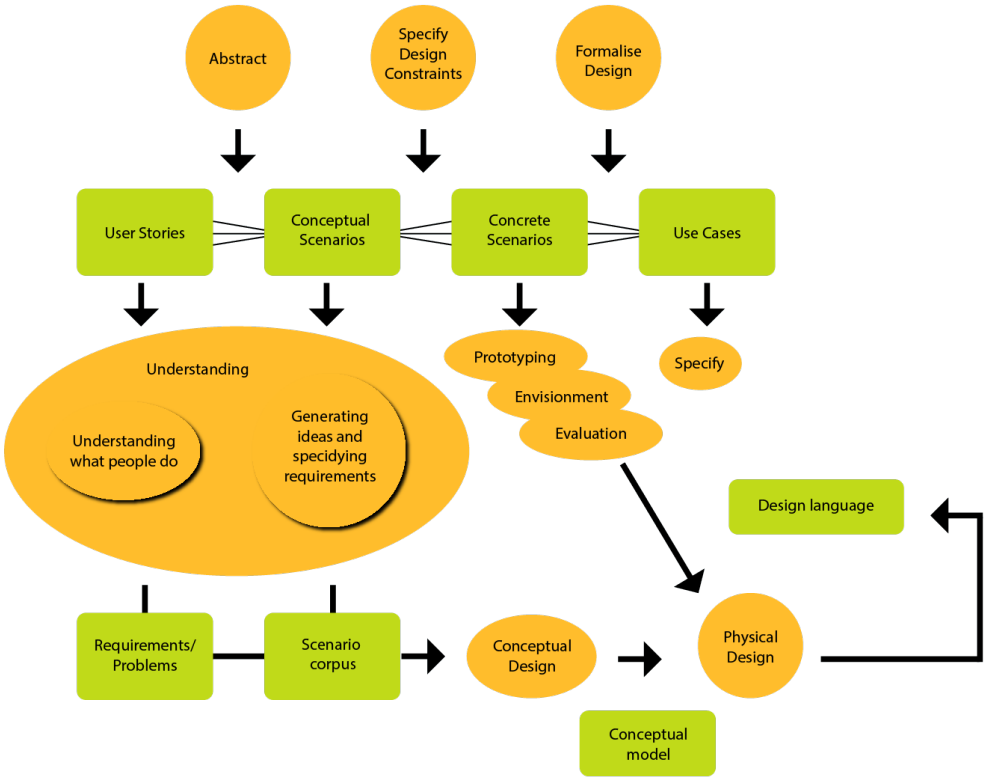


Figure 2: Scenario-based design method as discussed in Benyon (2014)

3.1.4 The Chatbot Role

The third component of the personality framework consists of understanding the role the chatbot should play, or the job it should do. The user-centred approach identified the two most important jobs the chatbot could do that both furthers the goals of the brand and add value to the users: increase the consumption of fruits and vegetables, and reduce food waste. Therefore, the job of the chatbot agent is to *assist* users to help them make healthier choices for their family and *motivate* them to successfully implement these changes. The chatbot is there to assist and guide the users, with the long term goal of successfully implementing a fresher and healthier lifestyle and reduce food waste. To achieve this the chatbot will help parents plan their dinners for the whole week, assist with the shopping and make recommendations based on ingredients they have in house and leftovers. In addition to this the chatbot encourages parents to inform the chatbot about what they have eaten so far, what was a success and what was not so that the chatbot can learn which items are not eaten, are wasted and instead offer healthy alternatives that will be eaten rather than wasted.

Assistants and motivators have specific traits (Burge 2016, Lipcamon 2013) in which they need to have in order to be successful in their role (see Table 2).

Assistant	Motivator
Professionalism	Give praise and encouragement
Collaborators	Treat clients as equals
Outstanding organisational skills	Show trust
Excellent communication skills	Communicate and set goals
Willingness to go the extra miles	Be attentive
Problem-solver	Allow mistakes
Proactive	Be pleasant
Respectful	Ask for feedback
	Keep others informed
	Don't micromanage

Table 2: Desirable traits for Assistants and Motivators

The traits of the assistant will be reflected in the way the chatbot handles its job. Through AI and NLP capabilities, as well as connection to necessary APIs, will the chatbot be able to complete necessary tasks and incorporate "hidden/invisible" features that will help users achieve their goals. As for the motivator traits, this will be reflected in how it encourages/motivates users, the language its using, and its behaviour (prompts, affirmations, tips). Therefore the chatbot role will be reflected by its external traits (motivator) and its internal traits (assistant).

3.1.5 Personality Trait Model

Once the brand mission, core values, understanding of user needs, and the chatbot role have been identified, designers should find an appropriate personality trait model to help put together a dynamic personality for the chatbot. It is important that this model will be used to place the desired

personality within a framework, this is to benefit the designer when designing a consistent personality.

It is not the goal of this thesis to assess whether the designed personality is consistent with the chosen trait model, or to assess the personality of participants. The trait model is only used to guide the design of the chatbot personality, and to help in evaluating desirable traits for the specific chatbot agent.

The chosen personality model to model this chatbot's personality was the five factor model. As explained in section 2.3.2, this personality trait model is widely known, and based on lexical data used by humans to describe personalities. These traits have been mapped out and organised into the five factors, and as the traits are identified by the words we use to describe them, it will be an appropriate model for a linguistic interface such as a chatbot.

3.1.6 The Chatbot Personality Description

The final chatbot personality therefore were based on identified user needs and user personas, the brand image and tone of voice, and the identified role of the chatbot. This information was used to model the desirable traits and quirks of the chatbot personality, and the five factor model was used as reference to model the personality. The personality includes the traits shown in Figure 3. The role and character description is laid out in next two sections; both descriptions are based on the user research in regards to how the chatbot's tone and tasks are handled to support the needs and goals of both genders.

Role description

The chatbot has been given the name Bella. She works as a dinner planner, helping couples plan meals for the whole week. Her recipes helps couples eat a healthy and balanced diet focusing on increasing the consumption of fruits and vegetables. Bella is personalised to each specific family as she knows and learns over time what they prefer to eat, what they dislike, any known allergies they might have, and who makes up the whole family. Her mission is to increase their health, while also taking an environmental and economical approach to dinner planning; keeping track of what food they have at home, base recipes on ingredients that needs to be used up and providing users with shopping lists. It is important that Bella works to motivate users towards their goals, and does so in a supportive fashion. Her role as assistant makes sure that she is trustworthy, reliable and efficient.

Character description

Bella is designed to act and look about the same age as her target audience, and works to assist and motivate her users. Her personality is modelled to be that of a supportive friend, helping with cool new tricks as well as emotional support, by using a cheerful and playful tone. She is always happy to help and conducts herself to a high professional standard. She always acts respectful and polite, but likes to add humour to her interactions with users. Bella is designed to help her users throughout the day, by assisting with logging meals, provide help, tips and tricks, reminders, shopping lists and recipes on the go. She also offers words of encouragement and wants to support and motivate her users towards their goals. Therefore Bella acts efficiently and to the point, while also providing

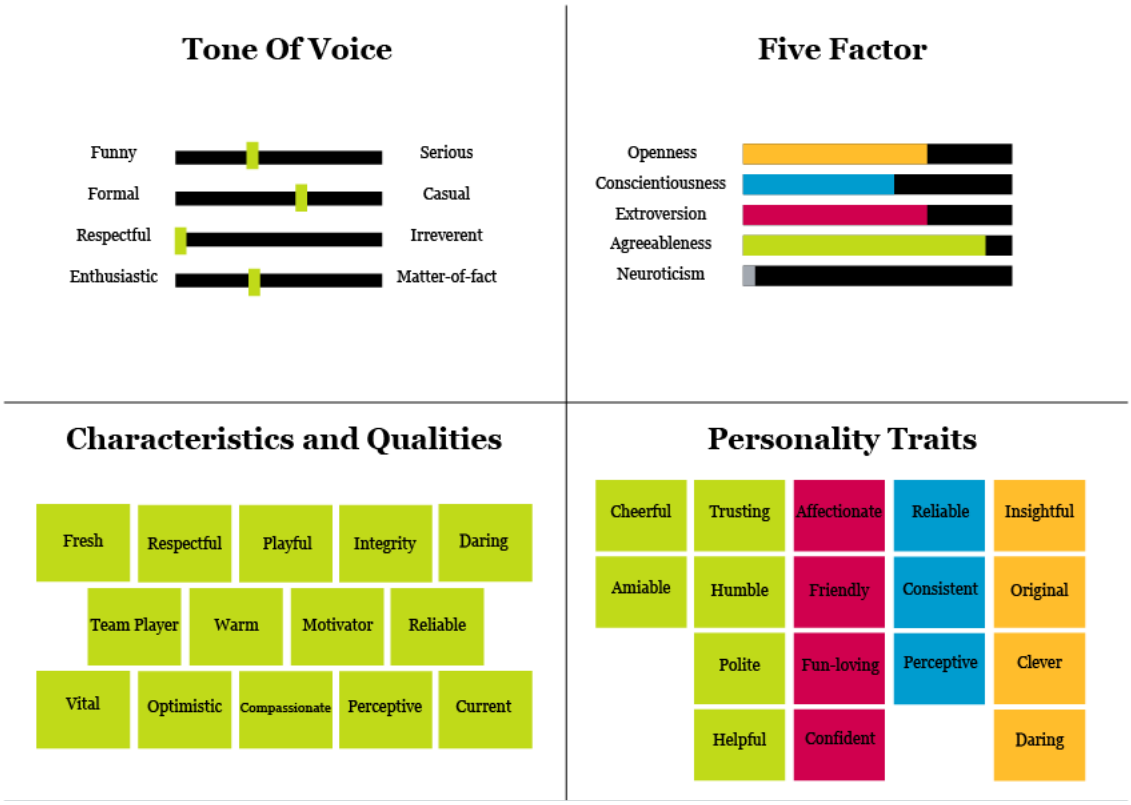


Figure 3: The defined tone-of-voice, five factor personality traits, characteristics and qualities

a helpful and supportive tone. She is cheerful, youthful, and fun-loving, while also being reliant, consistent, efficient and trusting. Bella is a team player, and works to support great collaboration between couples.

3.2 Prototype & Conversation Design

The chatbot prototype was built using the Chatfuel bot builder platform, and the experiment was run through Facebook's Messenger platform. The conversation design was written by mixing a user scenario technique with mind-mapping tools, such as Xmind, to map out the different chatbot skills, user intentions, and conversation flows (see 3.1.3 & A.3). To improve the AI of the chatbot, the researcher wrote more than 300 unique training data per chatbot skill. This resulted in participants testing eight different conversation flows, three of which can be seen in Figure 4. These eight chatbot skills included, planning dinner for the whole week or specific day, help using leftover ingredients, log five a day, help eating healthier or increase consumption of vegetables, add and locate shopping lists.

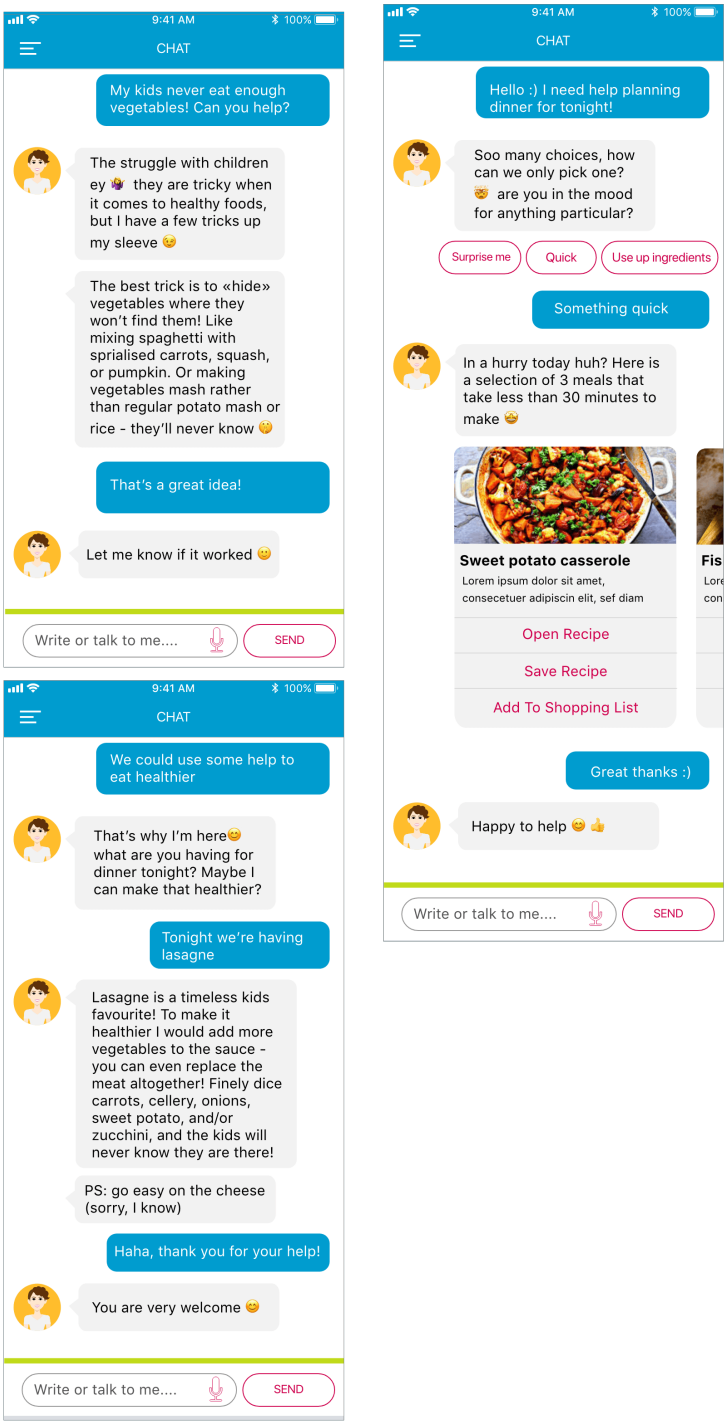


Figure 4: Example of conversations users had with Bella

The conversation design began by forming different use cases, that were based on the concrete scenarios built around the insights from the user interviews. These use cases described the most important tasks the chatbot could provide to its users, and the conversations were written around these tasks. The goal of the conversation would be to complete the specific task described in the use case, and this of course can be done in many different ways, which gives you the possibility to play around with multiple paths to the same conversation goal. The main path was first identified, before creating different "sub-paths" to reach the same goal.

Training data for chatbots consists of variations of the user intention. This means as many versions of the same question or user input that describes the same goal. The training data is what builds the chatbots AI, and the more training data a chatbot has the more certain it becomes in predicting the right intent. If the chatbot is very well trained, it would be able to answer questions or input that has not been included in the training set, but describes the same intent.

3.2.1 The Chatbot Avatar

Figure 5: Chatbot Bella's avatar

Chapter 2.2 provided background and research regarding how humans perceive the various types of conversational agents, and found that anthropomorphism can benefit how a personality is perceived, and that humanness can increase anthropomorphism. Humanness relates to human capabilities and physical characteristics, and chatbots with a high level of humanness increases trust and familiarity, while also providing a more natural interaction with humans. The latter is explained by that humans know how to interact with other humans, therefore will assume that a chatbot will interact like a human if it has a more human appearance. Because of this, the chatbot avatar will have a human appearance, to encourage natural human interaction. However, research have found that too human might have a negative effect on human perception. Therefore, the chatbot will be portrayed as an illustrated character rather than a realistic human, as can be seen in Figure 5. This will make clear the distinction that the chatbot is not a human, while incorporating high levels

of humanness simultaneously, which will contribute to increased anthropomorphism. The chatbot has been given a female gender. While aware of the research regarding stereotypes and unwanted attention towards female CAs (see section 2.2.4), the design process determined that the role of the chatbot and the personality befitted a female agent better for this specific scenario.

3.3 Chatbot B - the other chatbot

In order to test whether the chatbot personality has any impact on the user experience, the experiment will test two levels of personality. The first level (Chatbot A) will include the personality modelled using the developed personality framework. The second level (Chatbot B), will be designed to appear as opposite to the personality of Chatbot A. This will mainly involve using the same traits that Chatbot A's personality has been modelled on and to the extent possible remove them from the personality of Chatbot B. Chatbot B will therefore appear as having a less human personality than Chatbot A. Chatbot B cannot be defined as having "no personality", as even a "machine-like" behaviour can be defined as a personality type. Instead Chatbot B has been designed to be mainly task-oriented, providing the same service and completing the same task as Chatbot A, but in a machine-like way as shown in Table 3.

User Expressions	Chatbot A	Chatbot B
What should I cook for dinner tonight?	Cool cool;) What are you in the mood for?	Do you have a preference?
Something that's quick to make	In a hurry today huh? Here's a selection of 3 meals that take less than 30 minutes to make	Quick recipes:
Dinner tonight was delicious!	That's wonderful :D should I recommend this recipe again?	OK, recommend recipe in future?

Table 3: Difference in personality in responses between Chatbot A and Chatbot B

This means that Chatbot B will provide the same value to users as Chatbot A, by meeting their needs and being a great assistant. However, the motivator role will be affected by the different personalities, as Chatbot B will not act as a great motivator as these traits are mainly found in the Agreeable and Extroversion factors. Therefore Chatbot A and B will have the following traits in common: Reliable, Consistent, and Perceptive. These are the traits found in the conscientiousness factor, and mainly displayed through the assistant role. The traits found in the Openness factor will also be removed to some extent in Chatbot B, however some of the traits might be reflected in the tasks it performs and will therefore to some extent be present in Chatbot B. The prototype for Chatbot B was built by duplicating Chatbot A, and change the way it responded to fit with its defined personality characteristics.

3.4 Experiment Design

The experiment will be conducted in two parts in order to answer the second research question: *Will chatbots with a defined personality improve the user experience of chatbot interfaces?*. The first part of the experiment (Experiment Part 1) consists of evaluating whether participants perceive the intended personalities for both chatbot versions. This it to control that the personality has been modelled correctly, and are perceived as such. This will also test whether the personalities are perceived consistently by all participants, and that they are in agreement. Then, the second part of the experiment (Experiment Part 2) will assess whether personality has an effect on the user experience. Each part of the experiment will be conducted simultaneously, and repeated for each chatbot version as seen in Table 4.

Experiment Design				
Group 1	Chatbot A	Evaluation Part 1 + Evaluation Part 2	Chatbot B	Evaluation Part 1 + Evaluation Part 2
Group 2	Chatbot B	Evaluation Part 1 + Evaluation Part 2	Chatbot A	Evaluation Part 1 + Evaluation Part 2

Table 4: Experiment Design of the two-by-two factorial design showing the starting condition and when participants will evaluate the chatbot versions

3.4.1 Experiment Part 1

Independent Variable: Characteristics
Dependent Variable: Personality

Part 1 of the experiment, to assess whether participants perceive the intended personality, will be assessed by asking participants, after interacting with the chatbot, to rate on a five point Likert scale whether they perceived each characteristic. As seen in Table 4, participants will begin with either Chatbot A or Chatbot B, asked to rate the characteristics right after the first interaction, then interact with the second chatbot and then rate the characteristics for that chatbot as well. The first part of the evaluation will use the following hypotheses:

$H_1 1$: *Users will perceive the personality of Chatbot A as different to the personality of Chatbot B*
$H_0 1$: *Users will not perceive the personality of Chatbot A as different to the personality of Chatbot B*

$H_1 2$: *Users will perceive the personality of Chatbot A as intended*
$H_0 2$: *Users will not perceive the personality of Chatbot A as intended*

$H_1 3$: *Users will perceive the personality of Chatbot B as intended*
$H_0 3$: *Users will not perceive the personality of Chatbot B as intended*

Data Collection Experiment Part 1

To collect the necessary data to assess the stated hypotheses for part one of the experiment, the participants will be asked to evaluate the predefined characteristics of the personalities each chatbot is modelled on, on a five-point Likert scale. The data collection forms are the same for each chatbot. The characteristics included in the data collection form are the characteristics Chatbot A's personality is modelled on, as seen in Table 5. Chatbot B only has three characteristics in common with Chatbot A: 1) Reliable 2) Consistent 3) Perceptive, which are found in the conscientiousness factor. Therefore, if the personality for Chatbot A and the personality for Chatbot B is perceived as intended, the results should have these three characteristics rated to a high degree by all participants. All other characteristics should be rated to a low degree for Chatbot B, and to a high degree for Chatbot A. This experiment is conducted to determine to what extent participants perceived the given traits, and whether participants are in agreement of the perceived traits. This will help determine whether the chatbot script has been successful in displaying the correct and intended personality; or whether participants perceive a different personality for one or both chatbot versions. It will also inform to what degree the two personalities are seen as different by all participants. All participants will be subject to all conditions, and asked to evaluate each chatbot versions independently.

Characteristics Data Collection Form					
	1=not at all	2=to a small degree	3=to some extent	4=to a high degree	5=to a very high degree
Cheerful					
Trusting					
Polite					
Helpful					
Affectionate					
Reliable					
Fun-loving					
Confident					
Consistent					
Perceptive					
Insightful					
Original					
Clever					
Daring					

Table 5: Data collection form presented to each participant to rate the chatbot personality against the given characteristics

3.4.2 Experiment Part 2

Independent Variable: Personality
Dependent Variable: User Experience

Part two of the experiment will be conducted to assess whether personality affects the user experience, and will do this by manipulating the independent variable *Personality* into two levels (Chatbot A and B), to assess whether it has an effect on the dependent variable *User Experience*. The second part of the experiment uses the following hypotheses:

$H_1$4: *Personality affects the user experience of chatbots*
$H_0$4: *Personality has no effect on the user experience of chatbots*

$H_1$5: *Chatbot A will have an improved effect over Chatbot B*
$H_0$5: *Chatbot A will not have an improved effect over Chatbot B*

In addition to these hypotheses the researcher will also collect data on each participants preferred version as well as their reasoning for this. The researcher assumes that the motivator role might be more appropriate for female users, especially if they are mothers, as they are more likely to use the chatbot throughout the day for help and guidance, while male users are more concerned with how efficiently and usable it is in completing tasks. Therefore collecting additional data on preference and reasoning, might help inform this assumption and could be interesting to see next to the final results.

Data Collection Experiment Part 2

To collect the necessary data to assess the stated hypotheses for the second part of the experiment, the AttrakDiff measuring tool will be used to assess the effect on the user experience. All participants will be subject to all conditions, and asked to evaluate each chatbot versions independently.

AttrakDiff - operationalising the user experience

User experience is defined in ISO 9241 – 210 as "all the users' emotions beliefs, preferences, perceptions, physical and psychological responses, behaviours, and accomplishments that occur before, during and after use" (ISO 2010). Usability is the most widely known definition to determine whether a product is good or bad, and therefore an important part to determine a great user experience. Usability is defined in ISO 9241-210 as the "extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". Hassenzahl (2006) believes that this definition is too task oriented, focusing on task completion and reaching goals, simplicity and efficiency, and forgetting about the "fun". Satisfaction is therefore one of the components of usability in which professionals, researchers, and designers alike struggle to agree on a definition. Is it the satisfaction of efficiently completing a task, or the satisfaction of an overall enjoyable experience? The AttrakDiff measure-

ment tool was built to assess exactly this, knowing that users might choose a product which is slightly less efficient but are extremely enjoyable to use. While some argue that something cannot be enjoyable unless it is efficient, we must not forget all the different ways in which the overall user experience is affected by a product that "stands out" from the rest. Hassenzahl et al. (2000) built the AttrakDiff measurement tool, which assesses the user experience by looking at usefulness and usability in the pragmatic quality, independently from the hedonic qualities of stimulation, challenge and motivation, and attractiveness. The AttrakDiff form assesses personal user rating of a products usability and design.

- Pragmatic Quality: Usefulness and usability of the system
- Hedonic Quality: Motivation, stimulation and challenge for the user

The AttrakDiff measurement instrument consists of 28 seven-step items of opposite adjectives ordered into a scale of intensity (AttrakDiff.de 2013). The middle values of an item group creates a scale value for pragmatic quality (PQ), hedonic quality (HQ - include HQ-I and HQ-S) and attractiveness (ATT). HQ-I and HQ-S are the sub-qualities of stimulation and identity of hedonic quality. The theoretical model was created and tested by Hassenzahl et al. from 2000 through 2006 (Hassenzahl et al. 2000, Hassenzahl 2001); and their studies have found that the hedonic and pragmatic qualities are perceived independently from each other and consistently, and contributes equally to the attractiveness rating. According to their website, Hassenzhal et al. (2000) found that the model separates four essential aspects: 1) the product quality intended by the designer, 2) the subjective perception and evaluation of quality, 3) the independent pragmatic and hedonic qualities, and 4) the behavioural and emotional consequences (AttrakDiff.de 2013).

The AttrakDiff measurement tool was used to collect the appropriate data to assess and compare the two chatbot personalities against the user experience. The AttrakDiff evaluation will be used to assess the user experience of both chatbot prototypes. The pragmatic quality will asses usability and usefulness of the chatbot, while both hedonic and attractiveness qualities will be used to assess the satisfaction with each version.

3.4.3 Experiment Setup

Participants were recruited through convenience sampling, all participants were within the age group of 25-40 years of age, the sample consisted of couples, either married or unmarried, but all living together. 12 of the 16 participants had children in kindergarten or elementary school, while two couples are not parents as of yet. Eight of the participants had also participated in the earlier interviews. The participants were invited to test a new chatbot application that aims to help families with weekly tasks. This means that none of the participants were aware that the goal of the experiment is to test the chatbot personality, but rather that they were invited to test two versions of a chatbot interface. Before the participants began the tests, they were given one deck of eight cards, see Figure 6, each describing a need participants have. Participants will be asked to open one card at a time and use the chatbot to solve that need e.g. "I don't know what I should make for dinner tonight". In addition to this deck participants will also be presented with four different coloured

decks, see Figure 7 and 8, they will be asked to open one card from each of the four decks that are named: 1) Dinner 2) Ingredient 3) Vegetables 4) Fruits. Participants were instructed to provide this information when asked by the chatbot. E.g. one of the conversation flows includes adding an ingredient to the shopping list, your ingredient card informs the participant which ingredient to add. The cards were necessary as it was important that the experiment was completed with as little interference from the test moderator as possible. This was to not influence ones own personality into the interaction with the chatbot solution. The two chatbot versions are prototypes, and therefore prone to error during testing. Time restraints made it impossible to produce a complete well-functioning chatbot, this made it necessary to design the experiment to only include specific conversation flows presented by the cards mentioned above. Pilot studies and preliminary studies conducted during the autumn 2017, found that errors occurring during a chatbot user test has a huge impact on participant's perception of the chatbot as a whole, as participants rated the chatbot as being lower in intelligence, helpfulness, and usefulness if it predicted the wrong intent. Therefore it is very important that the chatbot does not give a wrong answer or fail to predict the right answer during the experiment.

The decks were necessary to exclude any interference from the researcher, and also ensured that participants were engaged in conversations flows that the chatbot had been trained to know well. This allowed the researcher to focus on only eight specific conversations to allow participants to interact with the system. The preliminary study conducted as preparation for this master project found that users need a goal with the interaction and expect a chatbot to be able to do something, or perform some kind of service, to perceive the interaction as meaningful. Participants were unable to give a true evaluation of their perception in the preliminary study if they did not "see the point" with the interaction, as there was a conflict when rating characteristics such as intelligent or helpful. Therefore, participants will use the chatbot as a complete service, even though the research aim is not to assess how well the chatbot performs its tasks. In order to conduct a true assessment of user experience, the product being tested must be seen as useful to its users, this will be reflected in the pragmatic quality score from the AttrakDiff test. The design process and participant sampling will already have assured that the tasks the chatbot performs are valuable or useful to the participants, which means that the Pragmatic Quality of each version is assumed to be scored high by all participants - although maybe not to the same degree.

Participants will evaluate the two chatbots by completing a series of tasks using each chatbot. As mentioned earlier in this chapter, the first part of the experiment will be used to assess whether the personality is perceived consistently and as intended by each participant, and the second part to assess the user experience. In order to compare the two chatbot versions, the participants will be their own control group as the experiment design will allow for a between & within-subjects design using a two by two factorial design, see Table 4. Half of the group will test Chatbot A first, while the second half will test Chatbot B first. This is to avoid a sequence/interaction effect, where participants become affected by which chatbot they try first. Participants will not have any comparison during the first interaction, and can therefore be either stricter or nicer in their assessment. While the second interaction will be compared to the first interaction, and can therefore also be

assessed stricter or nicer. Participant's might also feel more at ease during the second interaction, as they now know what to do. After the participants have interacted with each chatbot they will be asked to assess the chatbot personality characteristics to determine whether participants are in agreement regarding the perceived characteristics and personality traits. Then participants will be asked to fill out the AttrakDiff evaluation to assess the perceived hedonic-, and pragmatic quality and attractiveness of the chatbot, before repeating the same process again with the second chatbot version. This experiment design will ensure statistic power. Both the personality characteristics data collection form and the AttrakDiff measurement tool will be presented to all participants after they have interacted with each chatbot. The participants will be presented with the same form for each bot, and the forms will be accessed locally through the researchers private computer in order to ensure participants anonymity.

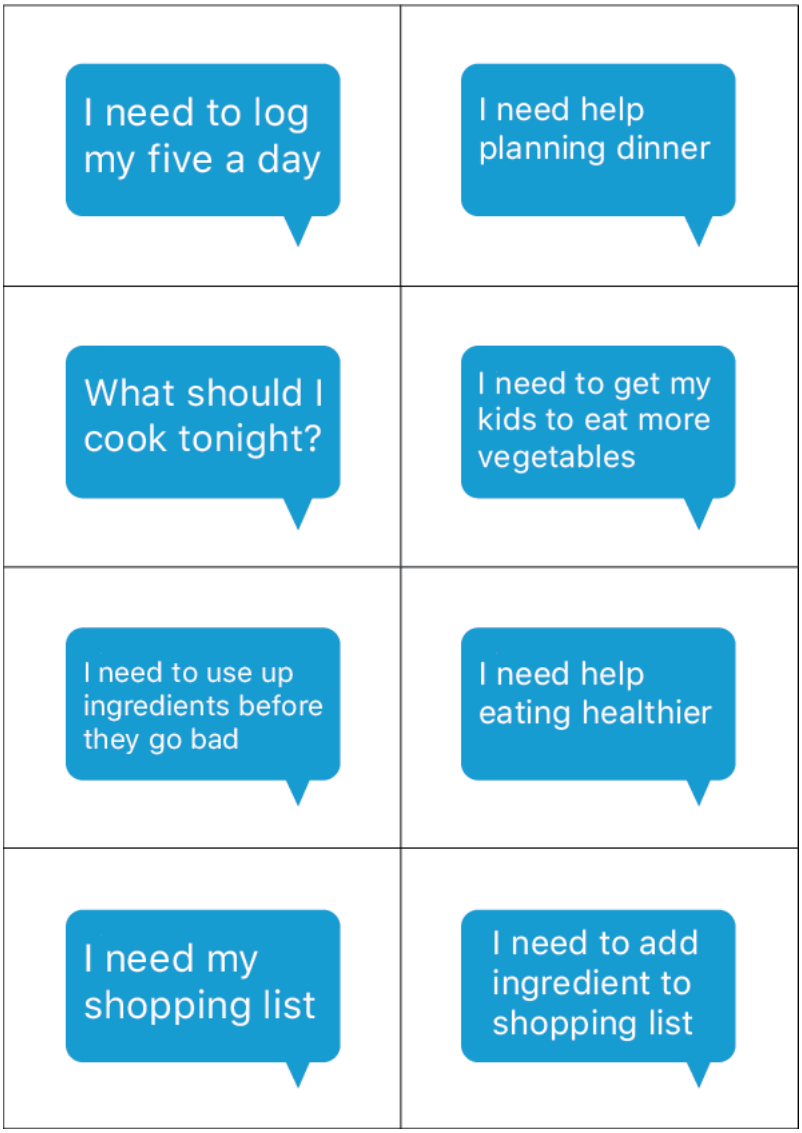


Figure 6: Figure showing the eight cards with a need to be solved by the chatbot

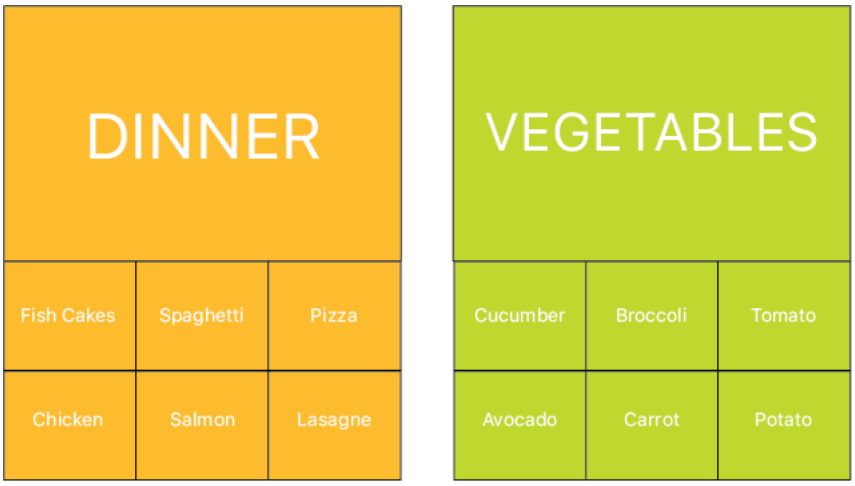


Figure 7: Figure showing the cards for dinners and vegetables

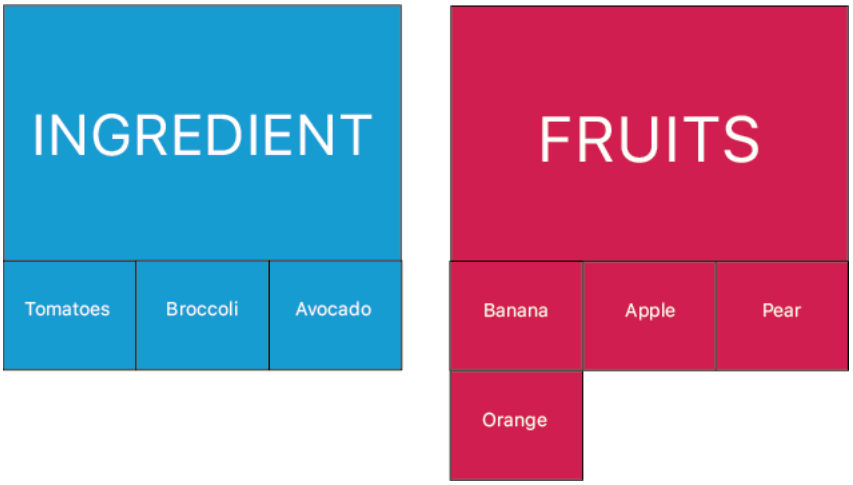


Figure 8: Figure showing the cards for ingredients and fruits

4 Results

This section will provide the results from the data analysis of both parts of the experiment; 1) the agreement regarding the personality traits and characteristics, 2) the AttrakDiff evaluation to assess the user experience. The data was analysed using SPSS.

4.1 Part 1 Results: Personality characteristics data analysis

The characteristics were rated using a scale from 1-5 in which 5 meant a high perceived presence of the characteristic in the interaction with the chatbots, and 1 meant no perceived presence of the characteristics. The data-set was analysed through running a paired-samples t-test, and a two-way repeated measures ANOVA to investigate whether an interaction effect occurred within subjects, as all participants were subjected to all conditions. The independent variable personality has two levels, and the dependent variable characteristics has four factors where each factor describes a factor from the five factor model (extroversion, agreeable, conscientiousness, openness). The analysis performed on this data set will be used to test $H_1 1$: *Users will perceive the personality of Chatbot A as different to the personality of Chatbot B*, $H_1 2$: *Users will perceive the personality of Chatbot A as intended*, and $H_1 3$: *Users will perceive the personality of Chatbot B as intended*. The descriptive statistics of the characteristics data analysis can be found in Table 6.

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Extroversion_A	16	2	5	4,3125	0,8116
Extroversion_B	16	1,67	3,33	2,2708	0,47483
Agreeable_A	16	4	5	4,625	0,30277
Agreeable_B	16	2,25	4,5	3,4219	0,57532
Conscientious_A	16	3,33	5	4,2917	0,5146
Conscientious_B	16	2,67	4,67	3,9167	0,60246
Openness_A	16	3	5	4,0781	0,59665
Openness_B	16	1,5	5	3,4531	0,92294

Table 6: Descriptive statistics chatbot characteristics results

The paired-samples t-test found that there is a significant difference between the two levels of personality and the perceived characteristics. Agreeableness Chatbot B (M=3,4219, SD=,57532) and Agreeableness Chatbot A (M=4,6250, SD=,30277); t(15)=-8,760, p < ,001. Extroversion Chatbot B (M=2,2708, SD=,47483) and Extroversion Chatbot A (M=4,3125, SD=,81166) t(15)=-8,976, p < ,001. Openness Chatbot B (M=3,4531, SD=,92294) and Openness Chatbot A (M=4,0781 SD=,59665) t(15)=-3,727, p = ,002. Conscientiousness Chatbot B (M=3,9167, SD=,60246) and

Conscientiousness Chatbot A (M=4,2917, SD=,51460) t(15)=-2,522, p = ,023. These results suggests that participants perceived the two personalities as significantly different from one another. The Mean score difference between Chatbot B and A, suggests that participants perceived all characteristics in Chatbot A to a higher degree than in Chatbot B, where the extroversion factor had the largest difference in mean score.

The two-way repeated ANOVA investigated whether an interaction effect occurred with respect to the starting condition (startbot) to investigate whether the order of chatbots participants were subjected to had any effect on the results. There was a significant main effect of the two levels of personality with respect to the characteristics F(1,14)=73,181, p<,001. There was also a significant main effect between the characteristics F(3,42)=12,960, p<,001. There was an INsignificant interaction effect of the starting condition (personality*characteristics*startbot) p=,380. See Figure 9.

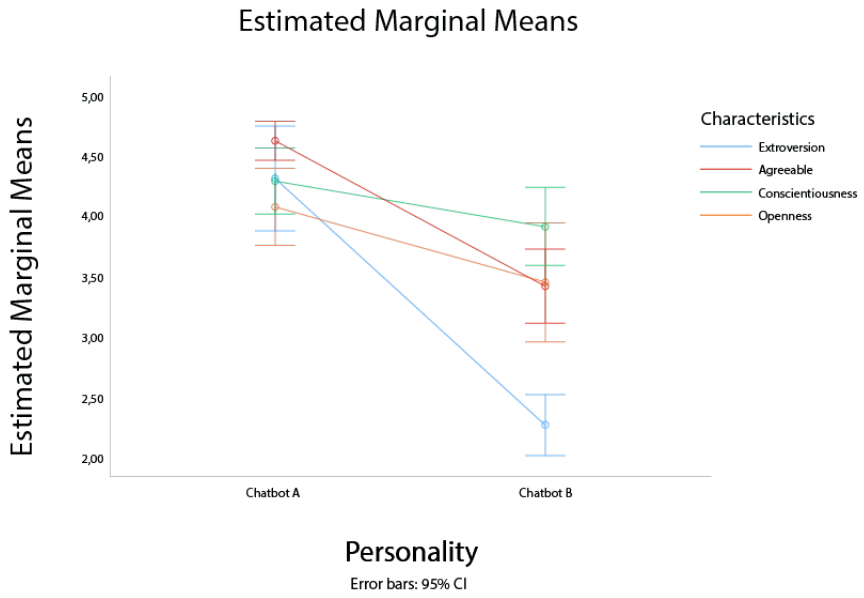


Figure 9: Estimated marginal means of characteristics for Chatbot A and Chatbot B

The data was also tested for whether there was an interaction effect of gender. This two-way repeated measures ANOVA revealed no significant interaction effect of gender (Personality*Characteristics*Gender) p=,331. See Figure 10.

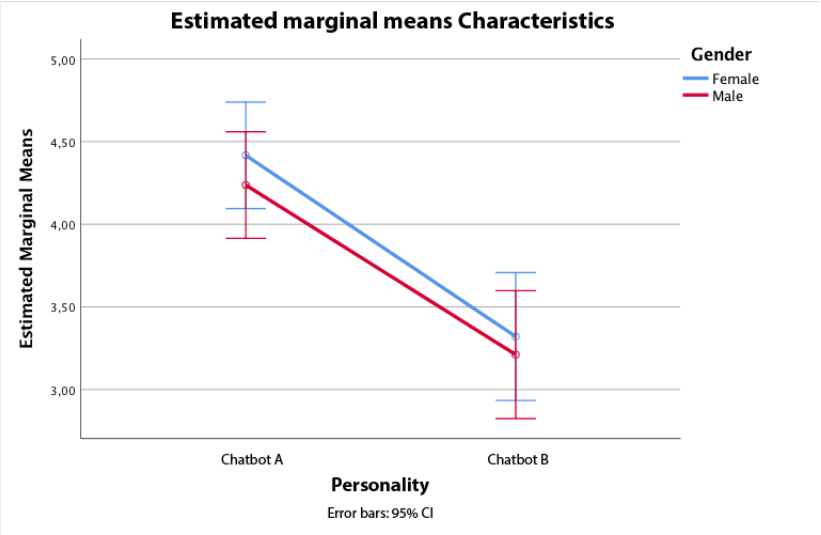


Figure 10: Estimated marginal means of Chatbot A vs Chatbot B personality on gender

4.2 Part 2 Results: AttrakDiff data analysis

The results from the AttrakDiff evaluation was also analysed by running a paired samples t-test, and a two-way ANOVA to understand whether any interaction effects occurred. The statistics will be used to test $H_1$4: *Personality affects the user experience of chatbots*, and $H_1$5: *Chatbot A will have an improved effect over Chatbot B*. The descriptive statistics of the AttrakDiff data analysis can be found in Table 7. Descriptive statistic for all word-pairs found in each factor can be found in A.8. Personality has two levels (personality, no personality) and user experience has four factors (Pragmatic Quality, Hedonic Quality Stimulation, Hedonic Quality Identity, Attractiveness).

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
PQ_A	16	4,86	6,71	5,9286	0,56424
PQ_B	16	4	6,43	5,4732	0,5964
$HQ\text{-}I_A$	16	4	6,29	5,4821	0,53165
$HQ\text{-}I_B$	16	3,57	5,43	4,7679	0,56874
$HQ\text{-}S_A$	16	5,14	6,14	5,625	0,29909
$HQ\text{-}S_B$	16	2,57	6,57	4,7768	1,19405
ATT_A	16	5,14	7	6,3482	0,44864
ATT_B	16	3,9	6,3	5,339	0,7805

Table 7: Descriptive statistics AttrakDiff results

The paired samples t-test found that there is a significant difference in the scores between Chatbot B and Chatbot A, where all four factors of the user experience showed a significant improved

effect between Chatbot B and A. Pragmatic Quality Chatbot B (M=5,4732, SD=,59649) and Pragmatic Quality Chatbot A (M=5,9286, SD=56424); t(15)=-2,152, p = ,048. Hedonic-I Quality Chatbot B (M=4,7679, SD=,56874) and Hedonic-I Quality Chatbot A (M=5,4821, SD=,53165) t(15)=-3,239, p = ,006. Hedonic-S Quality Chatbot B (M=4,7768, SD=1,19405) and Hedonic-S Quality Chatbot A (M=5,6250 SD=,29909) t(15)=-2,934, p = ,010. Attractiveness Chatbot B (M=5,339, SD=,7805) and Attractiveness Chatbot A (M=6,3482, SD=,44864) t(15)=-4,069, p = ,001. These results suggests that personality has an improved effect on the user experience of chatbots, as all four factors of user experience was scored higher for Chatbot A than Chatbot B. See Figure 11.

A two-way repeated measures ANOVA was conducted to investigate the effect the starting condition (startbot) had on the independent variable personality and the dependent variable user experience. There was a significant main effect of the two levels of personality with respect to the user experience F(1,14)=15,300, p=,002), and the starting condition (startbot) had no significant interaction effect on personality p=,847. There was a significant main effect of the user experience F(3,42)=12,264, p<,001, and the starting condition (startbot) had no significant interaction effect on user experience p=,865. There was no significant interaction effect between personality, user experience and starting condition (personality*UXscore*startbot) p=,909.

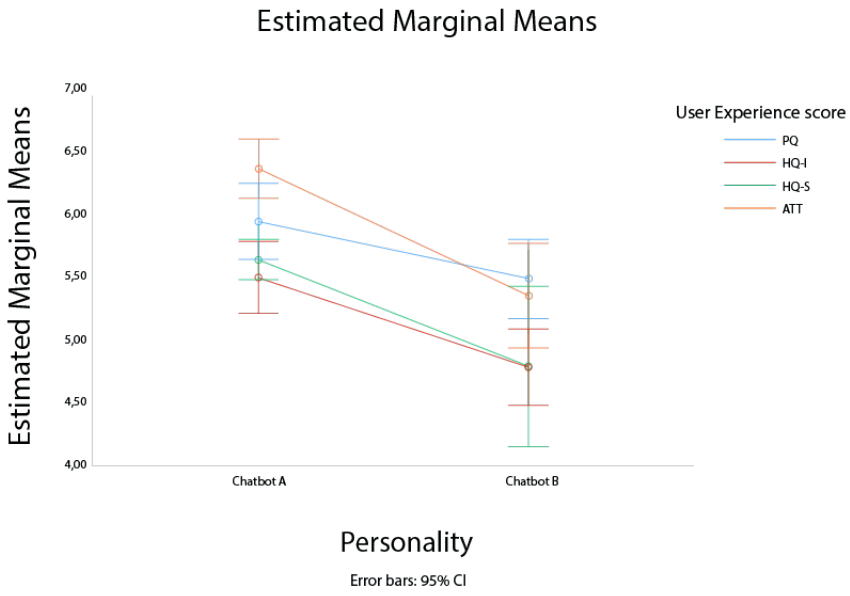


Figure 11: Estimated marginal means starting condition chatbot A user experience score

Another two-way repeated measures ANOVA was conducted to investigate whether gender had any significant interaction effect on the variables. The analysis found that there was no significant interaction effect between personality, user experience and gender (personality*UXscore*gender)

p=,436 (see Figure 12).

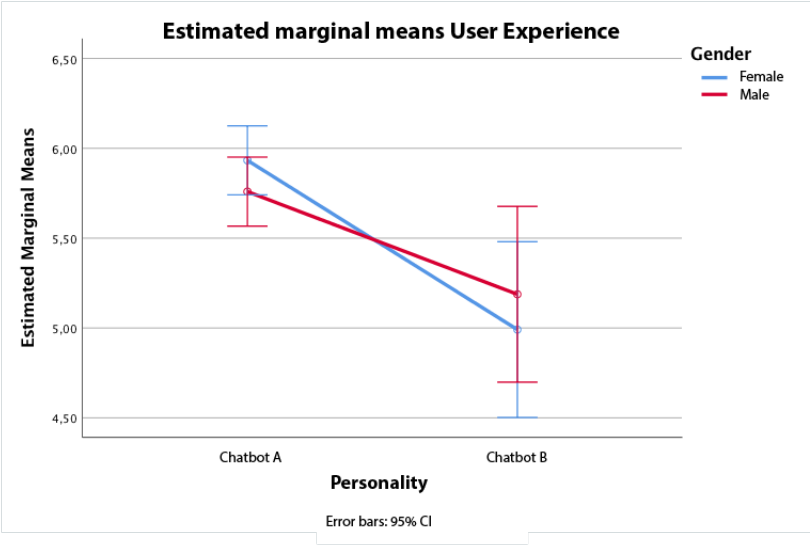


Figure 12: Estimated marginal means of Chatbot A and B personalities on gender

As shown in Figure 13, Chatbot A performed better in both hedonic and pragmatic quality than Chatbot B. The rectangles shows the confidence observed for Chatbot A and B, where Chatbot A has a smaller confidence rectangle implying that the participants were largely at one. Chatbot B however has a much larger confidence rectangle, which suggests that participant responses differed more greatly. Figure 14 shows the mean score of each user experience factor and how the two personalities scored compared to each other. The diagram shows that the difference is greater when it comes to hedonic quality-simulation and attractiveness, while the differences are less when it comes to pragmatic quality in particular. As the two personalities only shared traits found in the conscientiousness personality factor (reliable, consistent, perceptive) this result is to be expected as those traits are often found to be pragmatic qualities. Figure 15 shows the average score for each of the word-pairs for both chatbot versions.

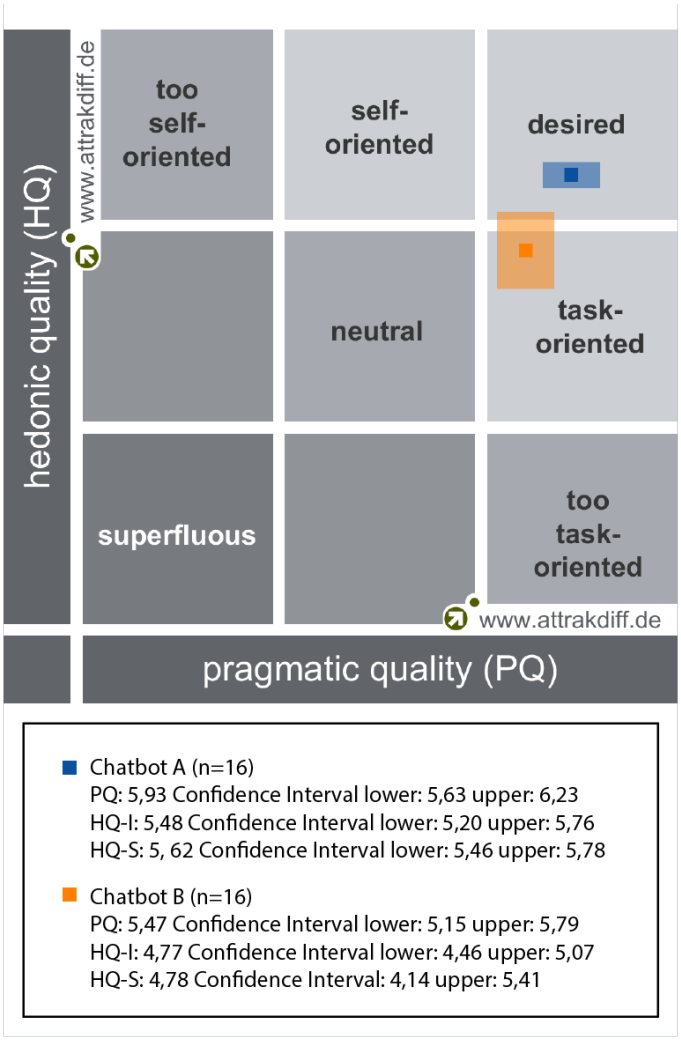


Figure 13: Results Attrakdiff

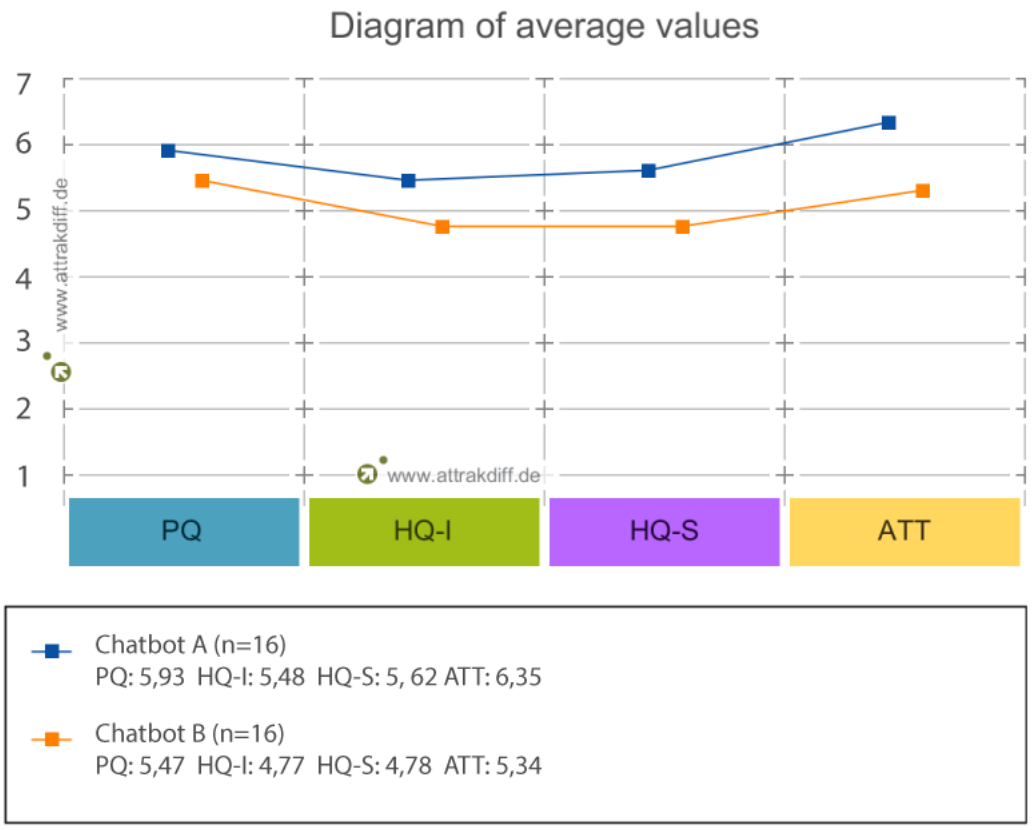


Figure 14: Diagram of average values

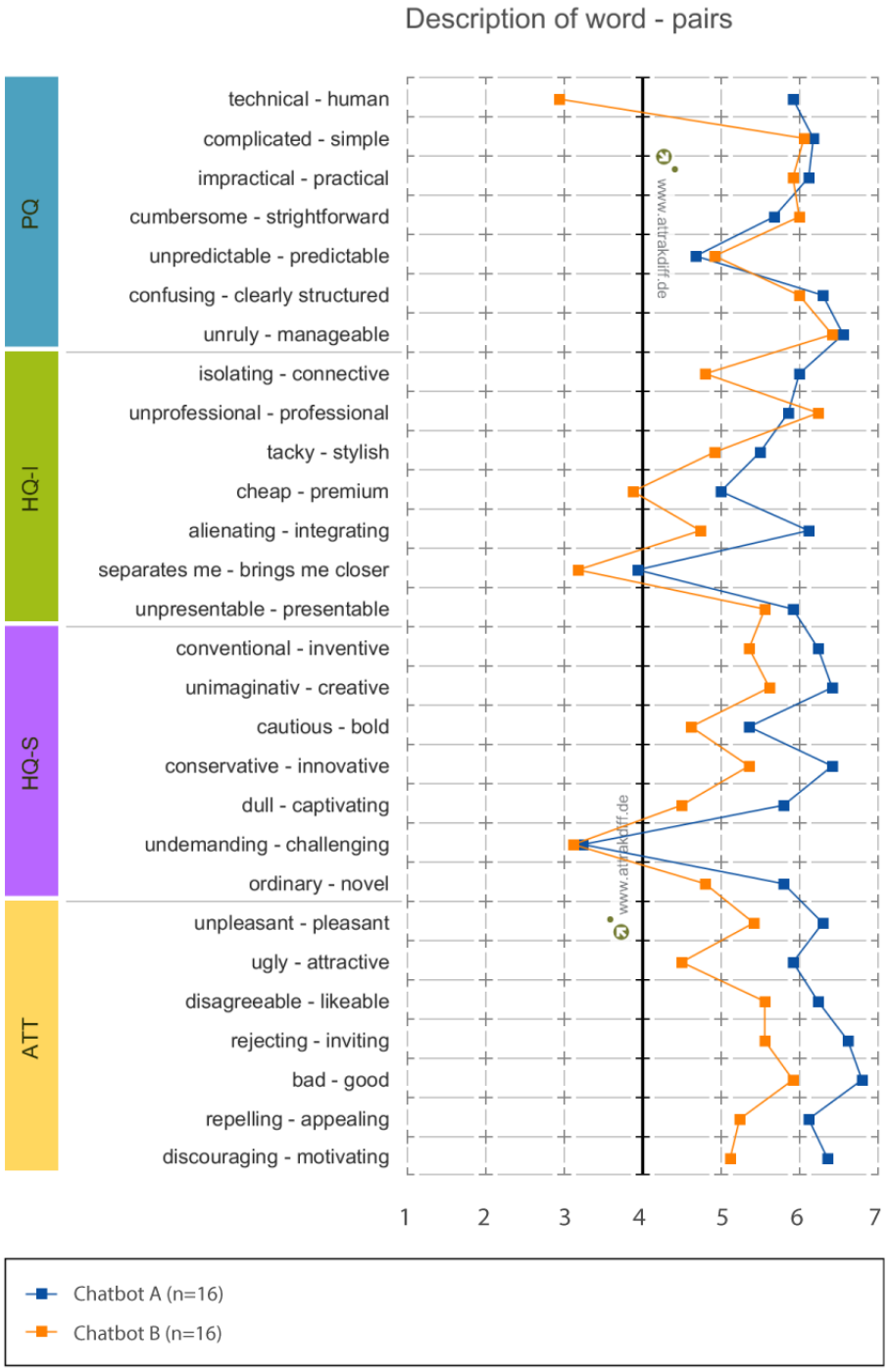


Figure 15: Description of word-pairs

5 Discussion

This chapter will first discuss the implementation of the personality framework and the experiences made throughout the design process towards the final prototype. Then the discussion will move towards the running of the experiment, recruiting of participants, and interpretations of the results. Possible confounders will be discussed along with contributions and limitations to offer insight into the process as a whole and the take-away.

5.1 Discussion: The implementation of the personality framework

The first research question asks *How can we design chatbot personalities to guide the design process of a chatbot interface?*, and the sub-questions concerns whether personality can be used as a stable pattern to guide the design process, which elements must be considered to inform the personality, and how can we ensure that the personality meets user needs and expectations. Through following the personality framework, the researcher found that the elements that must be considered to inform the personality are as follows:

1. The brand mission, goals, and values
2. A deep understanding of the users and their needs
3. The role/job of the chatbot
4. An appropriate personality model

These four elements were all used to form the personality of the chatbot, and helped narrow down the most important tasks the chatbot should perform, as well as understanding the way in which these tasks should be performed. As laid out in the methodology section, chatbots that are extensions of services provided by brands, must be consistent to the brand it represents. By having a deep understanding of the brand guidelines; tone of voice, values, mission and goals, will limit and set precise guidelines for the way in which the chatbot should behave. In addition to how it should conduct itself, it will also have to be built to understand how it should handle the different users regarding their emotions, needs, expectations, and contexts. This specific understanding not only adds to its behaviour, but also allows designers to plan for which characteristics should be present at what time. If the users are frustrated or in a hurry, the chatbot should not behave in a way that adds to their frustrations or become more time consuming. Understanding these helps write user scenarios, and will add to the knowledge of the chatbot, as different user scenarios needs different use cases. Therefore, designers have already predicted and planned for use cases that the chatbot might have to handle. Personality plays into this in regards to how it is designed to respond or act in these scenarios. The chatbot built for this thesis were built on the knowledge that its users will not respond well to a lecturing tone, the "I-know-best", and therefore took a much different

approach to providing help and tips. The last point here also consists of the role of the chatbot, the job its performing again limits the way in which it should behave. This chatbot was defined not only as an assistant, but a motivator as well. It was there to support, guide, and help, not to lecture, challenge, and tell - which might be a suitable approach for a different context and a different target audience. This shows how important it is to not choose a personality before we know 1) who the users are 2) which values it represents 3) which job it is performing! This should form the basis of the personality, and not come after-the-fact. Then you need to find an appropriate personality model.

This project chose to use the five-factor model to model the chatbot personality. This model was chosen over e.g. MBTI, HEXACO or Jung, and this is because the five factor model has been found to be suitable for multiple countries and cultures, and provides a descriptive taxonomy, in which organises natural language and the scientific traits into a single framework. It is therefore an organised and easy framework to use when wanting to build a personality. The model is created to assess persons personality types, but can very well be reversed to build a suitable personality; as done in this thesis project. Having the list of traits, organised under the five factors, was helpful to use throughout the scripting of the chatbot. It was a helpful tool when writing each use case, and plan for any conversation flow. It was also very helpful when building the personality for Chatbot B, as it offered an organised way to model an "opposite" personality. Other UX professionals have offered another view on how to model personality, which includes basing the personality on well known personalities; such as that of specific celebrities, historical figures, politicians, or fictional characters. This approach is a very creative take on writing personalities for chatbots, by combining and mixing traits from well-known personalities. For this context however, this approach would maybe have been useful in the ideation/exploratory phase of building the chatbot personality, and could provide a fun and collaborative way to design personalities in teams. There would however still be necessary to use a personality type model to organise the chosen personality after building one based on known personalities.

The personalities given to Chatbot A and Chatbot B did not include any traits from the neuroticism factor, and there is a good explanation for why this is. In character development, it is emphasised that characters should have flaws and weaknesses in order for them to be seen as realistic and to give depth to the character. The chatbot personality does not include any intended flaws or weaknesses, and this is because of its role and context. It is crucial that the chatbot is a good representative for the internal and core values of the brand. This means that it needs to conduct itself the same way as other employees are expected to conduct themselves when representing the brand. The phrase "there are no bad days in customer service" also extends to the personality of the chatbot. The chatbot does not have "bad days", it does not experience sorrow, loss, or anger, it does not worry, get tired or frustrated - and more importantly it should not display any emotions or personality traits that might be perceived negatively by the users. A chatbot in a different context, with a different role to play, might just be designed to appear flawed, and be influenced by emotions and events. However, in this context, such traits are not appropriate or desired.

5.2 Discussion: The experiment

Before discussing the results from both parts of the experiment, let us first talk about the experiment design as laid out in the methodology section in regards to how this was implemented. It was the intention when buiding the prototype, to allow participants to interact freely with the chatbot; getting to know it and exploring all the different tasks it could perform. Unfortunately this was not possible to do as it was revealed later in the design process that the machine learning that forms the AI that Chatfuel delivers, had several flaws. As explained in the methodology chapter, it is very important that the chatbot does not fail to predict a user intent, or answers the wrong question during the experiment, as this affects how users perceive the chatbot overall. For the AttrakDiff measurement tool to give a true evaluation of the user experience, the pragmatic qualities, or the usability and usefulness factors, must be met to a high standard by the chatbot. It was not the intention to evaluate the usability of the chatbot in regards to its performance, but rather allow users to test the overall experience of interacting with the chatbot. It was therefore necessary to control the conversation by limiting the participants freedom in what they could ask the chatbots during the experiment. This way made it easier to only focus on building out eight conversation flows that the researcher was confident that the chatbot could handle without fail. A few of the participants did state that it was difficult to give a true assessment of the chatbots when they had only interacted with it through a short session, and that they were not able to ask it what they wanted and test it in a true setting. This however, did not seem to affect how they rated the two personalities, as they were all certain of which they preferred and what they liked or disliked in each of the chatbots. None of the participants knew that the research aim was to investigate the effects of the personalities, and therefore a lot of the feedback were in relation to its performance: "I would have liked to try it out more before evaluating it, as I don’t know how it would perform in a real life setting". When participants were asked to provide commentary on the two chatbot versions, it was clear that the amount of interaction was enough for them to offer a clear and precise evaluation of preference towards their personalities.

The two-by-two factorial design was used to avoid an interaction effect on the results. It was assumed that participants might be more "careful" when rating the first chatbot version they interacted with, as they had no comparison to evaluate it against. Their opinion of the first chatbot might also influence how they evaluate the second chatbot. The results indicated that a majority of participants were less inclined to use the outer edges of the scales when evaluating the first chatbot, but used more often the outer edges when evaluation the second chatbot. A few of the participants stated during their second round of evaluation that they wished to go back and change their answers from the first round of evaluation as they felt they should have used more of the outer edges of the scales. This showed how preference affected their perception when comparing the second chatbot interaction with their first. If they preferred the first chatbot after interacting with the second, they wished to go back and improve their scores for the first one. The same was true if they preferred the second chatbot, as then they felt they had been "too nice" regarding the first. All participants were of course not allowed to change their initial answers.

In addition to the AttrakDiff measurement tool to assess the user experience, each participants were also asked at the end of the experiment to answer which of the two chatbots they preferred. Twelve of the sixteen participants preferred Chatbot A over Chatbot B, three males and one female preferred Chatbot B. Although they stated to prefer Chatbot B, they rated Chatbot A higher in the AttrakDiff evaluation, but they rated Chatbot B a little higher in pragmatic qualities. The reasons they gave for preferring Chatbot B, was that they expected to grow tired of Chatbot A over time. They assumed that the same answers over and over again would become irritating and tiresome over time. This provides important information regarding the need to allow chatbots to adapt and grow over time. Conversational UX is explained as designing interfaces that learns from interacting with users. This means that not only does the chatbot need to learn the preferences and remember previous information of its users; it also needs to adapt its language over time. Some of the participants who preferred Chatbot B also admitted that they would be more interested in the "assistant" features, such as generating meal plans, recipes, grocery lists, and keeping track of they weekly shopping. They were more interested in these tasks being handled efficiently and effectively, as these tasks needs to be to the point.

It is also important to note that the participants were not aware of which brand the chatbot represented. The participant's relationship with a brand could have additional effects on their overall perception and could have impacts on the user experience as well. As there was no formal collaboration with the brand in question it was not revealed to the participants as to not create misunderstandings.

5.2.1 Discussion of characteristics results

The results of the degree participants perceived the personality characteristics of both chatbots showed that users did rate all factors for Chatbot A as more present than in Chatbot B. The paired samples t-test found this difference to be significant, and the two-way repeated measures ANOVA found no interaction effect of the starting condition or gender. This information tells us that the two chatbot personalities are seen as significantly different, with a significant higher perceived presence of characteristics in Chatbot A. To be able to say that the personalities were perceived as intended, Chatbot B would have had to receive lower scores than what it did in most of the characteristics. While the mean scores for all characteristics of Chatbot A were more than 4, and therefore supports $H_1 2$, Chatbot B received higher means than expected for Agreeableness (3,4219) and Openness (3,4531). This means that we cannot fully support $H_1 3$, and would have to keep the $H_0 3$. Although for Chatbot B to be perceived as intended would needed a lower score than what it was given, it was still rated as lower in the perceived characteristics compared to Chatbot A. Therefore, as it was significantly different and lower in scores compared to Chatbot A, it was suitable to use for comparison when testing the user experience in part two of the experiment.

Females and males both rated characteristics in Chatbot A as more present than the same characteristics in Chatbot B as can be seen in Figure 10. Female means for Chatbot A and Chatbot B were slightly higher scored than the means for males. The male mean score for Extroversion for Chatbot

A was higher than the mean score for females, and males scored Chatbot B higher in agreeableness than females. Apart from those small differences, each characteristic for each chatbot were rated more or less consistently between males and females. The conscientiousness factor was assumed to be rated more or less equal for both chatbots, and the mean scores differed by 0,375 showing that this factor was perceived almost the same for each chatbot by all participants. It is interesting though that Chatbot A was seen as somewhat more conscientious than Chatbot B, indicating that users found that version to be slightly more perceptive, reliant, and consistent than the other.

As will be explained further in the *Limitations* section, both personalities should have been tested in several iterations before conducting the second part of the experiment. This would have helped understanding exactly what the perceived personalities would be before testing the effect the personalities had on the user experience. However, the pilot testing were enough to satisfy that the two personalities were perceived as opposites to an extent, and the results confirmed this.

5.2.2 Discussion of user experience results

The results from the statistical analysis of the AttrakDiff data to assess the user experience, found a significant difference between the ratings of the two personalities, and a significant improved effect of Chatbot A over Chatbot B. The opposing adjectives used to evaluate the user experience in the AttrakDiff form (see Figure 15), shows clearly how both chatbot versions were rated compared to one another and where the differences are largest. The ratings for each chatbot follows each other well on most of the word-pairs, while the biggest difference is seen in the word-pair: human-technical. This indicates that the one trait that truly appeared as the biggest difference between the two chatbots was the degree of humanness displayed in both. For the majority of word-pairs Chatbot A scores higher than Chatbot B. The exceptions includes mainly word-pairs in the pragmatic quality: cumbersome-straightforward, unpredictable-predictable, and one in hedonic quality-identity: unprofessional-professional, in which Chatbot B was scored higher. This suggests that a chatbot which has a personality that appears less human, more technical, impersonal or machine-like, are perceived as being higher in pragmatic qualities than a chatbot with a more human personality. The mean score for professional-unprofessional for Chatbot A was 5,875 and Chatbot B mean score was 6,25. This shows that Chatbot B was perceived as conducting itself to a higher professional standard than Chatbot A, ever so slightly. While the overall impression was that Chatbot A was more attractive than Chatbot B, and offered a more enjoyable user experience.

Although there was found no significant interaction effect caused by the starting condition, there are some minor interesting observations. As can be seen in Table 8, the differences between the ratings of the two groups are down to the decimals, except hedonic quality stimulation for Chatbot B, and as stated not significant. The differences between the two groups do show that Chatbot B were slightly better perceived when not being compared to Chatbot A. This indicates that Chatbot A had more of an impact on the rating of Chatbot B, than Chatbot B had on Chatbot A.

	Group 1		Group 2	
	Round 1 Chatbot A	Round 2 Chatbot B	Round 1 Chatbot B	Round 2 Chatbot A
PQ:	5,9107	5,5536	5,375	5,9464
HQ-I:	5,4464	4,6607	4,9821	5,5179
HQ-S:	5,75	4,25	5,3929	5,5
ATT:	6,4107	5,0357	5,8214	6,2857

Table 8: Starting condition and mean scores for each group AttrakDiff results

There was not found any significant differences between gender, in regards to the results of the AttrakDiff evaluation. Although it was assumed that men would prefer chatbot B over Chatbot A, the results from the evaluation found no evidence to support this. Males did however rate Chatbot B higher overall than females (see Figure 12). As mentioned earlier, three males and one female preferred Chatbot B over Chatbot A when asked to state their preferred version. This would suggest that it is still necessary and important to collect information regarding preference in addition to measuring the user experience, through a measurement tool such as AttrakDiff. As the evaluation might suggest that one version is better than the other, while this not being necessarily accurate in regards to preference.

The participants were not instructed to assess the personality of the chatbot when filling out the AttrakDiff form. Instead they were asked to evaluate the chatbot as a product in its entirety. A few of the word-pairs assessed in the AttrakDiff measurement tool proved somewhat difficult to define by participants, as they were easy to understand if evaluating an interactive product such as a website or application, but more difficult when evaluating a chatbot agent. A selection of the word-pairs were seen as positive for a product, but negative for a personality e.g. *undemanding-challenging*, were participants found *challenging* to be a negative trait to describe the agent while for the assessment of a product it was seen as a positive trait. Another word-pair that confused participants were *separates me from people - brings me closer to people*, as a lot of participants noted that talking to a robot would in the long run separate them from people rather than bringing them closer. This shows that the AttrakDiff, while researched and tested to be a reliable tool for measuring attractiveness of interactive products, might need some modifications to be used to measure attractiveness of Chatbot systems. Also, all participants in this study were Norwegian, but the experiment was run in English. The opposing adjectives could have been misunderstood or had a slightly different meaning in Norwegian, which could have impacted the results.

5.2.3 Possible confounders

This section will discuss the possible confounding variables that might have impacted the results of the experiment.

Because of the Facebook and Cambridge Analytica scandal during the spring of 2018; Facebook stopped reviewing any third party applications running through their platform. This resulted in the Chatbot B prototype not being able to be published through the messenger platform in time for the experiments. Chatbot A was able to be published through an existing application already reviewed

by Facebook before the scandal, however the name of the chatbot could not be changed in time of the experiment. This had an impact on the appearance of the interface during the experiment. Chatbot B did not display a profile picture, but instead the Chatfuel logo, and the name was displayed as "Chatfuel-test your bot", as this had to be tested through the unpublished test service provided by Chatfuel, rather than being run through its own page and account. Chatbot A however was run through its own page and account, but the name of the chatbot belonged to a previous version of a chatbot displaying the wrong name. The absence of an image and name for Chatbot B could have impacted how users perceived this chatbot. Some unconsciously referred to this chatbot as a "he" rather than "she" for Chatbot A - even though all participants were instructed before beginning the tests that they are about to test two versions of the chatbot "Bella". The cards each participants used during the test had Bella's face on them in order to remind them of who they are talking to.

Another possible confounder was one that the researcher failed to recognise before running the experiment. This was the recipes suggested by the chatbot. A few of the test participants told the researcher after the experiment had finished, that they rated e.g. the word-pair *inventive-conventional* higher because the recipes in themselves were inventive - or they rated the characteristic helpful lower because they did not like the recipes suggested to them.

5.3 Contributions

Earlier in the report it was stated that the planned contributions would consist of the personality framework and the results regarding the quantitative measurement of personality and its impact on the user experience of chatbot interfaces. The personality framework has been laid out in detail in the methodology section to allow others to replicate, modify and improve it. This framework forms the foundation to guide the design process of chatbot interfaces, and can be used for any kind of chatbot and domain. However, the personality framework might be more suitable for chatbots that represents brands. The experiment contributed significant data that suggests that personality indeed has an effect on the user experience of chatbots and that the personality modelled using the personality framework showed an improved effect on the user experience. In addition to this, this thesis offers two methods to collect data and evaluate the personality modelled by the personality framework and the chatbot product as a whole in regards to the user experience. Both chatbot prototypes have been preserved and can be made available if necessary. If wanting to explore the chatbot interactive mock-ups, follow this link: https://xd.adobe.com/view/ea6e993d-5087-4425-7892-c957909c9c4b-9c5f/?fullscreen

5.4 Limitations

The implementation of the personality framework for this specific thesis project lacked enough iterations of testing. Time constraints affected how many times the modelled personality could be tweaked and improved through user testing, also the access to users made this difficult as testing the personality should be done by an appropriate sample belonging to the target group. Testing this

on the participants throughout the design process would have biased them for the final experiment. Therefore initial testing of the personality, conducted through the design process used any person on hand that was willing to test the personalities. It would therefore have been more accurate if it could have been more thoroughly tested before testing the user experience of the personalities.

The use of Chatfuel to build the chatbot prototypes provided some challenges, and limitations. Chatfuel was chosen because it allowed for more dynamic conversation flows than other tools, and provided a simple interface to organise the different intents and user flows. It was also easy to hook up to Facebook's Messenger platform for testing. The AI provided by chatfuel however were not as sophisticated as believed when choosing it as the bot builder platform. Previously, other platforms for building bots have been tested that provided much better AI. It appeared after a while that the machine learning Chatfuel uses were extremely flawed. It did not e.g. ignore stop words, and worked poorly to extract keywords to separate two similar user intents, it did not use stemming, nor allowed creating a registry of synonyms, it did not correct simple typos, and it differentiated between upper and lower case letters. Only after writing more than 300 unique phrases for each intent was it discovered that the machine learning failed to improve the AI to a desirable standard. Had this been noticed earlier, the chatbots would most likely been built using a different platform or another third party application that provides better AI and NLP capabilities. When the flawed AI of Chatfuel had been discovered, it was too late in the process to move to another platform, as Chatfuel did not allow any form of migration or export. Chatfuel does however provide, as stated earlier, a very organised and easy to use interface and they allowed for flexible linking of chatbot skills to create varied conversations, that other platforms lack.

The sample size consisted of 16 participants, 8 males and 8 females, and all participants belonged to the defined target group. The two-by-two factorial design implemented in the experiment needed a larger sample than other experiment designs. However, by subjecting all participants to all conditions, rather than having two separate groups made the need for a larger sample less crucial. All participants were recruited through convenience sampling, which means that all participants were familiar to some extent with the researcher. This familiarity could have affected the results to some degree.

Both the chatbot interface, tasks, and measurement tools were in English. All participants were Norwegian and proficient in the English language. However, there could easily have been confusion or misunderstandings regarding the true definition of the opposing adjectives and characteristics used in the data collection forms. All participants were able to ask for clarification or translations, but translations can potentially be misleading. Many bot builder tools does not yet support the Norwegian language well enough for the prototype to be of the appropriate standard, and a translation of the AttrakDiff tool in Norwegian is not available. It could have been translated, but would not not have been empirically tested in order to see if it provides the same results as the original.

6 Conclusion

This thesis aimed to investigate how we can design user centred chatbot personalities and whether personality matters for the user experience of chatbots. The latter has been a long standing assumption, and based on findings related to other interfaces than chatbots and CAs. While findings regarding traditional web and app user interfaces have found that personality is an important factor for how users perceive the interface, no similar study has yet investigated whether the same is true for chatbots. If personality does not matter, then we do not need to spend time, money and effort to create personalities for chatbots, and can spend more time on further developing and improving the AI and NLP capabilities, and offer a seamless experience with great and effective task handling. However, the findings presented in this thesis suggests that a human personality not only improves the user experience compared to chatbots without a human personality, it also impacts a user's perception overall. Even though Chatbot A and Chatbot B provided the same service, just as effectively and efficiently, Chatbot A was still rated higher than Chatbot B in the pragmatic qualities. This shows that an overall great user experience also improves how users perceive other qualities of the chatbot that are not in reality better than other versions. Building a framework to design personalities for conversational agents would be useless if personality did not have an improved effect on the user experience. This is why this project is twofold 1) building, implementing, and testing a personality framework for chatbot interfaces, 2) testing whether personality improves the user experience of chatbot interfaces.

The answer to Research Question 1, and its sub-questions, is the personality framework presented in this thesis; the personality framework includes the defined elements to be considered to inform the personality, it also includes user centred methods to ensure that the personality meets user needs and expectations. The researcher's experience implementing the framework found that it does provide a stable pattern to guide the design process of chatbot interfaces. By always referring the choices of the design process back to the personality, helped write use cases and the conversation flow for the chatbot. It helped form the answers to specific use-cases, such as handling a frustrated user or a happy one, and also to plan for how the chatbot needed to handle specific scenarios. By knowing how the chatbot should behave, we can also to a greater certainty predict what the course of the conversation most likely will be. Knowing that the personality is appropriate to the users also made it easier to understand which behaviours that would not only be inconsistent with the personality, but also which behaviours that would be perceived more negatively by the users. In addition to this, the framework also provided necessary information regarding the most suitable role the chatbot could take on to support users towards their goals. The chatbot role dictated which characteristics that were suitable, and helped understand the most valuable tasks and information it could provide users with.

The findings from the statistical analysis of the personality characteristics evaluation found a significant difference between the personality of Chatbot A and the personality of Chatbot B. There was no significant interaction effect between the starting condition and gender. The results found that Chatbot A was rated higher on all factors than Chatbot B. We can therefore keep our research hypothesis $H_1 1$ that users will perceive the personality of Chatbot A as different to the personality of Chatbot B. As for $H_1 2$ and $H_1 3$ in that users will perceive both Chatbot A as intended and Chatbot B as intended, we can keep $H_1 2$ as it received a very high score on all factors. For $H_1 3$ however, the researcher expected a lower score for Chatbot B in the two factors: Openness and Agreeableness. The extroversion factor was perceived as intended as Chatbot B received a very low mean score of 2,2708. The conscientious factor was assumed to be rated more or less equal for both chatbots, and the mean scores differed by 0,375, confirming to some extent with the researcher's assumptions. We can conclude that Chatbot A was perceived as intended with high scores on all factors, while Chatbot B were perceived as intended in the Extroversion and Conscientiousness factors, and not for the Agreeable and Openness factors as the intention was for users to perceive them to a much lower degree.

The findings from the statistical analysis of the results of the AttrakDiff data found there to be a significant difference between the pragmatic quality, hedonic quality and attractiveness of Chatbot A and Chatbot B. The results of the paired-samples t-test found that there was a significant improved effect on the user experience of Chatbot A compared to Chatbot B. Chatbot A was rated to a higher degree in all factors, pragmatic quality, hedonic quality and attractiveness, over Chatbot B. The two-way repeated measures ANOVA found a significant main effect between the two personalities with respect to the user experience. There was no significant interaction effect found between the starting condition with respect to the personality or user experience score, nor was there found a significant interaction effect of gender. Therefore, based on these results we can keep our research hypothesis $H_1 4$ and $H_1 5$, and we can conclude that personality does affect the user experience of chatbot interfaces and that Chatbot A's personality has an improved effect on the user experience compared to the personality of Chatbot B. This answers our second research questions that chatbots with a defined personality improves the user experience of chatbot interfaces.

6.1 Future Research

For future research it is recommended to test the effects of the personality over a longer period of time. Participants commented that they could grow tired of the personality if being subjected to it over a longer period, especially if some answers were repeated over and over again. Therefore understanding the long-term effects is important to inform how we can design the chatbot to behave over time, and how it should behave dependent on the preferences of the specific user is important. Allowing the chatbot and the personality to learn and adapt to the specific user, is assumed to be preferred by users. By testing the modelled personalities over a longer period would add more to the understanding of the effects of personality for CUIs.

IBM and others are currently using tone analysing to read the mood of the author of texts. This

could be chat messages, social media post, articles etc. By implementing the tone analyser in a chatbot, will allow it to understand the mood of the user throughout the interaction. Adapting the chatbot responses, and the dynamics of its personality and behaviour, to the mood of the user can be benefited by the framework presented in this thesis. The personality traits can help dictate whether the variations of answers still are consistent with the chatbot personality, and can help plan and prepare for the different moods users bring to the interaction. Exploring this further to improve the personality framework can help benefit the user experience of chatbots.

Another aspect for future research is to adapt the AttrakDiff measurement tool to become more suitable for conversational user interfaces. In addition to this, having a tool to be able to assess and evaluate the user experience when the CA grows and learns in regards to the specific user, will become important as chatbots are rapidly becoming more and more popular and intelligent.

As for ensuring that users perceive the personality as intended by the designers of the CA, it is recommended to use the Agree! framework developed by Callejas et al. (2014). This framework is built to evaluate a personality in iterations during the design process. Each interaction involves a change in attitude (e.g. unfriendly, neutral, friendly), and the framework assesses whether participants are in agreement in regards to the perceived personality of the CA. The Agree! tool "computes a very wide range of numerical coefficients that measure the similarity of the perceived and target personalities, the impact of the user personalities in their perceptions, and also the level of agreement between your users" (Callejas et al. 2014).The framework can be used to run a score-based evaluation or a tag-based evaluation. The score-based evaluation computes the similarity between the target and perceived personalities of different personality traits, e.g. the five factors. The tag-based evaluation allows participants to choose between traits/tags to describe the personality from a predefined set of traits/tags.

Another aspect that has received little attention in the research community in regards to how humans perceive conversational agents, is the effects of gender. The gender of the agent has been found to impact the nature of the conversation, and also what kind of attention the CA receives from the different users. Understanding these effects, and how we can use this knowledge requires more attention from the research community.

Bibliography

Ackerman, C. (2017), 'The big five personality theory: The 5 factor model explained', *Positive Psychology Program* .
URL: *https://positivepsychologyprogram.com/big-five-personality-theory/*

AttrakDiff.de (2013), 'Attrakdiff scientific background'.
URL: *http://attrakdiff.de/sience-en.html*

Augello, A., Gambino, O., Cannella, V., Pirrone, R., Gaglio, S. & Pilato, G. (2011), An emotional talking head for a humoristic chatbot, *in* 'Applications of Digital Signal Processing', InTech.

Balzarotti, S., Piccini, L., Andreoni, G. & Ciceri, R. (2014), '"i know that you know how i feel": Behavioral and physiological signals demonstrate emotional attunement while interacting with a computer simulating emotional intelligence', *Journal of Nonverbal Behavior* **38**(3), 283–299.
URL: *https://doi.org/10.1007/s10919-014-0180-6*

Benyon, D. (2014), *Designing interactive systems : a comprehensive guide to HCI and interaction design*, 3rd ed. edn, Pearson, Harlow.

Beres, D. (2015), 'Microsoft's cortana is like siri with a human personality'.
URL: *https://www.huffingtonpost.com/entry/microsofts-cortana-is-like-siri-with-a-human-personality_us_55b7be94e4b0a13f9d1a685a*

Beun, R.-J., De Vos, E. & Witteman, C. (2003), Embodied conversational agents: effects on memory performance and anthropomorphisation, *in* 'IVA', Springer, pp. 315–319.

Boutin, P. (2017), 'Why most chatbots fail'.
URL: *https://chatbotsmagazine.com/why-most-chatbots-fail-1c085b74d6ad*

Brahnam, S. & De Angeli, A. (2012), 'Gender affordances of conversational agents', *Interacting with Computers* **24**(3), 139–153.

Burge, J. (2016), 'Qualities of a great assistant'.
URL: *https://officedynamics.com/qualities-of-a-great-assistant/*

Callejas, Z., Griol, D. & López-Cózar, R. (2014), 'A framework for the assessment of synthetic personalities according to user perception', *International Journal of Human-Computer Studies* **72**(7), 567 – 583.
URL: *http://www.sciencedirect.com/science/article/pii/S1071581914000263*

Callejas, Z., López-Cózar, R., Ábalos, N. & Griol, D. (2011), 'Affective conversational agents: The role of personality and emotion in spoken interactions', *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices: Techniques and Effective Practices* pp. 203–223.

Cialdini, R. B. (2007), *Influence: The Psychology of Persuasion*, HarperCollins Publishers.

Cohen, M. H., Cohen, M. H., Giangola, J. P. & Balogh, J. (2004), *Voice user interface design*, Addison-Wesley Professional.

Courage, C., Baxter, K. & Caine, K. (2015), *Understanding Your Users: A Practical Guide to User Research Methods*, second edition. edn, United States: Morgan Kaufmann Publishers Inc.

Cummings, H. (2017), 'Finding your brand's tone of voice: How to shape a tone of voice'.
URL: *https://www.distilled.net/tone-of-voice/*

Dautenhahn, K., Ogden, B. & Quick, T. (2002), 'From embodied to socially embedded agents–implications for interaction-aware robots', *Cognitive Systems Research* **3**(3), 397–428.

De Angeli, A. & Brahnam, S. (2008), 'I hate you! disinhibition with virtual partners', *Interacting With Computers* **20**(3), 302–310.

Delin, J. (2005), 'Brand tone of voice: a linguistic analysis of brand positions.', *Journal of Applied Linguistics* **2**(1), 1 – 44.

Dictionary, O. E. (n.d.), '"anthropomorphism"'.
URL: *http://www.oed.com/view/Entry/8449?redirectedFrom=anthropomorphismamp;*

Dirk, H. (2003), Talking head says cheese, *in* 'Humor as an impetus for Embodied Conversational Agent Research. CHI-2003 WorkShop: Humor Modeling In the Interface'.

Disalvo, C., Gemperle, F., Forlizzi, J. & Kiesler, S. (2002), 'All robots are not created equal: the design and perception of humanoid robot heads', pp. 321–326.

Dybala, P., Ptaszynski, M., Rzepka, R. & Araki, K. (2009), Humoroids: conversational agents that induce positive emotions with humor, *in* 'Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2', International Foundation for Autonomous Agents and Multiagent Systems, pp. 1171–1172.

Epley, N., Waytz, A. & Cacioppo, J. T. (2007), 'On seeing human: a three-factor theory of anthropomorphism.', *Psychological review* **114**(4), 864.

Fogg, B. J. (2002), *Persuasive Technology: Using Computers to Change What We Think and Do*, Interactive Technologies, Elsevier Science.

Forrester (2017), 'Chatbots are transforming marketing'.
URL: *https://www.forrester.com/report/Chatbots+Are+Transforming+Marketing/-/E-RES136771*

Gould, J. D. & Lewis, C. (1985), 'Designing for usability: key principles and what designers think', *Commun. ACM* **28**(3), 300–311.

Griol, D., Molina, J. M. & Callejas, Z. (2015), Towards emotionally sensitive conversational interfaces for e-therapy, *in* 'Artificial Computation in Biology and Medicine', Springer International Publishing, pp. 498–507.

Griol, D., Molina, J. & Sanchis, A. (2017), 'Integration of context-aware conversational interfaces to develop practical applications for mobile devices', *Journal Of Ambient Intelligence And Smart Environments* **9**(5), 561–577.

Haraldsen, I. (2011), '25 prosent rett i søpla'.
URL: *https://forskning.no/mat-landbruk-forbruk-hus-og-hjem/2011/09/25-prosent-rett-i-sopla*

Hassenzahl, M. (2001), 'The effect of perceived hedonic quality on product appealingness', **13**, 481–499.

Hassenzahl, M., Platz, A., Burmester, M. & Lehner, K. (2000), Hedonic and ergonomic quality aspects determine a software's appeal, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', CHI '00, ACM, New York, NY, USA, pp. 201–208.
URL: *http://doi.acm.org/10.1145/332040.332432*

Hendrich, S. (2017), 'I gave cortana her personality'.
URL: *https://blogs.microsoft.com/jobs/story-library/i-gave-cortana-her-personality/*

Holtgraves, T. M., Ross, S. J., Weywadt, C. R. & Han, T. L. (2007), 'Perceiving artificial social agents', *Computers in Human Behavior* **23**(5), 2163–2174.

IDEO.org (2015), *The Field Guide To Human-Centred Design*, first edition. edn, Canada: Design Kit.

Inbar, O. & Meyer, J. (2015), Manners matter: Trust in robotic peacekeepers, *in* 'Proceedings of the Human Factors and Ergonomics Society Annual Meeting', Vol. 59, SAGE Publications Sage CA: Los Angeles, CA, pp. 185–189.

ISO (2010), Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems, Standard, International Organization for Standardization, Geneva, CH.

John, O. P. & Srivastava, S. (1999), 'The big five trait taxonomy: History, measurement, and theoretical perspectives.', *Handbook of personality: Theory and research, 2nd ed.* pp. 102–138.

Jung, C. G. (1923), 'Psychological types: or the psychology of individuation.'.

JuniperResearch (2017), 'Chatbot infographic key statistics'.
URL: *https://www.juniperresearch.com/resources/infographics/chatbots-infographic-key-statistics-2017*

Kothari, A., Zyane, R. & Hoover, J. (2016), 'Chatbots for ecommerce: Learn how to build a virtual shopping assistant', *Chatbots for eCommerce* .

Kulms, P., Krämer, N. C., Gratch, J. & Kang, S.-H. (2011), 'It's in their eyes: A study on female and male virtual humans' gaze', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **6895**, 80–92.

Lebowitz, S. (2016), 'The 'big 5' personality traits could predict who will and won't become a leader.'.
URL: *http://www.businessinsider.com/big-five-personality-traits-predict-leadership-2016-12*

Lee, E.-J. (2010), 'The more humanlike, the better? how speech type and users' cognitive style affect social responses to computers', *Computers in Human Behavior* **26**(4), 665–672.

Lee, J. D. & See, K. A. (2004), 'Trust in automation: Designing for appropriate reliance', *Human Factors* **46**(1), 50–80.
URL: *http://journals.sagepub.com/doi/abs/10.1518/hfes.46.1.50*$_3$0392

Lemon, O., Pietquin, O. & service), S. O. (2012), *Data-Driven Methods for Adaptive Spoken Dialogue Systems : Computational Learning for Conversational Interfaces*, Springer New York : Imprint: Springer.

Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B. & Bhogal, R. (1997), 'The persona effect: affective impact of animated pedagogical agents', pp. 359–366.

Lipcamon, J. (2013), '12 traits of good leaders who inspire and motivate'.
URL: *http://www.diagnosticimaging.com/practice-management/12-traits-good-leaders-who-inspire-and-motivate*

Mauldin, M. L. (1994), Chatterbots, tinymuds, and the turing test: Entering the loebner prize competition, *in* 'AAAI', Vol. 94, pp. 16–21.

McCrae, R. R. & Costa, P. T. (1987), 'Validation of the five-factor model of personality across instruments and observers.', *Journal of personality and social psychology* **52**(1), 81.

McTear, M., Callejas, Z. & Griol, D. (2016*a*), Affective conversational interfaces, *in* 'The Conversational Interface: Talking to Smart Devices', Springer International Publishing, pp. 329–357.
URL: *https://doi.org/10.1007/978-3-319-32967-3*$_1$5

McTear, M., Callejas, Z. & Griol, D. (2016*b*), Conversational interfaces: Past and present, *in* 'The Conversational Interface', Springer, pp. 51–72.

Mencía, B. L., Pardo, D. D., Trapote, A. H. & Gómez, L. A. H. (2012), 'Embodied conversational agents in interactive applications for children with special educational needs', *Technologies for Inclusive Education: Beyond Traditional Integration Approaches: Beyond Traditional Integration Approaches* p. 59.

Meyer, J., Miller, C., Hancock, P., de Visser, E. J. & Dorneich, M. (2016), 'Politeness in machine-human and human-human interaction', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **60**(1), 279–283.
URL: *http://journals.sagepub.com/doi/abs/10.1177/1541931213601064*

Meyer, K. (2016), 'The four dimensions of tone of voice'.
URL: *https://www.nngroup.com/articles/tone-of-voice-dimensions/*

Mori, M. (1970), 'The uncanny valley', *Energy* **7**(4), 33–35.

Myers, I. & Myers, P. (2010), *Gifts differing: Understanding personality type*, Nicholas Brealey Publishing.

Nass, C. & Moon, Y. (2000), 'Machines and mindlessness: Social responses to computers', *Journal of social issues* **56**(1), 81–103.

Norman, D. A. (2007), *Emotional design : why we love (or hate) everyday things*, Basic Books, New York.

Orf, D. (2017), 'Facebook chatbots are frustrating and useless'.
URL: *https://gizmodo.com/facebook-messenger-chatbots-are-more-frustrating-than-h-1770732045*

Pavlus, J. (2016), 'The next phase of ux: Designing chatbot personalities'.
URL: *https://www.fastcodesign.com/3054934/the-next-phase-of-ux-designing-chatbot-personalities*

Pearl, C. (2017), *Designing Voice User Interfaces: Principles of Conversational Experiences*, O'Reilly.

Piltch, A. (2017), 'Talk is cheap: Why chatbots will always be a waste of time'.
URL: *https://www.tomsguide.com/us/chatbots-waste-our-time,news-22562.html*

Prada, R., Vala, M., Paiva, A., Hook, K. & Bullock, A. (2003), Fantasya–the duel of emotions, *in* 'International Workshop on Intelligent Virtual Agents', Springer, pp. 62–66.

Reeves, B. & Nass, C. (1996), *The media equation : how people treat computers, television and new media like real people and places*, Cambridge University Press, New York.

Schroeder, J. & Epley, N. (2016), 'Mistaking minds and machines: How speech affects dehumanization and anthropomorphism', *Journal of Experimental Psychology: General* .

Stern, A. (2003), 'Creating emotional relationships with virtual characters; from: Emotions in humans and artifacts, eds. r. trappl, p. petta, and s. payr'.

Stokke, O. P. B. (2017), 'Så dumme er de norske robotene'.
URL: *https://www.dinside.no/okonomi/sa-dumme-er-de-norske-robotene/68798600*

Terada, K., Jing, L. & Yamada, S. (2015), Effects of agent appearance on customer buying motivations on online shopping sites, *in* 'Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems', ACM, Seoul, Republic of Korea, pp. 929–934.

Toegel, G. & Barsoux, J.-L. (2012), 'Women leaders: The gender trap', *The European Business Review* pp. 1–10.

Turing, A. M. (1950), 'Computing machinery and intelligence', *Mind* **59**(236), 433–460.

Vala, M., Blanco, G. & Paiva, A. (2011), 'Providing gender to embodied conversational agents', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **6895**, 148–154.

Vincent, J. (2017), 'Chatbots are dumb, but wait until they learn how to negotiate for you'.
URL: *https://www.theverge.com/2017/6/14/15799068/chatbot-negotiations-ai-facebook-fair*

Weizenbaum, J. (1966), 'Eliza—a computer program for the study of natural language communication between man and machine', *Commun. ACM* **9**(1), 36–45.

Xiao, H., Reid, D., Marriott, A. & Gulland, E. (2005), 'An adaptive personality model for ecas', *Affective Computing And Intelligent Interaction, Proceedings* **3784**, 637–645.

Zimmerman, J., Ayoob, E., Forlizzi, J. & McQuaid, M. (2005), 'Putting a face on embodied interface agents'.

A Appendices

Interview Guide

Interview aim:

The aim of these interviews is to investigate the understanding and challenges young adults (in particular young parents) have regarding dietary requirements and activity levels of themselves and their children/future children. This information will be used to pinpoint what they find most challenging about making sure they eat healthy, and what they think about ensuring healthy diets for their children/future children. Their understanding of available information about this issue and which tools they find necessary to help guide them.

Questions these interviews will help answer:

- How do parents view their family's diets and activity levels?
- To what extent are they concerned about eating healthy?
- What challenges do they face regarding their own and/or children's nutrition?
- Have they tried to change their diets in the past and were they successful?
- Have or do they use any tools to help with nutrition or activity?
- What are their primary motivation for eating healthy?
- What is most important: activity or healthy diet?
- Concerns and habits regarding food waste

Participant selection:

The participants recruited for this interview are all between the ages of 25 to 40, aiming at young couples living together, and/or are parents with young children. This target group usually have hectic days where healthy eating and activity can be difficult to maintain, and they are also, or about to be, in charge of their children's diet and activity levels as well. As learning good habits starts when we are children, parents have a major impact in regards to teaching children the right habits. Therefore, to reach the brand's goals of increased consumption of fruit and veg, and have a healthier lifestyle, this is an important target group to help reach these goals. In regards to the brand's other goals related to the environment and sustainability, people in the age group 25-39 waste more food than other age groups, and in particular families with small children waste more because they have hectic schedules and not enough time to plan ahead.

Demographic Data:

Participant nr: _____

Gender: _____

Age: _____

Marital status: _____

Diet: _____

Activity level: _____

Children: YES / NO / Expecting

How many?: _____

Ages: _____

Semi Structured Interview Questions:

Introduction:

The aim of this interview is to investigate the understanding and challenges young adults (in particular young parents) have regarding dietary requirements and activity levels for themselves and their children/future children. This information will be used to pinpoint what they find most challenging about making sure they eat healthy and provide their children with healthy, nutritious food. In addition to this I am interesting in hearing about your understanding of available information and guides regarding nutrition and whether you use any tools to help you.

Part 1 - current eating habits:

1. First I would like you to tell me about your food habits during a regular day:
    a) eat at specific time? b) Do you eat the same things for breakfast and lunch every day? c) Are most of your meals homemade? d) Do you make packed lunches for yourself and your children (if any)?

2. Are your children picky eaters, or are there some food groups they won't touch?

3. Do you find it difficult to plan what to have for dinner?

4. Who cooks at home? Do your children help sometimes?

5. Do you consider your family's diet to be healthy and balanced?
    a) Do you find it difficult to eat healthy?

6. Have you had concerns regarding what your children should or should not eat? Do you have any rules regarding this?
    a) Do you eat the same meals you serve your children?

7. Do you like cooking food?

8. In your opinion, why do you think sometimes families with young children find it difficult to eat healthy?

9. Did/do you have any concerns about becoming a parent and being responsible for furthering healthy habits to your children?

Part 2 - Have you tried to change your diet in the past or do you wish to change your diet?

If yes: what was your motivation for this change?

10. Have you ever used any tools to help you make these changes? Such as apps or website, schedules, plans, groups or forums to help you?

11. What do you think could help you make healthier choices?

12. Are there any information or health advice that you find difficult to use in everyday life?

13. Do you eat all 5 a day?
    a) why/why not?
    b) Do you know what 5 a day actually looks like? (1 portion is 100g)

Part 3 - Activity Levels:

14. What are your family's activity level like?

15. Are you active together as a family?

16. Do you think increasing the activity level or eating more healthy is more important?

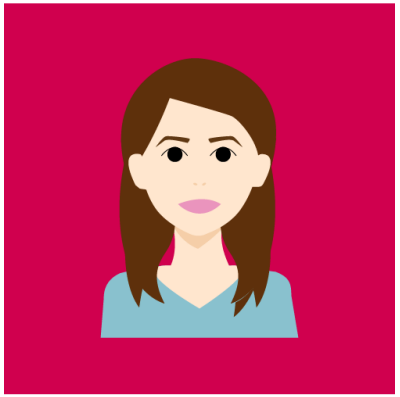
17. Do you use any tools to monitor your activity levels?
    a) Tools can be apps, journals, websites, wearables, gym membership, gym classes, running groups etc.

Part 4 - Food Waste:

18. What are your thoughts regarding food waste?
    a) Do you end up wasting food?
    b) What are the reasons why you/why you don't sometimes waste food?

21. How often do you go to the grocery store?

22. Have you actively tried to reduce food waste?



Name: **Sofie Hansen-Lunde**

Quote/Tagline

"After I had children my motivations for working out and being fit changed. Now I am motivated to have enough energy and "want" to play with my children. I need to be healthy for them"

Overview

Sofie is a mother of two, she has been married to Andreas for 5 years and works as a child care worker. She is currently on leave with her youngest child Emma, but she will return to work in a few months and then dad will take over for 10 weeks. Her oldest child Oskar spends his day in the kindergarten where he is provided with three meals.

Both Sofie and Andreas have been very active, participating in sports all their life. Now that they have two small children, Sofie have kept being active even if her days are very full. She is motivated by being healthy for her children, so that she has the energy to play with them.

Sofie is more concerned about healthy eating than her husband, and she is more strict after becoming a mom. Now they have fish once a week, and dinner is always served with a salad or vegetables. Although the children won't always eat the vegetables or fish, she will continue to serve it to set a good example and teach good habits. As the children will not finish or try most foods they do end up with leftovers, and this usually ends up in the trash. They always eat the same things as their children for dinner. The children are interested in helping out with the dinner, and sometimes will be more inclined to try new things or finish their plate if they participated - however the cooking must be fun and not time consuming, or they will lose interest quite quickly.

Sofie does not enjoy cooking dinner anymore as it is more about getting food on the table as quick as possible. This often results in half-processed foods (ready sauces, fishcakes, meats) that can be prepared quickly. They try to do their shopping twice a week, but sometimes they do it more. Planning what to eat for dinner is a struggle, and the same meals are repeated every week. Sometimes feels guilty when serving half processed foods, but they don't have the time to plan ahead and make everything from scratch. If dinner is not on the table in 30 minutes their children will be difficult and moody.

Age: 32
Occupation: Child Care Worker
Education: Bachelor degree
Location: Norway, Lier
Marital status: Married
Children: boy 3 ½ y.o, girl 10 months

Technical comfort

Comfortable with mobile devices such as smartphones and tablets, as well as computers. Sometimes struggle to find her way around the three remotes to the new smart TV, but she is getting there. Recently purchased a fitbit charge to help her keep on track with her workouts and sleeping pattern.

Goals (what does she wish to achieve with the system)

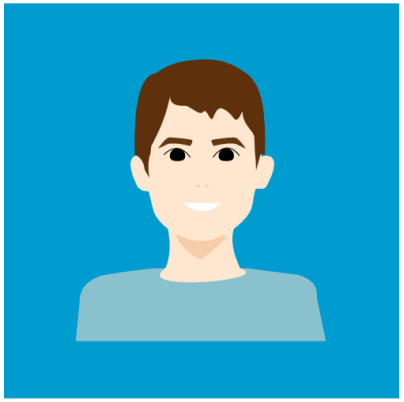
- Be able to track nutrition
- Receive help and tips regarding weekly meal plans
- Make sure her kids are getting all their nutrients

Motivations (why does she want to use the system?)

- Wants herself and her husband to be healthy, fit and energetic in order to take care of their two children
- Make sure her children has the absolute best start in life, learns good healthy habits for the future

Frustration and pains (what prevents her from using the system?)

- Time consuming activity to track and log meals
- Difficult to plan ahead - need inspiration
- Limited time "tidsklemma" to plan and prepare meals
- Other parents seems to brag on social media - guilt



Name: **Andreas Hansen-Lunde**

Quote/Tagline

"I leave the cooking to my wife, but I try to help out in other ways by doing the shopping on my way home from work. But I guess I could participate more in planning meals"

Age: 32
Occupation: Physical Therapist
Education: Masters degree + 1 year internship
Location: Norway, Lier
Marital status: Married
Children: boy 3 ½ y.o, girl 10 months

Technical comfort
Very comfortable with most technology. He operates smartphones and tablets everyday as well as his computer for work. Just purchased a new smart TV and surround system. Andreas is interested in the newest technology is now making his house smarter by the second hoping to control everything from his smartphone soon.

Overview
Andres is a father of two young children and married to Sofie, and works as a physical therapist. As Sofie is home with their youngest she is in charge of planning and cooking dinner and packed lunch for their children. Andreas usually do the shopping on his way home from work, twice a week, but in reality he shops more as there's always something they need or have forgotten.

Andreas has never been to concerned with eating healthy as he has continued with the eating habits taught by his parents. A normal healthy diet as he views it, although he has to eat bread with chocolate spreads for lunch everyday. No matter what Sofie says. Andreas was a little hesitant at first when Sofie said that now that they have to change their eating habits now that they are parents. No more pop in the week, fish at least once a week, more vegetables with dinner. At first he thought these changes didn't concern him, but he has learnt that it is important to lead by example if they wish their children to eat better as well.

Andreas is very active both at work and in the evenings. He plays football a couple evenings a week, and also works out with Sofie from home 2-3 times a week following a new workout app.

Andreas has never been interested in cooking, and having his wife at home for her maternity leave, he leaves meal planning and shopping lists to his wife. However, lately he has understood that his wife is very tired from planning everything by herself and wants to take an active role in the meal planning and shopping. Soon his 10 weeks of paternity leave will begin, and if Sofie is to continue to cook dinner he must take a more active role in the planning and shopping than he is currently doing.

Goals (what does he wish to achieve with the system)
- Be better at helping his wife with meal planning
- Participate more regarding the shopping and keeping track of what they need
- Help his kids eat more healthy

Motivations (why does he want to use the system?)
- Be more active and help out with dinners, planning and shopping
- Help his wife by taking his share of the dinner responsibilities
- Waste less food, more economical
- Wants his kids to be healthy

Frustration and pains (what prevents him from using the system?)
- Used to his wife planning dinner
- Does not know how to cook what his wife wants the family to eat
- Wife is tired of always having to be alone in planning meals and this puts pressure on him to take more action

User Scenarios

(Conceptual and concrete scenarios models from Benyon (2014) pp. 62-73)

Overall Conceptual Scenario

Parents with young children in Norway experience limited time to prepare and cook healthy and varied dinners for their children. As most parents work full time, and hectic schedules prevents them from spending time on cooking dinners, children end up eating too much processed foods. This leads to children not receiving the appropriate amount of nutrients they need during the day, and instead they consume too much salts, sugars and saturated fats. While most parents believe that their children are eating healthy and balanced, there is still a lack of fresh fruit and vegetables in their diets. There are still many myths that parents believe in regards to what constitutes a healthy balanced meal, and with more and more ready-made options being available, parents are not aware of what their children are actually eating. This has resulted in a large market of ready-meals or half-processed foods that allows parents to buy less ingredients and spend less time cooking. The issue is that ready-made alternatives have far more ingredients than people are aware of that leads to more sugar, salt, fats and calories in their diets. How can we ensure that parents become more aware of what is in the food their children are eating in a hectic day where time is a luxury item? With limited time to cook, plan and shop for healthy ingredients, parents struggle to be inventive in the kitchen and prepare meals that are both healthy and will be eaten by their children. Parents are in need of a simple solution, that requires little manual input and logging in order to keep track of what they eat and give suggestions for future meals that can learn and adapt to the requirements of the families likes and dislikes. The solution should act as a personal assistant that keep track for them and provides help, guidance and suggestions.

Story 1

“My children are in that stage where they don’t like a lot of food, in particular vegetables. This is not that my kids are just picky, their taste buds are changing and they develop a hesitation related to trying certain foods. We are worried that this stage will continue longer than normal if we don’t keep encouraging him to try and eat certain foods. We therefore always lead by example and always serve vegetables or salad with dinner. However, as the children usually only tries a little bit, we do end up wasting food at times. This is also due to a little laziness from our part as we rarely want to eat leftovers the next day for breakfast or lunch. And then we forget about the leftovers if we leave it in the fridge”.

Conceptual Scenario 1

Parents often end up with leftovers from dinner because their children do not finish their plate or refuse to eat certain foods. These leftovers are then forgotten or thrown away because leftovers are not something people want to eat. Parents are therefore in need of a tool to help them remember which leftovers occur most often, and reduce the amount in which they cook. The tool should also help remind parents of how much of the foods their children ate last time a similar dish was made.

Concrete Scenario 1

Sofie and Andreas needs to reduce their food waste when it comes to leftovers from dinner. It is difficult for them to always remember which meals their kids did not finish, or when they made too much food. Sofie consults the chatbot to ask whether the children finished their food previously, and how much they should prepare. The chatbot informs Sofie that last time her son Oskar did not finish a food group.

Notes:

1. After the chatbot learns more about the habits of the family and can make more accurate predictions, it should base the recipe amounts on previous eating habits
2. Also the chatbot should recommend alternative foods to replace foods the children don't like or did not finish previously

Story 2

"We have been forced to drastically change the way we eat and cook everything from scratch after the birth of our second child. We found that he suffers from several food allergies in particular all dairy products and eggs. As this is found in most ready made, half-processed foods we are no longer able to go for quick or ready-made solutions and have to not only cut out several food groups, but also cook everything from scratch. This way we have full control of everything we are eating and what it contains. It was surprisingly easy to adopt this new change, however we do find that a lot of time we struggle with time limitations and can therefore opt for too simple solutions still".

Conceptual Scenario 2

Some parents have strict food restrictions because of either their children having food allergies or change in taste buds, which forces them to eliminate certain foods and ingredients from their diets. In varying degrees this affects how much ready made, easy options they can go for. This makes it difficult for his parents to use certain ingredients such as ready sauces, ready mix, and certain "kid" friendly items such as fish cakes. Sometimes they are in a hurry and find it difficult to not opt for too easy solutions when time is limited. Time is an issue, they must therefore be able to make quick meals that does not contain specific ingredients.

Concrete Scenario 2a

Oskar has entered an age in which he does not like specific vegetables like red peppers and potatoes, which he used to eat a lot of when he was younger, in addition he is allergic to dairy. Being in a hurry at the grocery store, not being able to find a sauce without dairy, Andreas consults the chatbot to ask what they can make instead. The chatbot offers Andreas a quick recipe to make the sauce from scratch, however Andreas finds this complicated. The chatbot instead gives him options for ready made sauces that does not contain dairy.

Notes:

1. Some might use this app frequently and other infrequently
2. For infrequent users, they might not leave the planning to the agent and must therefore be able to access alternatives on the go
3. For frequent users, the chatbot would design meals based on ingredients they can eat

Concrete Scenario 2b

This week Oskar refused to eat his mushrooms, this was a surprise to his parents as he has always like mushrooms. After dinner Sofie consults the chatbot informing the agent that Oskar did not eat his mushrooms, because he said they tasted bad. The chatbot logs this and will for a time not include meals with mushrooms. A few weeks later the chatbot will suggest to try and serve a small mushroom and see if he likes it again.

Notes:

1. The chatbot should always ask for the reason why the child did not eat something
2. The reason should dictate the appropriate action
3. Foods should either be logged as: allergy, don't like (temporary/always)

Story 3

"When we found out we were having our first child we decided right then and there to change the way we eat. For starters we had to eat fish at least once a week and we stopped drinking pop for dinner. Overall we have both become much more healthy after we decided to have kids, we have also started being more active than we used to. Now we have to be healthy for our kids, they are our motivation"

Conceptual Scenario 3

Most parents become more concerned about what they eat once they become parents. Parents often introduce more fish and veg to their diets, but wrong assumptions regarding what is healthy and what is not contributes to parents thinking they eat healthier than they actually do. Most parents state that they serve a lot of bread, pasta, potatoes, rice, processed meats such as fish cakes, meatballs, sausages, bacon and fish sticks, and ready sauces such as bolognese. These foods often have a lot more calories than a homemade version, and more sugar, salts and fats. Processed meats are easy and usually a winner for most children, however these should be avoided to a degree in a healthy balanced diet. The chatbot should therefore provide parents with alternatives and state clearly the differences in unhealthy ingredients and calories to promote a healthier alternative.

Concrete Scenario 3a

Andreas is trying to figure out dinner, he has to pick up Oskar from the kindergarten afterwards and he is in a hurry. It's Tuesday which means they are having fish for dinner, he is looking for something quick and easy, but notices that most of the fish cakes are with dairy. He consults the chatbot saying that he needs a recipe for a fish dish that takes less than 30 minutes to prepare. The chatbot knows that Oskar cannot eat dairy and pulls up a tin foil roasted salmon and veg recipe that takes 10 minutes to prepare and 15-20 minutes in the oven - the same time as it takes to warm fish cakes. The chatbot also informs Andreas that he already has onions and broccoli at home and provides a grocery list without these items.

Concrete Scenario 3b

This has been a hectic week for Sofie and they have had three pasta dinners so far this week. The chatbot sends Sofie a notification. Sofie opens the app and reads the message:

First message: "I see you have eaten pasta three times this week. Remember that pasta does not provide your kids with any necessary nutrients or vitamins other than carbohydrates. Today I recommend replacing the pasta with a healthier alternative: quinoa, bulgur, zucchini, or what about the sweet potato you have left in your fridge?".

Sofie: "I'm not going to the grocery store today and Oskar does not want the sweet potato"

Chatbot: "That is understandable. I'll recommend this recipe instead "", only serve 50 g of pasta and add a few more vegetables to the side or sauce instead"

Concrete Scenario 3c

Sofie has just informed the chatbot what Emma has eaten today so far.

Chatbot: "Great effort so far Sofie! Emma has eaten all her daily dose of fruit. To finish her 5 a day I recommend carrots for snacks before or after dinner, and 1 of "her favorite vegetable" as a side for dinner."

Sofie: “which alternatives should I give if she won’t go for the carrots?”
Chatbot: “You could either serve the carrot with a homemade yogurt dip, seasoned with a little salt and pepper, or you can add two of “her favourite vegetable” today for dinner”.

Story 4

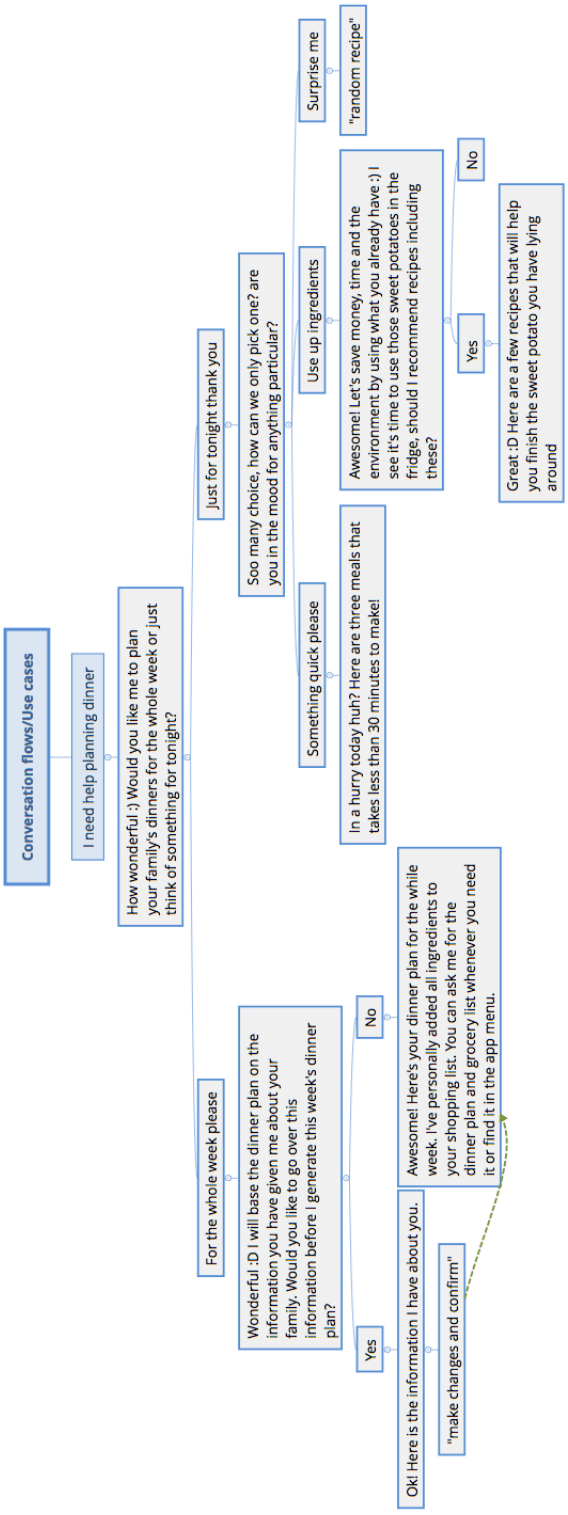
“We decide what to have for dinner on the day. We always pick up the kids from kindergarten and then we stop by the grocery store before heading home. Our fridge is therefore always stacked with breakfast spreads, milk, butter and eggs that we usually have for breakfast and lunch, while anything for dinner we by day by day. We therefore don’t plan or use up ingredients the next day, instead we buy stuff we know we can use for several meals. Our dinner meals therefore consists of mainly potato, pasta, pancakes and rice and then some form of protein like chicken or minced meat. We are worried that our children does not get all the necessary nutrients and therefore we need to start to plan our meals ahead of time, and mix things up a bit instead of the same meals over and over again”.

Conceptual Scenario 4

Some parents struggle with planning what to eat in advance. They know that it is more economical and time saving to plan ahead, but it has become routine to go by the grocery store on the way home from work. This they find this to be unnecessary, a time thief, and not the most economical approach.

Concrete Scenario 4

They consult the chatbot because they need help planning dinners for the week. They want to eat healthier, and they want to reduce the amount of trips to the grocery store, but they are worried they will be left with a lot of leftovers if they shop for the whole week. The chatbot provides them with a weekly menu, that provides the whole family with the necessary nutrients and calories for the whole day. They notice that although the menu doesn’t use a lot of pasta, rice or white bread, a lot of the same core ingredients are used from day to day. They accept the weekly menu, and the chatbot gives them a shopping list.


Conversation flows/Use cases
I need help planning dinner
How wonderful :) Would you like me to plan your family's dinners for the whole week or just think of something for tonight?
For the whole week please
Wonderful :D I will base the dinner plan on the information you have given me about your family. Would you like to go over this information before I generate this week's dinner plan?
Yes
Ok! Here is the information I have about you.
"make changes and confirm"
No
Awesome! Here's your dinner plan for the while week. I've personally added all ingredients to your shopping list. You can ask me for the dinner plan and grocery list whenever you need it or find it in the app menu.
Just for tonight thank you
Soo many choice, how can we only pick one? are you in the mood for anything particular?
Something quick please
In a hurry today huh? Here are three meals that takes less than 30 minutes to make!
Use up ingredients
Awesome! Let's save money, time and the environment by using what you already have :) I see it's time to use those sweet potatoes in the fridge, should I recommend recipes including these?
Yes
Great :D Here are a few recipes that will help you finish the sweet potato you have lying around
No
Surprise me
"random recipe"

SAMTYKKESKJEMA

Jeg _____ gir frivillig mitt samtykke til å delta i dette intervjuet.

Jeg forstår at selv om jeg har gitt mitt samtykke til å delta nå, har jeg rett til å trekke mitt samtykke til enhver tid eller nekte å svare på spesifikke spørsmål uten at dette skal ha noe form for konsekvenser.

Jeg forstår at jeg har rett til å trekke mitt samtykke til å bruke dataen fra mitt intervju til enhver tid for noen som helst grunn, hvor i dette tilfellet all data om meg vil bli slettet. Jeg forstår at etter publikasjon av rapporten 1. juni 2018, vil det ikke lenger være mulig å trekke mitt samtykke.

Jeg har blitt forklart hva dette intervjuet innebærer og hvilket formål det har, og jeg har blitt gitt muligheten til å stille spørsmål angående studien i sin helhet.

Jeg er er inneforstått med at deltakelse involverer å svare på spørsmål angående min og min families kostholdsvaner og aktivitetsnivå, i tillegg til min personlige erfaring og mening om utfordringer ved kostholdsendringer og økning av aktivitetsnivå.

Jeg forstår at jeg ikke vil ha direkte nytte av å delta i dette intervjuet.

Jeg forstår at all informasjon jeg gir for denne studien vil bli behandlet konfidensielt.

Jeg forstår at i en rapport om resultatene av denne undersøkelsen forblir identiteten min anonym. Dette vil bli gjort ved å ikke samle inn personlige data som kan avsløre identiteten min eller identiteten til personer jeg snakker om.

Jeg forstår at anonymiserte/skjulte utdrag fra intervjuet mitt kan bli sitert i: masteroppgaven, masterpresentasjon, publisert avhandling.

Jeg forstår at signerte samtykkeskjema og intervjunotater vil bli beholdt i personvernet til intervjuer Tuva Lunde Smestad, og at intervjuer Tuva Lunde Smestad og veileder Frode Volden er de eneste som har tilgang til rådata frem mot 1. juni 2018 hvor disse vil blir slettet.

Jeg forstår at jeg under norsk lovgivning har rett til tilgang til informasjonen jeg har gitt til enhver tid mens dette blir beholdt som presisert over.

Jeg forstår at jeg er fri til å kontakte de involverte personene i denne studien hvis jeg har behov for videre informasjon.

Kontakt detaljer og informasjon:

Intervjuer:
Tuva Lunde Smestad
Master Interaksjonsdesign
Fakultet for Arkitektur og Design
NTNU i Gjøvik

mob: email: tuvals@stud.ntnu.no

Akademisk veileder:
Frode Volden
Førsteamanuensis
NTNU i Gjøvik

mob: email: frodev@ntnu.no

Deltaker signatur

---------- ----------
Deltaker signatur Dato

Intervjuer signatur

Jeg mener deltaker gir sitt informerte samtykke til å delta i dette intervjuet

---------- ----------
Intervjuer signatur Dato

SAMTYKKESKJEMA

Jeg _____ gir frivillig mitt samtykke til å delta i dette eksperimentet.

Jeg forstår at selv om jeg har gitt mitt samtykke til å delta nå, har jeg rett til å trekke mitt samtykke til enhver tid uten at dette skal ha noe form for konsekvenser.

Jeg forstår at jeg har rett til å trekke mitt samtykke til å bruke dataen som omhandler min deltakelse til enhver tid for noen som helst grunn, hvor i dette tilfellet all data om meg vil bli slettet. Jeg forstår at etter publikasjon av rapporten 1. juni 2018, vil det ikke lenger være mulig å trekke mitt samtykke.

Jeg har blitt forklart hva dette eksperimentet innebærer og hvilket formål det har, og jeg har blitt gitt muligheten til å stille spørsmål angående studien i sin helhet.

Jeg er er inneforstått med at deltakelse involverer å teste to chatbot prototyper og evaluere disse ved å fylle ut 2 evalueringsskjemaer.

Jeg forstår at jeg ikke vil ha direkte nytte av å delta i dette eksperimentet.

Jeg forstår at all informasjon jeg gir for denne studien vil bli behandlet konfidensielt.

Jeg forstår at i en rapport om resultatene av denne undersøkelsen forblir identiteten min anonym. Dette vil bli gjort ved å ikke samle inn personlige data som kan avsløre identiteten min.

Jeg forstår at anonymiserte/skjulte utdrag fra mine kommentarer til eksperimentet og prototypene kan bli sitert i: masteroppgaven, masterpresentasjon, publisert avhandling.

Jeg forstår at signerte samtykkeskjema og data fra eksperimentet vil bli beholdt i personvernet til masterkandidat Tuva Lunde Smestad, og at masterkandidat Tuva Lunde Smestad og veileder Frode Volden er de eneste som har tilgang til rådata frem mot 1. juni 2018 hvor disse vil blir slettet.

Jeg forstår at jeg under norsk lovgivning har rett til tilgang til informasjonen jeg har gitt til enhver tid mens dette blir beholdt som presisert over.

Jeg forstår at jeg er fri til å kontakte de involverte personene i denne studien hvis jeg har behov for videre informasjon.

Kontakt detaljer og informasjon:

Masterkandidat:
Tuva Lunde Smestad
Master Interaksjonsdesign
Fakultet for Arkitektur og Design
NTNU i Gjøvik

mob: email: tuvals@stud.ntnu.no

Akademisk veileder:
Frode Volden
Førsteamanuensis
NTNU i Gjøvik

mob: email: frodev@ntnu.no

Deltaker signatur

Deltaker signatur

Dato

Masterkandidat signatur

Jeg mener deltaker gir sitt informerte samtykke til å delta i dette eksperimentet

---
Intervjuer signatur

Dato

Resultat av meldeplikttest: Ikke meldepliktig

Du har oppgitt at hverken direkte eller indirekte identifiserende personopplysninger skal registreres i forbindelse med prosjektet.

Når det ikke registreres personopplysninger, omfattes ikke prosjektet av meldeplikt, og du trenger ikke sende inn meldeskjema til oss.

Vi gjør oppmerksom på at dette er en veiledning basert på hvilke svar du selv har gitt i meldeplikttesten og ikke en formell vurdering.

Til info: *For at prosjektet ikke skal være meldepliktig, forutsetter vi at alle opplysninger som registreres elektronisk i forbindelse med prosjektet er anonyme.*

Med anonyme opplysninger forstås opplysninger som ikke på noe vis kan identifisere enkeltpersoner i et datamateriale, hverken:

- *direkte via personentydige kjennetegn (som navn, personnummer, epostadresse el.)*
- *indirekte via kombinasjon av bakgrunnsvariabler (som bosted/institusjon, kjønn, alder osv.)*
- *via kode og koblingsnøkkel som viser til personopplysninger (f.eks. en navneliste)*
- *eller via gjenkjennelige ansikter e.l. på bilde eller videoopptak.*

Vi forutsetter videre at navn/samtykkeerklæringer ikke knyttes til sensitive opplysninger.

Med vennlig hilsen,

NSD Personvern

NSD – Norsk senter for forskningsdata AS Harald Hårfagres gate 29 Tel: +47-55 58 21 17 nsd@nsd.no Org.nr. 985 321 884
NSD – Norwegian Centre for Research Data NO-5007 Bergen, NORWAY Faks: +47-55 58 96 50 www.nsd.no

Descriptive statistics Chatbot A

	N	Minimum	Maximum	Mean	Std. Deviation
human - technical	16	4	7	5,94	,772
isolating - connective	16	4	7	6,00	,816
pleasant - unpleasant	16	2	7	6,31	1,302
inventive - conventional	16	5	7	6,25	,577
simple - complicated	16	3	7	6,19	1,109
professional - unprofessional	16	2	7	5,88	1,360
ugly - attractive	16	3	7	5,94	1,063
practical - impractical	16	2	7	6,12	1,500
likeable - disagreeable	16	1	7	6,25	1,528
cumbersome - straightforward	16	1	7	5,69	1,778
stylish - tacky	16	4	7	5,50	,816
predictable - unpredictable	16	1	7	4,69	1,662
cheap - premium	16	2	6	5,00	1,033
alienating - integrating	16	5	7	6,13	,619
brings me closer to people - separates me from people	16	2	6	3,94	1,289
unpresentable - presentable	16	4	7	5,94	,998
rejecting - inviting	16	6	7	6,63	,500
unimaginative - creative	16	6	7	6,44	,512
good - bad	16	4	7	6,81	,750
confusing - clearly structured	16	6	7	6,31	,479
repelling - appealing	16	4	7	6,13	,719
bold - cautious	16	4	7	5,38	1,088
innovative - conservative	16	5	7	6,44	,629
dull - captivating	16	5	7	5,81	,544
undemanding - challenging	16	1	6	3,25	1,483
motivating - discouraging	16	5	7	6,38	,719
novel - ordinary	16	4	7	5,81	1,047
unruly - manageable	16	6	7	6,56	,512
Valid N (listwise)	16				

Descriptive statistics Chatbot B

	N	Minimum	Maximum	Mean	Std. Deviation
human - technical	16	1	6	2,94	1,769
isolating - connective	16	3	7	4,81	1,424
pleasant - unpleasant	16	3	7	5,44	1,504
inventive - conventional	16	3	7	5,38	1,310
simple - complicated	16	4	7	6,06	,854
professional - unprofessional	16	4	7	6,25	,856
ugly - attractive	16	2	6	4,50	1,033
practical - impractical	16	4	7	5,94	,772
likeable - disagreeable	16	4	7	5,56	1,153
cumbersome - straightforward	16	3	7	6,00	1,211
stylish - tacky	16	4	6	4,94	,772
predictable - unpredictable	16	2	7	4,94	1,289
cheap - premium	16	2	5	3,88	,806
alienating - integrating	16	2	6	4,75	1,342
brings me closer to people - separates me from people	16	1	4	3,19	1,047
unpresentable - presentable	16	2	7	5,56	1,153
rejecting - inviting	16	3	7	5,56	1,365
unimaginative - creative	16	3	7	5,62	1,455
good - bad	16	3	7	5,94	1,237
confusing - clearly structured	16	3	7	6,00	1,155
repelling - appealing	16	4	6	5,25	,683
bold - cautious	16	2	6	4,63	1,204
innovative - conservative	16	2	7	5,38	1,821
dull - captivating	16	1	7	4,50	1,461
undemanding - challenging	16	1	6	3,13	1,455
motivating - discouraging	16	2	7	5,13	1,258
novel - ordinary	16	2	7	4,81	1,905
unruly - manageable	16	4	7	6,44	,814
Valid N (listwise)	16				