Voice Assistants are Social Actors – An Empirical Analysis of Media Equation Effects in Human-Voice Assistant Interaction

Inaugural-Dissertation

zur Erlangung der Doktorwürde (Dr. rer. nat.) der

Fakultät für Humanwissenschaften

der

Julius-Maximilians-Universität Würzburg

Vorgelegt von

Florian Alexander Schneider

Würzburg, 20.06.2022

Erstgutachter: Professor Dr. Frank Schwab

Zweitgutachter: Professor Dr. Peter Ohler

Betreuer: Professor Dr. Frank Schwab

Danksagung

Besonderer Dank geht an Prof. Dr. Frank Schwab, der mir während des gesamten Prozesses der Doktorarbeit uneingeschränkt und jederzeit mit Rat und Tat zur Seite stand und durch konstruktive Kritik und kreative Vorschläge Verbesserungsmöglichkeiten aufzeigte. Endlich habe ich Zeit, all die Filme und Serien nachzuholen, die er mir während der Endphase dieser Dissertation empfohlen hat.

Ebenfalls besonderer Dank gilt Prof. Dr. Peter Ohler für die Betreuung dieser Arbeit als Zweitgutachter, seine wertvollen Kommentare zu dieser Dissertation sowie möglicher Anschlussstudien und für die spannenden Einblicke in die Anfänge der deutschen Medienforschung. Ich wünsche viel Spaß in der kommenden ruhigeren Zeit, sei es mit oder ohne Roadtrip auf der Harley.

Meinen Bachelor- und Masterstudentinnen und -studenten danke ich für ihre Hilfe bei der Datenerhebung.

Contents

List of Figures

List of Tables

Abbreviations

AI	Artificial Intelligence
ASR	Automatic Speech Recognition
CASA	Computers are Social Actors
CMC	Computer-Mediated-Communication
DM	Dialog Manager
ECA	Embodied Conversational Agent
HAI	Human-Agent Interaction
HCI	Human-Computer Interaction
HRI	Human-Robot Interaction
IoT	Internet of Things
ML	Machine Learning
NLG	Natural Language Generation
NLP	Natural Language Processing
NLU	Natural Language Understanding
SDS	Spoken Dialogue System
TTS	Text-to-Speech

1 Abstract

Ownership and usage of personal voice assistant devices like *Amazon Echo* or *Google Home* have increased drastically over the last decade since their market launch. This thesis builds upon existing *computers are social actors* (CASA) and *media equation* research that is concerned with humans displaying social reactions usually exclusive to human-human interaction when interacting with media and technological devices. CASA research has been conducted with a variety of technological devices such as desktop computers, smartphones, embodied virtual agents, and robots. However, despite their increasing popularity, little empirical work has been done to examine social reactions towards these personal stand-alone voice assistant devices, also referred to as *smart speakers*. Thus, this dissertation aims to adopt the CASA approach to empirically evaluate social responses to smart speakers. With this goal in mind, four laboratory experiments with a total of 407 participants have been conducted for this thesis. Results show that participants display a wide range of social reactions when interacting with voice assistants. This includes the utilization of politeness strategies such as the interviewer-bias, which led to participants giving better evaluations directly to a smart speaker device compared to a separate computer. Participants also displayed prosocial behavior toward a smart speaker after interdependence and thus a team affiliation had been induced. In a third study, participants applied gender stereotypes to a smart speaker not only in self-reports but also exhibited conformal behavior patterns based on the voice the device used. In a fourth and final study, participants followed the rule of reciprocity and provided help to a smart speaker device that helped them in a prior interaction. This effect was also moderated by subjects' personalities, indicating that individual differences are relevant for CASA research. Consequently, this thesis provides strong empirical support for a *voice assistants are social actors* paradigm. This doctoral dissertation demonstrates the power and utility of this research paradigm for media psychological research and shows how considering voice assistant devices as social actors lead to a more profound understanding of voice-based technology. The findings discussed in this thesis also have implications for these devices that need to be carefully considered both in future research as well as in practical design.

Zusammenfassung

Die Verbreitung und Nutzung von persönlichen Sprachassistenten wie *Amazon Echo* oder *Google Home* haben seit deren Veröffentlichung im Laufe des letzten Jahrzehnts stark zugenommen. Diese Thesis baut auf existierender *computers are social actors* (CASA) und *media equation* Forschung auf, die sich mit sozialen Reaktionen auf Medien und technologische Geräte befasst, die normalerweise nur in der Mensch-Mensch Interaktion auftreten. CASA Forschung wurde bereits zu einer Bandbreite an technologischen Geräten durchgeführt, darunter Desktopcomputer, Smartphones, virtuelle Agenten und Roboter. Trotz ihrer zunehmenden Popularität wurde bisher wenig empirische Forschung zu sozialen Reaktionen auf Geräte wie die genannten Sprachassistenten, auch *Smart Speaker* genannt, durchgeführt. Deshalb ist es das Ziel dieser Dissertation, soziale Reaktionen auf Smart Speaker basierend auf dem CASA Ansatz empirisch zu evaluieren. Zu diesem Zweck wurden im Rahmen dieser Thesis vier Laborexperimente mit insgesamt 407 TeilnehmerInnen durchgeführt. Die Ergebnisse machen deutlich, dass Nutzer eine Bandbreite an sozialen Reaktionen in der Interaktion mit Sprachassistenten zeigen. Darunter die Verwendung von Höflichkeitsstrategien wie des Interviewer-Bias, was zu besseren Bewertungen eines Smart Speakers geführt hat, wenn dieser direkt am Gerät selbst bewertet wurde. Im Vergleich dazu fielen Bewertungen, die an einem separaten Computer abgegeben wurden, schlechter aus. Die TeilnehmerInnen zeigten außerdem prosoziales Verhalten gegenüber einem Sprachassistenten, nachdem eine Interdependenz und Teamzugehörigkeit induziert wurde. In einer dritten Studie wandten die TeilnehmerInnen Geschlechterstereotype auf Sprachassistenten an, basierend nur auf der Stimme, die das Gerät in der Interaktion verwendet hatte. Dies zeige sich sowohl in einer Bewertung des Geräts als auch durch konforme Verhaltensmuster. In einer vierten und letzten Studie zeigten die TeilnehmerInnen reziprokes Verhalten und halfen einem Smart Speaker Gerät, das ihnen zuvor bereits geholfen hatte. Dieser Effekt wurde außerdem durch die Persönlichkeit der TeilnehmerInnen moderiert, was ein starkes Indiz dafür liefert, dass individuelle Unterschiede relevant für die CASA Forschung sind. Folglich liefert diese Dissertation starke empirische Belege für ein *voice assistants are social actors* Paradigma. Sie demonstriert die Nützlichkeit dieses Paradigmas für

medienpsychologische Forschung und wie die Betrachtung von Smart Speaker Geräten als soziale Akteure zu einem vertieften Verständnis von sprachbasierten Technologien führen kann. Die Ergebnisse, die in dieser Dissertation diskutiert werden, haben Implikationen sowohl für zukünftige Forschung als auch für das praktische Design von Sprachassistenten.

2 Introduction

In the 2013 movie *Her*, Theodore, a sad and lonely writer who recently went through a breakup – played brilliantly by Joaquin Phoenix – buys a new voice assistant called *Operating System One*. During the first boot sequence, he is asked whether he wants the assistant to speak with a male or a female voice and quickly decides to go with the female version – voiced for the movie by actress Scarlett Johansson. During their first conversation, the voice assistant reveals ‘her’ name – Samantha – and they have a short conversation that ends with the main character remarking: “You seem like a person, but you are just a voice in a computer”. Throughout the movie, the main character starts to show more and more social behavior when interacting with Samantha. Besides clearly treating the voice assistant as though it was female based on the voice alone, he displays signs of politeness by not directly criticizing her or by apologizing if he said something offensive. He also starts ascribing certain human traits to her, telling her that she has a good sense of humor and is very smart. Eventually, he even falls in love with Samantha and develops a more intimate relationship with the voice assistant. While the last part leans a bit into science fiction, as the voice assistant depicted in the movie is certainly more advanced than most technologies currently available for consumers, the main characters’ initial reactions are a lot closer to reality. Even though he is consciously aware that Samantha is just “a voice in a computer”, he can’t stop himself from interacting with her in a very human way. And he is not the only one. Technological devices interacting with people via speech is not a particularly new concept in media. Even back in 1966, there was a computer using voice input and output in the *Star Trek* TV show (voiced by actress Majel Barrett-Roddenberry, after which Google later named its first developed voice technology in real life: *Majel*). Two years later in 1968, the movie *2001: A Space Odyssey* depicted a sentient talking computer named HAL 9000 that eventually went rogue. In 1982, the TV show *Knight Rider* depicted a talking supercomputer car named K.I.T.T. that was able to think, learn and interact with humans. Interestingly, both HAL 9000, as well as K.I.T.T., used male voices which have become increasingly rare in recent depictions of voice assistants. In general, however, voice assistants – often in combination with depictions of artificial intelligence (AI) – only got more prevalent in media, now being a staple in many novels, movies, TV shows, and video games. While many of these depictions mirror real-life technological development, they are often exaggerated and ahead of their time. As far as real-life goes, conversational interfaces are widely considered one of the breakthrough technologies of the 21st century. With the launch of *Siri*, a virtual assistant developed by *Apple Inc*. and launched together with the *iPhone 4S* in October 2011, voice-enabled technology first became available to the broad public. Siri has since been an integral part of the *iOS* operating system and Apple claims it is now used by over 500 million people worldwide (Wardini, 2022). Siri was the first publicly released digital assistant using a voice interface to interact with users and can answer questions, give recommendations, or execute instructions. Initially, these services were only available on smartphones. However, three years later in 2014, *Amazon* released the *Amazon Echo* (short *Echo*), a stand-alone device referred to as a smart speaker. The Echo was released in combination with *Alexa*, a virtual assistant AI like Siri. The Echo contains a multitude of microphones for voice input, speakers for sound output and requires an internet connection to work, as Alexa is a server-based software. With this technology available, people were able to place the smart speaker anywhere they wanted within reach of Wi-Fi and an electrical outlet (McTear et al., 2016). By now, voice assistants are no longer a novelty and since the release of Siri and Alexa, many competitors have flooded the market: Google Assistant, Google Now, Cortana, Bixby, Alice, Celia, and Evi are just some of the many publicly available voice assistants. Worldwide sales numbers of smart speakers increase about 40% every year and are expected to reach 300 million units sold per year in 2025 not even including voice assistants integrated into other devices such as smartphones (Scott, 2021). What sets all these virtual assistants apart from other interactive technologies is the clear focus on voice and speech. We are used to interact with digital devices in a multitude of ways ranging from keyboards and mouses to touching screens, but none of them are natural to humans. Speech, however, is the most fundamental means of human communication and a very clear marker of humanness (Pinker, 1995). Thus, these devices are designed with the ability to display human-like characteristics – in this case, speech – and while this usually offers a more comfortable interaction between user and device it might also result in inappropriate or undesirable reactions. Compared to the depictions in media mentioned before, human-voice assistant interaction is still rather one dimensional and often one-sided, but due to rapid advancements in technology such as automatic speech recognition (ASR), natural language processing (NLP), deep learning (DL), machine learning (ML) and AI it is only a matter of time before that changes (McTear et al., 2016). Microsoft CEO Satya Nadella has coined the term *conversation-as-a-service* during a speech in 2016, which includes several future improvements in human-machine interfaces like giving technological devices and virtual interaction partners more human capabilities and at the same time making them more trustworthy, open, and respectful. Instead of mostly one-directional inputs usually consisting of the user entering a prompt and the device giving an answer, the goal is to allow true interaction in the form of a dialogue: “people-to-people conversations, people-to-digital assistants, people-to-bots and even digital assistants-to-bots. That’s the world you’re going to get to see in the years to come” (Della Cava, 2016, para. 2). What was once considered science-fiction in movies and TV shows might very well be a reality soon based on the rapid technological advances and wider adoption of voice assistant technologies in the years since. Thus, it is more important than ever to examine which mechanisms originally exclusive to human-human communication are transferred to these devices and what underlying psychological processes are relevant for this interaction. Which of these mechanisms are adopted in human-voice assistant communication and what are the resulting reactions and expectations towards these devices? Understanding why and when users will treat voice assistants as though they were human will not only lead to a deeper understanding of the technology itself but also will help facilitate the best user environment possible and even prevent possible abuse of automatic social reactions to technological devices.

2.1 Thesis Overview

The CASA paradigm originally examined social reactions to desktop computers, but has since been employed in human-machine communication, human-computer interaction (HCI), human-robot interaction (HRI), and human-agent interaction (HAI) (Gambino et al., 2020). Since the initial wave of CASA research, personal stand-alone voice assistants such as Amazon Echo and Google Home have entered the market and are constantly growing in popularity and distribution (Scott, 2021). These devices are highly relevant for CASA research, as they are specifically designed for social interactions using the fundamental means of human communication: speech (Pinker, 1995). Thus, the communicative ability of these devices far exceeds that of devices such as desktop computers which early CASA research and initial theories of people's social interactions with technology were based around (Guzman, 2019). Yet only a very limited number of empirical studies regarding these social interactions and users' reactions have been conducted as of this moment (Seaborn et al., 2021). Even fewer studies used objective behavioral measures to assess these reactions which is especially relevant as many of the underlying processes are theorized to be automatic and unconscious mechanisms that are hard to assess using only self-reports (Nass & Moon, 2000). Consequently, this thesis focuses on assessing social reactions, evaluations, and behaviors related to human-voice assistant interaction using both self-reports as well as behavioral measures. The smart speaker devices considered and used for this thesis are disembodied stand-alone voice assistants displaying only minimal social cues through voice. No additional visual social cues or anthropomorphology features are considered to clearly differentiate this research from literature on embodied conversational agents and social robots (Luger & Sellen, 2016). To achieve this goal, four experimental laboratory studies were conducted with a total of 407 participants. Four different core constructs from previous CASA literature were examined in human-voice assistant interaction: politeness and the interviewer-bias (Nass et al., 1999), team affiliation and prosocial behavior (Nass et al., 1996), gender stereotypes (Nass et al., 1997) and reciprocity (Fogg & Nass, 1997a). The moderating influence of participants' individual differences such as prior experience with voice-assistant devices, self-efficacy, anthropocentrism, willingness to suspend disbelief and personality traits were also investigated in these four experimental studies. Section 3 of this thesis aims to give an extensive overview of previous CASA research conducted on disembodied devices as well as the theoretical basis these experiments are based on. Section 4 describes the first empirical study conducted as part of the thesis. The study consisted of a replication of earlier work conducted by Nass et al. (1999) in which participants were shown to act politely toward a computer based on the interviewer-bias observed in human-human interaction (Finkel et al., 1991). The study was also designed to explore the question of whether prior experience with voice assistants influences social reactions or the assessment of the device. Experiment 2 (see section 5) examined the construct of group membership, more specifically team membership in human-voice assistant interaction. This experiment is based on work conducted by Nass et al. (1996) whose findings show that manipulating interdependence can induce team membership and thus influence subject's favorable assessments of a computer. The research design was expanded by including a behavioral measure in the form of prosocial behavior towards a voice assistant after interdependence was induced. Experiment 2 also explored the effects of deliberately implemented voice recognition errors and resulting failed states on behavior towards voice assistants and their assessment. Previous media equation research has also shown that people will apply gender stereotypes when interacting with various technological devices (Carolus, Schmidt, Muench, et al., 2018; Ernst & Herm-Stapelberg, 2020; Eyssel & Hegel, 2012; E. Lee et al., 2000; E. J. Lee, 2003; E.-J. Lee, 2008; Nass et al., 1997; Siegel et al., 2009). Experiment 3 (see section 6) was designed to replicate this research and transfer it to voice assistants using either a male or a female sounding voice. Conformity as an objective behavioral measure was introduced in addition to assessments of the device based on established dimensions of gender stereotypes. The fourth and final experiment (see section 7) was based on the principle of reciprocity. This was previously operationalized in HCI by Fogg and Nass (1997a). The aim of experiment 4 was to investigate whether prior helpful behavior from a voice assistant elicits reciprocal behavior by participants in return. Additionally, experiment 4 was focused on considering three additional individual factors that have been theorized to influence media equation effects but had not previously been examined in empirical CASA studies: personality, willingness to suspend disbelief, and anthropocentrism. Advances in technology between experiment 1 and experiment 4 allowed for the interactions between participants and voice assistants to gradually become more natural and less prone to technical and recognition errors as well as incorporating more markers of true interactivity such as information and responses being stored, referenced, and repeated at later points. The thesis concludes with a general discussion of findings in section 8. Implications for both design and future research are discussed.

3 Theoretical Background

This chapter begins with a look back at the beginnings of HCI and media equation research. The CASA paradigm as well as the term of media equation are introduced and defined, after which an overview of possible explanations for these phenomena is given. The second section provides a theoretical foundation of the most important social aspects of human-human communication that have subsequently been transferred to human-technology interaction as well as an exhaustive overview of previous CASA and media equation research and its implications for this dissertation. Since this thesis is focused on social reactions towards smart speakers and all four studies are voice-based, the third subchapter is an introduction to the topics of speech, voices, and their relevance in human communication. Evolutionary and psychological aspects of perception, categorization, and interpretation of voices are introduced. The fourth and last subchapter focuses on the devices used as the technological basis of this dissertation. Definitions and components of smart speakers, their functionality, and the underlying technological features are established, and previous research is analyzed.

3.1 Computers are Social Actors & Media Equation

3.1.1 The Beginnings of Human-Computer Interaction

In the field of HCI, computers and similar technological devices are not just an intermediary to transmit communication data from human to human but can act as interlocutors interacting with humans (Gunkel, 2012). These thoughts can be traced back to the 1950s and Alan Turing's famous *imitation game* (Turing, 1950). In his article *Computing Machinery and Intelligence*, Turing proposes that a machine is only truly intelligent if a user cannot tell it apart from a human counterpart. In his thought experiment, a person communicates with two communication partners via text-based messages. One of these communication partners is male, the other female, and the task is to correctly identify their respective gender. To do so, the person can ask both of them any question they want. However, one of the communication partners has been instructed to assist in making the right decision while the other was instructed to deceive the person asking the questions. Only if a computer can take the role of the interaction partner that deceives the person asking the questions, Turing considers it intelligent enough to pass the so-called Turing test (Turing, 1950). Another thought experiment that contemplated whether computers are intelligent enough to replace humans as interaction partners is the *Chinese room experiment* by John Searle (1980). It proposes that a person that only speaks and understands English is locked in a room. Another person outside the room can send text-based questions into the room, but the questions are written in Chinese. The person inside the room has a book with Chinese symbols that instructs them how to correctly respond to these questions just by picking the matching symbols from the book. Searle proposes that by using this method, the person inside the room can answer the questions without knowing Chinese and the person outside the room will not be aware of that fact. In this thought experiment, the person inside the room is a metaphor for a computer. Even though the person does not understand Chinese, with the correct tools they are still able to respond. Similarly, a computer does not necessarily need to understand human communication to imitate and partake in it (Searle, 1980). These two opposing thought experiments propose some of the most fundamental questions in the field of HCI that scholars are still pondering to this day: what does it mean to be human? Can technology imitate or even understand humans? And most importantly for this thesis: why and under which circumstances do people treat technological devices as if they were human?

Following Turing's thoughts that a machine could be more than just a means to transport human communication, researchers began to expand upon the distinction between the approach of communicating *through* a computer via computer-mediated communication (CMC) and communicating *with* a computer (Cathcart & Gumpert, 1985). As technology became more advanced, first simulations of HCI became possible. One of the initial programs to do so was called *ELIZA*. Developed by Joseph Weizenbaum in 1966, ELIZA was a text-based program that used NLP to deconstruct input sentences by keywords and constructed responses based on these keywords. While ELIZA was able to hold a basic dyadic conversation, it was unable to store information and thus could not reference earlier messages reducing the interactivity to one message at a time (Weizenbaum, 1966). In addition, every single keyword had to be manually added to a keyword dictionary along with rules on how to respond to it which was difficult due to the variability, ambiguity, and context-dependent interpretation of human languages (Hirschberg

& Manning, 2015). Still, according to Weizenbaum (1976), ELIZA did create the illusion of a human interaction partner among some people, especially those who did not have prior experience with computers. Other early programs that simulate human-computer interaction include *COACH*, a program designed to offer advice to users based on their browsing history (Selker, 1994) and *Letizia*, a program that tracks previously read webpages and uses that data to make browsing recommendations (H. Lieberman, 1997). With private computer technology being advanced enough to support these kinds of interactions in the early 1990s, scholars consequently began to study human-computer interaction from a psychological perspective.

3.1.2 The Computers are Social Actors Paradigm

Computer programs and agents becoming ubiquitous and creating at least a rudimentary illusion of being human interlocutors proposed a new question in the early 1990s: how do people react to those new programs and computers? To answer this question, Clifford Nass and his colleagues proposed the *computers are social actors* paradigm (e.g., Nass, Steuer, & Tauber, 1994; Nass & Steuer, 1993). It suggests that humans follow the same social rules and apply the same social scripts to human-computer interaction as they do in human-human interaction thus treating them as social actors. Experiments conducted as part of the CASA paradigm follow a particular research process that can be divided into four basic steps (Nass & Moon, 2000).

(1) A finding – usually from social psychology, sociology, or anthropology – literature regarding human-human interaction is chosen. One example would be the interviewer-bias (Finkel et al., 1991).

(2) One of the human interaction partners is replaced by a technological device, during early CASA research usually a computer. Thus, the device assumes the role of a social actor.

(3) The methodology of the original study is then replicated as human-technology interaction instead of human-human interaction. For the example of the interviewer-bias, the human interviewer would be replaced by a technological device such as a computer conducting the interview with human participants.

(4) The results obtained this way allow researchers to determine if social behavior is still applied during interaction with these devices.

Early CASA results confirmed that even minimal social cues like text or a name displayed on a monitor were enough for participants to assign a personality to a computer and follow the well-established principle of similarity-attraction: if a computer had a similar personality to their own, it was rated significantly better compared to a computer with a dissimilar personality (Nass, Moon, et al., 1995). Based on these initial findings, Byron Reeves and Clifford Nass then established the term *media equation* to explain people's reactions towards both physical features of virtual objects and social features of virtual objects (Reeves & Nass, 1996).

3.1.2.1 The Media Equation: Physical Features

Before examining social reactions towards media and technology based on social features, Byron Reeves and his colleagues focused their research on reactions toward physical features of media. This included the size of virtual objects (Detenber & Reeves, 1996), their visual fidelity (Reeves et al., 1993), and their motions and movement (Reeves et al., 1985).

Size. Detenber and Reeves (1996) argue that the size of objects in the environment is of evolutionary relevance to humans as big objects might represent both challenges and opportunities for survival. Thus, humans not only pay more attention to big objects but also recall them more easily. To test if this is also the case for media content, Detenber and Reeves (1996) presented participants with the same media content on either a large or a small screen. Results revealed that images presented on a large screen were perceived as more exciting and arousing and were recalled easier.

Fidelity. While the size of objects has direct implications for survival, fidelity is secondary for these judgments as rapid decisions about fight or flight can be made even based on the rough shapes of objects (Reeves et al., 1993). In an experimental setting, no significant differences were found for excitement, arousal, or recollection between images presented to participants in low and high fidelity (Reeves et al., 1993). Consequently, the authors argue that both real and virtual objects are processed by the same mechanisms.

Motion. Not just the size of objects has implications for survival and fight or flight mechanisms but also their movement. Even slight movements in our environment cause humans to orient towards that movement and pay attention to it to quickly decide on the appropriate behavior. This visual orienting response is again explained due to potential threats to survival moving objects (and especially big moving objects) represent (Reeves et al., 1985). As expected, participants paid more attention to objects in motion even when they were just presented on a screen and thus represented no real threat.

3.1.2.2 The Media Equation: Social Features

After providing initial evidence that both real and virtual objects are processed similarly by humans (Detenber & Reeves, 1996; Reeves et al., 1985, 1993), Reeves and Nass (1996) turned toward social features of media and technology. When a media device exhibits any cue or feature that could be interpreted as social, humans showed a tendency to react to it as though it was human. If a computer asks a question, we are inclined to answer. If it is polite, so are we. If a voice interface starts speaking, we automatically ascribe certain attributes to it based on the voice it uses – just like we would do for other humans. These are just a few examples of what the authors first called the media equation. Generally speaking, our reactions to and interactions with media devices are both fundamentally social and natural (Reeves & Nass, 1996). Reeves and Nass (1996) condense the term in one sentence: “media equals real life”. These interactions and reactions are described as unconscious and unavoidable and thus „[media equation] applies to everyone, it applies often, and is highly consequential“ (Reeves & Nass, 1996, p.5). Because these media equation effects are theorized to be universal, they are relevant for a multitude of research areas. Some of the early examples given are affective computing (Picard & Healey, 1997) and persuasive computing (for an overview, see Fogg, 2002) but they have also been transferred to a variety of teaching, e-learning, e-commerce and business contexts with a focus on how to achieve more comfortable experiences when using certain devices and interfaces (Nass & Yen, 2010). Reeves and Nass (1996) argue that understanding how and why people show social reactions towards computers is crucial as they become more and more capable to send affective and social cues. From a design standpoint, this can be used to improve usability by making the devices more understandable and interactions with them more enjoyable. Additionally, from an ethical standpoint, these social capabilities can also be used to manipulate users and influence their decision. This has been a focal point of persuasive computing for many years, as it can also lead to negative outcomes for users (e.g., Fogg, 1998, 2002; Ghazali et al., 2018; Heckman & Wobbrock, 2000; Siegel et al., 2009).

3.1.2.3 Social Cues and Social Actors

One of the most important questions in early CASA and media equation research was a far-reaching one: to what extent does a device or computer need to resemble a human interaction partner for these effects to manifest themselves? Initial research provided a rather surprising answer: even minimal social cues can trigger behavior usually reserved for human-human interaction. As mentioned, cues as simple as text displayed on a monitor as a communicative act or giving a computer a human name caused participants to ascribe certain human attributes like gender or personality to a computer and to display social reactions (Nass, Moon, et al., 1995). By no means does a computer need to look like a human, speak like a human or possess any form of AI to trigger these unwarranted social reactions. Reeves and Nass (1996) were able to show that it made no difference if a computer used voice output or text overlay to communicate. In both cases, the computer was rated more positively by participants if they had to rate a previous performance on the same computer compared to giving the rating on another computer. Again, participants followed a fundamental social behavior known exclusively from human-human interaction: they acted polite towards the computer (Reeves & Nass, 1996). Additional findings revealed that simply telling participants that they belong to the same team as a computer – and thus giving that computer a social role, that of a team member – resulted in significantly better ratings of that computer’s friendliness as well as performance compared to a computer that seemingly belongs to another (Nass et al., 1996). In this case, the only manipulation was a minimal cue in the form of perceived team membership. Still, it was enough to result in a better evaluation of a group member, an effect that has also been shown in social science literature (Wageman, 1995). To summarize, based on initial CASA and media equation research, if a computer was sending social cues or was assigned a social role, it triggered social behavior from humans. Thus, early CASA research concluded that a “rich human presentation” (Nass, Steuer, & Tauber, 1994, p. 77) is not necessary to trigger these social reactions and for people to treat computers as social actors. A social actor in this context is defined as an entity that can both adopt social characteristics and is able to give social answers or reactions (Nass & Brave, 2005).

3.1.3 Explanations for CASA & media equation

As mentioned before, technological devices used in early CASA research did not possess any physical human characteristics, thus making all social behavior towards them obsolete (Nass & Moon, 2000). One glimpse would have been enough to recognize that the interaction partner is not human and hence does not warrant social treatment. There is no logical reason to be polite to a computer or to rate it differently based on any social cues that it gives, be it personality, gender, or affiliation. Why then do results of CASA research point to the exact opposite? Why do Reeves and Nass (1996) claim that media equation is universal and almost unavoidable? There are quite a few possible explanations found in literature and this chapter aims to give an overview of the theorized causes for media equation as well as alternative explanations for media equation effects.

3.1.3.1 Demand Characteristics

In early media equation research, there was an argument that the way the computers interacted with participants was unusually social thus leading participants to assume that they were supposed to act like they were in a social situation and disregard the fact that they were interacting with a technological device (Nass & Moon, 2000). However, there is a very strong argument against demand characteristics: in almost all media equation experiments, participants were neither aware of the expected social responses nor the theorized concepts (e.g., similarity attraction, politeness, etc.) and were instead given cover stories usually referring to their evaluation of the computer during various tasks. Even if their participation in an experiment influenced their behavior, they had no way of knowing what exactly was expected of them. Additionally, in many media equation experiments the appropriate behavior would even be contrary to the reported effects (e.g., gender stereotypes).

3.1.3.2 Addressing the Programmer & Source Orientation

One of the simplest explanations for media equation effects is that social behavior shown by users is always addressed towards the programmer instead of the device they are interacting with. Consequently, any social reaction is not targeted at the device itself, which is merely seen as a surrogate, but towards the person who programmed it (Nass, Steuer, Henriksen, et al., 1994; Sundar & Nass, 2000). To test this explanation, Reeves and Nass (1996) conducted a laboratory experiment in which participants interacted with a desktop PC. In one condition, that PC always referred to itself (using the pronoun "I") and in the other condition, it referred to its programmer ("the programmer of this computer"). If users always address the programmer as previously stated, this manipulation should not yield any significant differences between both groups. However, Reeves and Nass (1996) found significant differences in how both groups rate the computer, stating that computer and programmer cannot be regarded as equal (Reeves & Nass, 1996). Even when there were no differences in the actual interaction with the computer (meaning it did not refer to itself or a programmer at all) and participants were only told at the beginning of the experiment that they would either interact with a computer or that they would interact with a programmer with the computer being a surrogate, there were significant differences between those groups (Sundar & Nass, 2000). Furthermore, if participants were directly asked if they are addressing the programmer when interacting with computers, this notion was uniformly denied (Nass & Moon, 2000). Subsequent research has also shown that when users are not oriented toward the computer as a source, they do not display any social responses (Eckles et al., 2009; Shechtman & Horowitz, 2003; Tourangeau et al., 2003). Thus, Solomon and Wash (2014) argue that an orientation towards the device itself as an immediate source is the default orientation.

3.1.3.3 Intentional Stance

Related to the explanation of addressing the programmer is the explanation of the intentional stance (Dennett, 1989). This approach states that when interacting with a complex entity that was obviously created by humans, users tend to ascribe humanlike goals and characteristics to that entity. This heuristic is used to better understand the entity, to explain, and especially to predict its behavior. Dennett (1989) himself uses the example of a chess computer to illustrate his point. Its behavior cannot be explained from a functional stance (function and interaction between its pieces) or a physical stance (the workings of fundamental physical connections). The best way to predict its behavior then is the intentional stance: we assume the chess computer has the desire to win the game of chess as well as the rationality to put that desire into action. In short, the chess computer is seen as an intentional system, because it reflects the desires, ideas, and intentions of its programmer or inventor and is therefore treated as a proxy (Dennett, 1989). In CASA research, this explanation of a proxy is often discarded because of the aforementioned studies (see section 3.1.3.2) pointing toward participants neither unconsciously nor intentionally addressing the programmer during interactions with computers (e.g., Nass & Moon, 2000).

3.1.3.4 Flow

Flow is traditionally considered as a state of immersion in a task (e.g., Csikszentmihalyi & Csikzentmihaly, 1990). The conventional approach hypothesizes that flow occurs between a user and a task which results in the user becoming highly immersed in the activity (Csikszentmihalyi & Csikzentmihaly, 1990). A newer approach to the concept of flow instead focuses on the interaction between the user, task, and an additional artifact (Finneran & Zhang, 2003). Flow can then occur as a process of interaction and omit the artifact, in the case of media equation the device itself (Pearce, 2005). By eliminating both the direct effect of the technological device as an artifact and the interaction between the user and the artifact, flow experience results in a higher sense of control and focused attention. Thus, when individuals are highly focused on their social interaction with a device, they experience a flow state in which their experience becomes satisfying (Csikszentmihalyi & Rathunde, 1993). However, flow would only account for a more pleasant and satisfying human-computer interaction. It does not explain any form of social reactions or behavior towards technological devices and thus is very rarely mentioned in media equation literature.

3.1.3.5 Anthropomorphism & Ethopoeia

One of the first explanations found in literature for social reactions towards computers and the predominant explanation for these responses prior to the advent of CASA research is anthropomorphism (Turkle, 2011; Winograd & Flores, 1987).

Anthropomorphism initially described the psychological phenomenon of consciously and actively attributing human characteristics to nonhuman objects and artifacts. This includes but is not limited to technical devices and since the objects are clearly nonhuman, was initially considered a social or psychological deficiency in early CASA research (Nass, Steuer, & Tauber, 1994). According to this definition of anthropomorphism, media equation then occurs because as a result computers are understood to be humanlike and thus are to be treated as such (Nass & Moon, 2000). Early anthropomorphism research even considered that not only is there a tendency to ascribe human traits to computers but to consider computers as human and consequently react to them in a social manner (Turkle, 2011). However, initial CASA literature argues against this explanation. With very few exceptions, most subjects that participated in CASA experiments were adults with higher levels of education and prior experience with computers. Therefore, it is highly unlikely that these participants believed desktop computers to be human. Additionally, after being explicitly asked, participants overwhelmingly stated that computers are not human and thus do not justify any social reactions (Nass & Moon, 2000) and that these reactions would not only be unnecessary but in fact inappropriate (Nass, Steuer, Henriksen, et al., 1994). The concept of anthropomorphism has since been refined and broadened and can now be defined as “the tendency to imbue the real or imagined behavior of nonhuman agents with humanlike characteristics, motivations, intentions, or emotions” (Epley et al., 2007, p. 864). Considering that since the initial wave of CASA studies, the capacity of technological devices to display humanlike features has increased substantially, anthropomorphism has since been adopted as a regular explanation for media equation effects in literature. This is especially relevant for virtual agents or robots that benefit greatly from technological advances such as increased computing power and have consequently adopted more visual anthropomorphic tendencies (Gambino et al., 2020). Anthropomorphism is also considered an explanation for social reactions in the field of human-AI interaction as it includes the ascription of a humanlike mind to anthropomorphized entities (X. Li & Sung, 2021; D. Park & Namkung, 2021; Shank et al., 2019). This includes both the ability to feel (experience) as well as the ability to act (agency) (H. M. Gray et al., 2007) or even theory of mind capabilities in technological entities such as robots (for an overview, see Söderlund, 2022). Newer studies have also considered the naturalistic usage of humanlike voices that assistants like Alexa and Google Assistant display as an element that leads to anthropomorphism (for an overview, see Seaborn et al., 2021). Nass and Moon (2000) mention an alternative explanation to anthropomorphism: ethopoeia. Ethopoeia is described as a state in which direct answers of a social nature are given to a disembodied, nonhuman entity while being aware that these responses are not appropriate (Nass, Lombard, et al., 1995). However, it should be noted that ethopoeia is merely a descriptive term for behavior linked to media equation and not an explanation. At no point do the authors explain the processes underlying ethopoeia and the term has rarely been mentioned in CASA literature since.

3.1.3.6 The Evolutionary Approach

When they first discussed the term media equation, Nass and Reeves (1996) focused on an evolutionary explanation for social reactions toward technological devices. From the perspective of evolutionary psychology, the human mind is "a set of information-processing machines that were designed by natural selection to solve adaptive problems faced by our hunter-gatherer ancestors" (Cosmides & Tooby, 1997, para. 2). Thus, the human brain developed in our ancestors' world and is a product of evolution designed by natural selection to serve survival and reproduction (Buss, 2015). In this world, where humans spend over 99% of evolutionary history, every rich social behavior was exclusive to humans and every perceived object was a real physical object. Everything that seemed like a real person was a real person. Because these very simple principles were valid for almost all evolutionary history, automatic reactions based on these heuristics had an evolutionary advantage in the environment of evolutionary adaptedness (EEA) (Bennett, 2018). A mechanism that automatically detects other humans using minimal resources was advantageous, as successful interaction with other humans was fundamental for survival and reproduction. These mechanisms are still the basis of human behavior to this day (Buss, 2015) and humans are not conscious of these adaptive mechanisms. Cosmides and Tooby (1997) refer to this phenomenon as *instinct blindness*. As these evolutionary processes take a significant amount of time to change, our living environment has changed greatly while the human brain still resembles that of human ancestors in the Pleistocene and is adapted to the environment of these times (Cosmides & Tooby, 1997). This means that the human brain is more capable to solve problems encountered in a hunter-gatherer society compared to problems encountered in the modern world such as interaction with technological devices (Reeves & Nass, 1996). This can result in a mismatch between evolved mechanisms and the current environment, culminating in maladaptive behavior: nowadays, modern media meets old brains that are not fully adapted to these new technologies (K. M. Lee & Jung, 2005). Our brain still expects social cues to originate from other humans. When a technological device conforms to social norms or sends social cues, it elicits automatic and unconscious reactions as if it were human (Reeves & Nass, 1996). The conscious recognition that social cues can originate from nonhuman sources is insufficient for our brain to overcome these automatic mechanisms even if they are inappropriate in modern human-computer interaction, which is why Reeves and Nass (1996) ultimately utilized human evolution as the fundamental explanation for media equation effects.

3.1.3.7 Presence

Based on the findings of Reeves and Nass (1996), Lombard and Ditton (1997) argue that media equation is the result of presence: the illusion that a mediated experience is not mediated. By giving social responses to technological devices, users are ignoring the mediated nature of the experience (Lombard & Ditton, 1997). Thus, social cues sent by a device result in it being treated as a social entity or actor. In conjunction with presence, willing suspension of disbelief was also considered as a possible explanation for media equation in CASA literature (K. M. Lee, 2004; Reeves & Nass, 1996). Willing suspension of disbelief argues that people deliberately choose to forget about the virtuality of media and technology for increased enjoyment thus resulting in the feeling of presence (Reeves & Nass, 1996). However, numerous empirical results challenge the willing suspension of disbelief argument as an explanation for media equation effects. It was shown consistently that people display social reactions to technological devices not consciously and willingly but rather naturally and unconsciously (e.g., Nass & Moon, 2000). After initial considerations by Reeves and Nass (1996), the explanation of presence was absent in CASA literature until K. M. Lee (2004) returned to the topic once again adopting an evolutionary stance. He argues that instead of willingly suspending their disbelief, people automatically apply so-called *folk-psychology modules* when interacting with media and technology. These modules are defined as “innate or rapidly developed knowledge about how the social world (or other minds) works" (K. M. Lee, 2004, p. 500) and are considered to be domain-specific. Based on findings of people applying these folk-psychology modules (e.g., reciprocal behavior, in-group favoritism, and detecting traits such as personality) to media and technology (e.g., Fogg & Nass, 1997a; Y. Moon & Nass, 1996a; Nass et al., 1996), he argues that by not realizing the virtuality of the social experience, people consequently experience social presence during these interactions (K. M. Lee, 2004).

3.1.3.8 Mindlessness

To differentiate media equation effects from early definitions of conscious anthropomorphism, Nass and Moon (2000) introduced mindlessness as an explanation for social reactions towards computers. The mindlessness approach argues that most semantic information during a social interaction never even reaches consciousness. Instead, social cues trigger behavior that is governed by rules and routines with an over-reliance on categories and distinctions drawn in the past (Langer & Abelson, 1972). This prompted Langer to develop the concept of mindlessness. In a state of mindlessness, an individual is focused on a small subset of contextual cues and ignores most of the present environment (Langer et al., 1985). These cues then trigger scripts that have been learned in the past which leads to behavior that is detached from immediate circumstances (Langer, 1989; Langer & Moldoveanu, 2000). Instead of actively considering all available information about the current situation, the previously learned scripts are applied. These social scripts usually contain mental models for human-human interaction and are activated during relevant situations (Schank & Abelson, 2013). Nass and Moon (2000) argue that humans have internalized these scripts to such an extent that they even apply them to computers and naturally orient themselves towards the social cues instead of the asocial cues of these devices (Nass & Moon, 2000). If, for example, a computer exhibits a social cue, in a state of mindlessness this social cue strongly suggests a human interaction partner thus triggering corresponding social scripts learned in human-human interaction while other aspects of the environment – such as the fact that the cue originates from a technological device – are ignored. Social scripts that are examined in media equation research are usually selected from social psychology, sociology, or anthropology literature (Nass & Moon, 2000). Mindlessness as an explanation for media equation effects was later revisited in a series of experimental HCI studies that provided support for the mindlessness approach (Johnson, 2006; Johnson et al., 2004; Johnson & Gardner, 2007, 2009). Still, while mindlessness offers a possible explanation for the cognitive processes that underlay media equation effects, it fails to clarify when and why mindless behavior happens and under what circumstances technological devices are merely treated as tools (Nass & Moon, 2000). In addition, as with all explanations, the mindlessness approach cannot account for all media equation behavior. In a more recent study, after getting greeted by a robot, participants started laughing instead of reacting to the social cue thus indicating that their reactions were not as mindless as researchers expected (Fischer, 2011). Based on these results, Fischer (2011) speculates that people might diverge in their tendency to be mindless during interactions with technology.

3.1.3.9 Conclusion

In summary, there is a multitude of approaches to explain media equation and people's social reactions towards technological devices. It must be noted that there is no direct empirical evidence for any of these approaches and most of them are based on inferences from indirect empirical evidence. In addition, there is a tendency in media equation literature to mix explanations of biological determinism and nature (e.g., Reeves and Nass, 1996) with nurture-based explanations of human reactions to technological devices such as learned social behavior[1]. It is neither the aim nor within the scope of this dissertation to argue for either of the approaches or to provide an ultimate explanation for media equation effects but rather to examine these effects in human-voice assistant interaction empirically and from an objective viewpoint.

3.2 CASA & Media equation - Previous Research

This section aims to give a comprehensive overview of previous media equation and CASA research focusing on disembodied devices. Media equation and CASA research can be divided by three different criteria: (1) by type of technological device used to replace a human interaction partner, (2) by core

[1] For reasons of clarity and brevity, origins of the nature/nurture debate regarding media effects are not expanded upon further in this dissertation (for an overview, see Sherry, 2004).

concepts of human-human interaction the experiment draws from, and (3) by type of social cue employed to elicit a social reaction. Most initial CASA experiments were text-based followed by voice-based manipulations. Prior to CASA research being transferred to embodied agents and devices such as robots, visual cues were rarely used. As for technological devices, the majority of previous experimental media equation research was done in the late 90ies to early 2000s thus limiting the available devices to mostly desktop computers. Very little research has been done to transfer the CASA paradigm to mobile devices and research has only recently picked up again with the previously mentioned rise of voice assistants. In addition to these technologies, this chapter aims to give a brief overview of media equation research related to social agents, embodied virtual agents, and social robots but there is a clear distinction in literature between these embodied technologies and the more traditional media equation approach as these technologies by design incorporate more traditional markers of humanness (such as faces and humanoid shapes) that were not the focus of traditional CASA literature (Luger & Sellen, 2016; Nass & Brave, 2005). In addition to type of device, media equation research can be roughly divided into categories that are all based on the core concepts of human interaction they draw from: (a) social norms and social rules that include politeness, flattery, praise, apologetic behavior, and reciprocity to name a few examples. (b) Group effects which include among others identity, interdependence, and team affiliation as well as gender (stereotypes) and the resulting conformity. (c) Traits which include for example personality and similarity attraction, but also complementary attraction and consistency attraction. As these core concepts and underlying mechanisms from social psychology, sociology and anthropology literature form the foundation of all CASA research and the foundation of the experiments conducted for this thesis, they are described and defined in the following sections. Due to the abundance of literature on these concepts and for the sake of clarity and brevity, this section focuses on the specific literature used as the theoretical base in previous CASA research. In addition, the most relevant CASA studies conducted for each concept are briefly summarized and observed media equation effects are noted. A complete overview of all studies conducted under the CASA paradigm for disembodied devices can be found in Table 1.

Table 1. Overview of previous media equation research focused on disembodied devices.

Examined criteria	Technology	Form of social cue(s)	Source
Politeness (interviewer-bias)	Desktop PC	Text-based	(Nass et al., 1999)
Politeness (interviewer-bias)	Personal Digital Assistant (PDA)	Text-based	(Goldstein et al., 2002)
Politeness (interviewer-bias)	Website	Text-based	(Karr-Wisniewski & Prietula, 2010)
Politeness (interviewer-bias)	Smartphone	Text-based	(Carolus, Schmidt, Schneider, et al., 2018)
Politeness (flattery)	Desktop PC	Text-based	(Fogg & Nass, 1997b)
Politeness (flattery)	Desktop PC	Text-based	(Johnson et al., 2004)
Politeness (apologetic statements)	Desktop PC	Text-based	(Akgun et al., 2005)
Politeness and Impoliteness	Smartphone	Voice-based	(Carolus, Muench, et al., 2019)
Personality (similarity-attraction)	Desktop PC	Text-based	(Y. Moon & Nass, 1996a, 1996b; Nass, Moon, et al., 1995)
Personality (self-serving bias)	Desktop PC	Voice-based	(Y. Moon & Nass, 1998)

Personality (complementary attraction)	Desktop PC	Visual (onscreen character)	(Nass et al., 2001)
Personality (personality inference)	Desktop PC	Voice-based	(Nass & Lee, 2001)
Personality (similarity-attraction)	Desktop PC	Voice-based	(K. M. Lee & Nass, 2003)
Interdependence (team affiliation)	Desktop PC	Text-based	(Nass et al., 1996; Nass, Fogg, et al., 1995)
Interdependence (team affiliation)	Desktop PC	Text-based	(Johnson & Gardner, 2007)
Interdependence (ethnicity)	Desktop PC	Visual (animated character)	(E. J. Lee & Nass, 1998)
Gender stereotypes	Desktop PC	Voice-based	(Nass et al., 1997)
Gender stereotypes (conformity)	Desktop PC	Voice-based	(E. Lee et al., 2000)
Gender stereotypes (social identification)	Desktop PC	Voice-based	(Morishima et al., 2002)
Gender stereotypes	Desktop PC	Visual (animated character)	(E. J. Lee, 2003)
Gender stereotypes	Smartphone	Text + visual (sleeve)	(Carolus, Schmidt, Muench, et al., 2018)
Reciprocity	Desktop PC	Text-based	(Fogg & Nass, 1997a)

Reciprocity (cultural differences)	Desktop PC	Text-based	(Takeuchi et al., 1998, 2000)
Reciprocity (self-disclosure)	Desktop PC	Text-based	(Y. Moon, 2000)
Assignment of roles	TV	Visual (labels)	(Reeves & Nass, 1996)
Assignment of roles	Desktop PC	Text-based	(Koh & Sundar, 2010)
Assignment of roles	Smartphone	Text-based	(K. J. Kim, 2014)
Distance	Desktop PC	Text-based	(Y. Moon, 1998)
Humor	Desktop PC	Text-based	(Morkes et al., 1998)
Frustration	Desktop PC	Text-based	(Klein et al., 1999)

3.2.1 Social Norms

A social norm is defined as a construct with far-reaching benefits that is used to describe and explain human behavior (Cialdini & Trost, 1998). The CASA paradigm states that humans have internalized these habits and behaviors to such an extent that they apply them to technological devices even though this behavior is neither justified nor does it provide any obvious benefits (Nass & Moon, 2000). This chapter aims to give an overview of the social norms, scripts, and behaviors that were transferred to human-voice assistant interaction in the studies conducted for this doctoral thesis.

3.2.1.1 Basics & Definitions

Norms have been conceptualized in many forms going back to so-called *folkways* (Sumner, 2019) which are defined as habitual rites that have been established in social groups because they facilitated the satisfaction of basic needs. Sherif (1936) describes norms as mutually negotiated rules for social behavior. In addition to widely accepted rules for desirable behavior, norms can also contain rules that prohibit or at least discourage unacceptable social behavior (Triandis, 1994). Social norms are formed through interaction with other humans and can but do not have to be explicitly stated. They offer guidelines on how to act in various social situations and on what actions to take to contribute to social order (Jackson, 1960). It is important to note that violations of these norms are sanctioned by social networks and not the judicial system (Blake & Davis, 1964). One of the most fundamental functions of social norms is to determine the availability of social interactions as well as their limits. While many different social interactions can take place in any given space, social as well as cultural norms dictate the form of these interactions in a significant way. If, for example, two friends are talking to each other, we expect a less formal interaction when compared to an employee talking to their superior in a professional context (Beniger, 1986). While many social norms are unique to different societies, cultures, or groups, there are also universal social norms. These universal norms are usually valid for most of if not all human-human interaction and are therefore ideally suited to be considered in human-technology interaction (Nass et al., 1999).

3.2.1.2 Politeness and Impoliteness

The social norm of politeness is considered an integral part of any human society and it is postulated to be universal (P. Brown & Levinson, 1978). Based on the politeness theory by P. Brown and Levinson (1978), politeness is defined as the utilization of verbal strategies aimed at respecting the feelings of the recipient. P. Brown and Levinson (1978) adopt the concept of 'face' from Goffmann (1967). Face represents an individual's public image he or she seeks to protect and contains two components: the need to (a) feel liked (positive face) and (b) not feel exploited (negative face) (Goffmann, 1967). Polite speech contains strategies like compliments, praise, or thanks aimed at respecting these needs. According to P. Brown and Levinson (1978), politeness indicates that a speaker is willing to mitigate so-called face-threatening acts against a recipient. There are two categories of face-threatening acts: negative face-threatening acts either restrict or disrespect the freedom of a conversational partner in any way, shape, or form. This includes putting the conversational partner under pressure through orders, instructions, or demands but also by making offers or promises with the intention of influencing the recipient's behavior. Positive face-threatening acts occur when the recipients' feelings or needs are ignored. This includes direct insults, accusations, disagreements, and interruptions but also ignoring communication rules like not correctly addressing a conversation partner based on gender, age, or status (P. Brown & Levinson, 1978). Not mitigating these face-threatening acts is considered as a violation of politeness norms. Impoliteness itself is composed of two components: First, the speakers' words are contradictory to the recipients' expectations of how the speaker should address him or her. Secondly, the speakers' words offend the listener. Additional context-sensitive factors like intention can increase the offence caused (Culpeper, 1996). These violations of politeness norms evoke negative emotional reactions that can result in social sanctions for the perpetrator (Goffmann, 1967). It should be noted that the politeness theory by P. Brown and Levinson (1978) is not without its deficits. Even though the authors claim that it is universal, cultural differences still influence how people deal with face-threatening acts and whether positive or negative politeness strategies are used predominantly (e.g., T. Holtgraves & Joong-Nam, 1990). As all studies conducted as part of this thesis employ German samples, cultural differences are not relevant for the purposes of this dissertation and thus will not be discussed further. The politeness theory and the concept of face as a social commodity (P. Brown, 2017; P. Brown & Levinson, 1978; Culpeper, 1996; Grice, 1975) have also been criticized for never clearly stating the benefits and the mechanisms of how face generates its value (Mühlenbernd et al., 2021) or how these concepts can be distinguished from evolved mechanisms (F. Schwab, personal communication, March 15, 2022).

3.2.1.3 Flattery

Flattery is a special form of politeness. It is defined as communicating positively to another person without considering the true qualities or skills of that person (Jones, 1964). As humans have an inherent desire to think positively about themselves, a person being flattered is inclined to believe that the flatterer is following the social norm of honesty. Therefore, flattery is often accepted as truthful even if the person being flattered knows that this is not the case. This also results in a positive affect for the person being flattered and a better evaluation of the flatterer (Fogg & Nass, 1997b; Jones, 1964).

3.2.1.4 The Interviewer-Bias

One specific example of a situation in which politeness norms are used that is particularly relevant for this thesis is the so-called interviewer-bias. The interviewer-bias assumes that individuals who are interviewed or questioned directly have a tendency to adjust their answers in a socially desirable way based on the perceived preferences of the interviewer (Sudman & Bradburn, 1974) which leads to questions being answered in a way that will be viewed favorably by the interviewer (Matthews et al., 2003). In addition, responses can be influenced by the interviewer's gender, ethnicity, or group membership and are often adapted accordingly (Finkel et al., 1991). Answers are also adjusted to avoid causing offense to the interviewer, which is seen as a direct and impolite violation of social norms (P. Brown & Levinson, 1978; Culpeper, 1996). Additionally, flattery may be used by the interviewee as a positive politeness strategy. As a result, responses to a third party tend to be more varied than responses made directly to the person who is conducting the interview or evaluation (Kiesler & Sproull, 1986; Nass et al., 1999).

3.2.2 Social Norms in Human-Computer Interaction

As mentioned, social norms like politeness are inherently human and are usually advantageous for all parties in human-human interaction and early CASA research has shown that social norms are also adopted in human-computer interaction (e.g., Fogg & Nass, 1997b; Nass et al., 1999; Nass, Steuer, & Tauber, 1994). This section gives an overview of the most important experiments and results.

3.2.2.1 Praise and Criticism in HCI

One of the first empirical experiments conducted as part of the CASA paradigm was based on the rule that in human-human interaction, praise given by others is more valid than praise by oneself (e.g., Jones, 1990; Meyer et al., 1986; Wilson & Chambers, 1989). To examine if this effect can also be observed in human-computer interaction, two laboratory experiments were conducted (Nass, Steuer, Henriksen, et al., 1994). In the first experiment, participants were tutored by a computer followed by a multiple-choice test based on the tutoring session. Afterward, the computer evaluated its own performance in the tutoring session either positively or negatively based on the experimental group. Participants were then asked to rate the computer's performance during all three sections (tutoring/test/evaluation). Results revealed that participants rated a computer that praised itself as more helpful and responsive compared to a computer that criticized itself (Nass, Steuer, Henriksen, et al., 1994). In the second experiment, the praise and criticism either originated from the same computer that provided the tutoring or a completely different computer. Results show that participants ascribed a different 'self' to each computer and evaluated a computer's performance more positively when it was praised by a different computer compared to a computer praising itself. In addition, a computer that criticized itself was perceived as friendlier than a computer that criticized another computer indicating that praise from others is more valid than self-praise and criticism from others is perceived as more impolite than self-criticism (Nass, Steuer, Henriksen, et al., 1994).

3.2.2.2 Politeness Strategies in HCI

One of the most important politeness strategies in human-human interaction is to apologize for undesirable communication outcomes in order to avoid frustration (Akgun et al., 2005; P. Brown & Levinson, 1978). To assess the effects of apologetic feedback given by a computer, participants were asked to play a guessing game with a computer. The computer provided participants with hints and asked them to guess the word these hints alluded to. This continued until participants guessed incorrectly 10 times, after which one round of the game ended. Participants played 10 rounds in total. For every correct guess, all participants received the same feedback. Every time they issued a wrong guess or a round ended in failure, they were given feedback by the computer according to one of two conditions: apologetic feedback or non-apologetic feedback. Subjects in the apologetic condition received an apology from the computer that indicated that the computer was responsible for the incorrect guess by not providing helpful hints while subjects in the non-apologetic group were simply told that their guess was wrong before being asked to continue. After the game ended, participants were asked to fill out a questionnaire to evaluate the computer they interacted with. Subjects that received apologetic feedback from the computer not only felt more respected and more comfortable, but they also perceived the computer to be more sensitive to their feelings when compared to the non-apologetic feedback condition. The authors conclude that apologetic feedback from a computer coincides with the effects of apologetic feedback in human-human interaction (Akgun et al., 2005). It is important to note that these effects were only assessed using subjective-self reports via a questionnaire. No behavior measures were employed by Akgun et al. (2005).

3.2.2.3 Interviewer-bias in HCI

In an experiment by Nass et al. (1999), subjects were asked to complete a text-based interactive learning unit using a desktop computer. Subjects were told that they would be presented with 20 randomly selected facts from a pool of 1000 facts on various topics. The computer then asked participants to rate these facts based on how much they knew about that fact. Subjects were told they would be given appropriate follow-up facts based on their ratings when in fact all participants were given identical facts. This tutoring session was followed by a multiple-choice test about the previously learned material. Afterwards, the computer evaluated the subjects' performance. Regardless of their performance, all subjects were told that they answered eight of the twelve questions correctly. In a last step, subjects were asked to evaluate the computer's performance either (a) on the same device, (b) on a different computer, or (c) via paper-and-pencil questionnaire. Results revealed an interviewer-bias: subjects who submitted their evaluation on the same computer rated it significantly better than subjects on a different computer or in the written questionnaire. However, in a debriefing session, all subjects indicated that the condition to which they were assigned had no effect on their responses. Subjects in the condition with the same computer indicated that they would have answered the same way if they had been required to give their feedback on a different computer. All subjects also indicated that they felt it would be unnecessary to show courtesy to a computer. The same experiment was repeated by Nass et al. (1999) using a speech-based system instead of a text-based one. Once again, subjects rated the assistance significantly worse on a different computer compared to evaluations given on the same computer. Nass (2004) argues that the mere presence of the ‘interviewer’ during the same-computer evaluation is sufficient to cause these effects compared to ratings given to a third party. The interviewer-bias was also shown to influence website ratings: participants were significantly more polite to a website when asked to rate the tutorial it gave them beforehand compared to ratings given on a different computer (Karr-Wisniewski & Prietula, 2010).

3.2.2.4 Flattery in HCI

Fogg and Nass (1997b) conducted a laboratory experiment to examine the effects of flattery in human-computer interaction. For that purpose, they had participants play a guessing game on a computer. To improve the game, the computer then asked participants to suggest questions for future iterations of the game. After giving a suggestion, participants were provided feedback by the computer based on one of three conditions. Both the 'sincere praise' and the 'flattery' condition received positive feedback. However, participants in the 'sincere praise' condition were also told that said feedback is based on their suggestion while participants in the 'flattery' condition were told that the feedback was chosen at random. In a third, 'generic feedback' condition, participants were simply told to start the next round of the game. Significant differences between both the 'sincere praise' and the 'generic feedback' as well as between the 'flattery' and the 'generic feedback' condition were found. Participants rated the computer's performance, their own performance, and their positive affect better in both the 'sincere praise' and the 'flattery' condition. Additionally, no significant differences were found between the 'sincere praise' and the 'flattery' condition. Participants showed the same reactions to positive feedback regardless of whether it was sincere or not – the same effect that flattery would have in human-human interaction (Fogg & Nass, 1997b). Johnson et al. (2004) replicated and expanded the experiment conducted by Fogg and Nass (1997b) by considering participants' previous experience with computers as an additional factor. The same methodological design was used. Participants were then asked for how many years they had been using desktop computers. Based on the answers given, they were separated into a 'high experience' and a 'low experience' group. Results revealed that participants with high experience reacted to flattery as expected. They believed that the computer was truthful, experienced more positive affect, and evaluated the computer's performance more positively. However, participants with low experience did not show any effects of flattery. Johnson et al. (2004) interpret these findings based on the mindlessness approach (see section 3.1.3.8): experienced users tend to engage in some form of ‘autopilot’ during their interaction with a computer and thus react to social cues automatically and without thinking. Experience is an important factor to consider for this thesis, as voice assistants are a comparatively new technology for most users and participants, which will be discussed in further detail later (see section 3.4).

3.2.3 Reciprocity

Reciprocity is one of the core concepts of being human (Cialdini, 2007). The principle of reciprocity can be condensed into one rule: people should help those who have helped them previously. Following this rule, reciprocity increases the chances of survival of the entire species (Gouldner, 1960). The human social system is based on the innate obligation of exchange, whether it be the sharing of help, food, or knowledge (Trivers, 1971). Consequently, one of the most important aspects of reciprocity is detecting cheaters and to reject them. Cosmides and Tooby (1992) argue that humans have even evolved a mind module very sensitive to reciprocal behavior to identify these cheaters. This is crucial, as cheaters do not reciprocate even after being afforded help by others. Consequently, without the detection of these cheaters, reciprocity can no longer be maintained (Cosmides & Tooby, 1992). In addition, reciprocity leads to a feeling of obligation to repay even when people did not explicitly ask for what they have previously received (Cialdini, 2007). Humans are constantly trying to balance the perceived social indebtedness through actions in order not to lose the societal status they earned. There is also strong evidence in literature that the norm of reciprocity is very powerful and not only valid across almost all human cultures but also a major theme of education and folk tales in many of them (Cialdini, 2007). One of the most cited studies on reciprocity has shown that subjects who received a gift – even if unsolicited – were more willing to buy something from the giver in return than subjects who had not received anything (Regan, 1971). The norm of reciprocity has also been shown influence self-disclosure. After receiving personal information from another person, people tend to disclose intimate information about themselves even to a person they are not close to (Dindia et al., 1997; Y. Moon, 2000). Generally speaking, there is a comprehensive body of research concerning the norm of reciprocity showing that people feel indebted to those who help them and obligated to reciprocate when they receive a favor of any form from someone (Hogg & Vaughan, 1995).

3.2.3.1 Reciprocity in HCI

In a series of experiments conducted by Nass and colleagues reciprocal behavior has been shown multiple times in human-computer interaction (Fogg & Nass, 1997a; Y. Moon, 2000; Takeuchi et al., 1998, 2000). In the initial experiment, Fogg and Nass (1997a) paired participants with a computer that either provided them with either helpful or unhelpful information during a web search. Afterwards, subjects were given the opportunity to help the same computer or a completely different computer during a second task focused on color perception. In this second task, participants were asked to order several color palettes based on their brightness. The number of palettes subjects chose to order was considered the indicator for the amount of help they were willing to give the computer. The authors predicted two main effects. (1) Subjects in the reciprocity condition who had received helpful information would show more reciprocal behavior towards the computer than the subjects in the unhelpful condition and (2) subjects would show more reciprocal behavior towards the computer that helped them before compared to a second, different computer. Both hypotheses were confirmed by the results obtained by Fogg and Nass (1997a). Participants not only ordered more palettes in the helpful condition but also devoted more time to the task. A similar experiment was conducted by Takeuchi et al. (1998). Instead of a web search, they opted to use the desert survival problem and had the computer provide participants with helpful or unhelpful information about the items they had to choose. The second task was identical to the previous study conducted by Fogg and Nass (1997a). Results revealed that participants again showed significantly more reciprocal behavior towards a computer that helped them previously and less to an unhelpful or a different computer (Takeuchi et al., 1998). The second study was replicated once again with a Japanese sample by Takeuchi et al. (2000). Based on the assumption that the Japanese culture is a collectivist culture resulting in behavior that is not individual-oriented but instead influenced by group-affiliation (Triandis, 2018), different results were theorized for Japanese participants. Results revealed that Japanese participants showed more reciprocal behavior in the helpful condition regardless of whether it was the computer that helped them before asking for help or a different computer. The authors interpreted these results as confirmation that Japanese participants act according to the social norms in their collectivist culture by showing reciprocal behavior not to an individual computer but by grouping the computers and reacting socially towards that group (Takeuchi et al., 2000).

In another experimental study, the concept of reciprocal self-disclosure was transferred to human-computer interaction (Y. Moon, 2000). Participants were interviewed by a computer that either disclosed some information about itself in the form of technological facts before asking them intimate personal questions (e.g. “What have you done in your life that you feel most guilty about?”) or asked the questions right away without giving any information about itself. Results revealed that subjects in the self-disclosure condition showed more reciprocal behavior by giving answers that were more intimate both in depth (how personal the given answers were) and breadth (how extensive the given answers were based on word count). An experimental study conducted by Velez (2015) examined the effects of team affiliation on reciprocal behavior. Participants were asked to play a video game with two teammates, one of which was a confederate while the other was a computer-controlled character. The manipulation was operationalized by having teammates be either helpful or not helpful during the game. In a second part, subjects were issued a prisoner’s dilemma in which they could donate money to quantify their reciprocal behavior depending on the perceived helpfulness of their teammates. Results revealed that subjects donated significantly more money following the interaction with a helpful teammate when compared to subjects in the unhelpful teammate condition. Additionally, a significant mediation effect was found for expectations of pro-social reciprocity. The author concludes that interacting with an unhelpful partner causes a decrease in prosocial behaviors as reciprocity expectations are disconfirmed (Velez, 2015).

3.2.4 Group Membership

Social identification is crucial for one of the most fundamental human principles: the more similarities between two humans, the more positive their disposition towards each other (Tajfel, 1974). These similarities extend to various in-group characteristics such as gender, ethnicity, or team membership. Some of these categorizations are inherently given and others are assigned, but all members of a group share the common need to belong (Tajfel et al., 1971). As the societal system of humans depends on group membership, building and maintaining these relationships is one of the core motives of an individual's actions (Leary & Cox, 2008). Tajfel and Turner (1979) define a group as a collection of individuals that are (a) members of the same social category, (b) share some emotional involvement in their common definition of themselves, and (c) have a similar evaluation of their group membership. Identification with a group positively impacts dispositions and behavior towards other members of that group. This includes the perceptions of other group members as more trustworthy, more likeable, and more intelligent (Tajfel, 1974; Tajfel et al., 1971). To identify the minimal conditions based on which an individual categorizes its belonging to a certain group, an experimental study was conducted in which participants were arbitrarily assigned to two groups by virtually meaningless properties, such as preference for an artist (Tajfel et al., 1971). Even this artificially induced minimal distinction between the groups led to ingroup favoritism and resulted in participants not only reporting a sense of belonging but also donating more money to other members of their perceived in-group (Tajfel et al., 1971). Even when explicitly told they were randomly assigned to a group, participants still resorted to competitive intergroup behavior and discrimination against perceived outgroups (Billig & Tajfel, 1973). There is a variety of explanations for positive attitudes and behavior towards others who are similar to oneself: social psychology proposes the factor of predictability, as other people who are similar to oneself are easier to predict which facilitates information processing and helps with the planning of behavior in social situations (Cantor & Mischel, 1979). Cognitive psychology argues for a simplified cognitive economy through similarity. Thoughts and behaviors of people similar to oneself are easier to understand without reflection thus reducing the cognitive load (Newell, 1994). Considering group membership from an evolutionary perspective, the survival and reproductive success of genetically similar others are almost as viable as one's own reproductive success and consequently, humans evolved to reliably detect genetic in-groupness (Dawkins & Davis, 2017). Two main factors that lead to an in-group bias and team formation in human-human interaction have been of special interest to the CASA paradigm: identity and interdependence (Nass et al., 1996; Nass, Fogg, et al., 1995).

3.2.4.1 Identity

One explanation of why individuals favor their ingroup is given in the Social Identity Theory (SIT) first mentioned by Tajfel (1974) and further developed by Tajfel et al. (1979). According to the SIT, people have both a personal identity and a social identity as part of their self-concept. The social identity is described as the awareness of membership in a certain social group and the value the individual assigns to it (Tajfel, 1974). Distinguishing between an ingroup, which is the group the person feels to belong to, and an outgroup provides “a system of orientation for self-reference: they create and define the individual’s place in society” (Tajfel & Turner, 1979, p. 40). As people strive to maintain a positive self-concept, which is closely related to self-esteem, they are constantly trying to positively evaluate their affiliation to their own group through social comparisons with other relevant groups (Tajfel et al., 1979). As mentioned in section 3.2.4, this social identity is easily manipulated even with minimal cues.

3.2.4.2 Interdependence

A second factor considered in CASA literature to induce the feeling of group membership and cause team formation is interdependence. Interdependence entails that a group or team member’s outcome is directly tied to the outcome of the entire group or team (Nass et al., 1996). Consequently, the success or failure of an individual is dependent on team performance. Literature indicates that individuals then classify themselves as members of a social group based on perceived interdependence among group members (Rabbie et al., 1989). Multiple studies support the hypothesis that interdependence is a crucial factor for ingroup-bias. Participants who were rewarded independently from other subjects did not show ingroup favoritism when allocating money (Karp et al., 1993). Interdependence also led to increased cooperation among group members (Crawford & Haaland, 1972) as well as a higher willingness to help each other (Van der Vegt & Van de Vliert, 2005; Wageman, 1995). When interdependence was salient, participants also perceived themselves as more similar to other group members and displayed more conformal behavior to group opinions (Mackie, 1986).

3.2.4.3 Identity and Interdependence in HCI

Since studying the effects of group membership is an important topic in social psychology, the effects of team affiliation based on identity and interdependence were replicated in early CASA research by conducting two laboratory studies (Nass et al., 1996; Nass, Fogg, et al., 1995). In the first experiment, participants were asked to complete a desert survival problem in interaction with a desktop computer. To induce identity and interdependence, participants were split into two groups. Subjects in the team condition were manipulated regarding identity by being told that they would work as members of the ‘blue team’ with a ‘blue computer’ and regarding interdependence by being told that the results of the desert survival problem would be evaluated as a team effort. Subjects in the individual condition were labeled as members of the ‘blue team’ but worked with a ‘green computer’. In addition, they were told that the results of the desert survival problem would be rated based only on their individual performance. The authors report that participants in the team condition not only displayed more conformal behavior towards the computer’s suggestions during the interaction but also ascribed more positive attributes to the computer in a subsequent evaluation. Participants also reported higher similarity to the computer based on team affiliation (Nass, Fogg, et al., 1995). The second experiment followed the same basic design, but identity and interdependence were manipulated separately by instead adopting a 2 (identity/non-identity) x 2 (interdependent/non-interdependent) design (Nass et al., 1996). Identity was constructed by color labels on the team’s name and the computers (e.g., ‘blue team’ working with the ‘blue computer’ vs. ‘blue individual’ interacting with a ‘green computer’). Interdependence was manipulated by telling the subjects either to be assessed individually (non-interdependently) or, in the case of interdependent subjects, that their joint performance with the computer was assessed at the end of the session. After working on a desert survival task, participants’ attitude was measured on the dimensions of perception of team affiliation, perceived similarity to the computer, the level of cooperation, openness to influence, and perceived information quality (Nass et al., 1996). Results revealed that identity did not influence any of the dependent measures, whereas interdependence had a significant effect on all of them (Nass et al., 1996). Creating dependency between users and computers led to higher levels of cooperation, higher conformity to the computer’s opinion, and higher ratings of friendliness and intelligence (Nass &

Moon, 2000). Contrary to Nass et al. (1996) who did not produce significant effects by manipulating identity in group work with computers, an experimental study by Johnson and Gardner (2007) produced effects based on the manipulation of identity in human-computer interaction. Being part of a human-computer team led participants to show different evaluations of the computer in terms of information quality, openness to influence, similarity, and cooperative behavior (Johnson & Gardner, 2007). The authors attribute these effects to the fact that in their experiment, unlike during the one conducted by Nass et al. (1996), the subjects were made aware of the existence of an outgroup.

Instead of creating a feeling of social identity through team affiliation, the effects of already existing groups have also been considered in CASA research. In an experimental study conducted in Korea, the effects of ethnicity as a social identity cue were examined (E. J. Lee & Nass, 1998). Korean participants were asked to read various social dilemmas to either a Korean computer agent or a Caucasian computer agent. The agent then provided participants with a solution to these dilemmas and a rationalization for that decision. Following this interaction, subjects were then asked to assess the computer agent and its decisions via a paper-and-pencil questionnaire. In addition, they were asked about their own solutions to the social dilemmas. Results confirmed that ethnicity had a significant effect in these interactions: participants that interacted with a computer agent matching their own ethnicity rated that agent to be more trustworthy and socially attractive compared to participants rating an outgroup agent. They also yielded to an ingroup agents' recommendations more often (E. J. Lee & Nass, 1998). Similar results were later reproduced with a comparison between African-American and Caucasian computer agents (Pratt et al., 2007).

3.2.5 Gender Stereotypes

3.2.5.1 Stereotypes

Stereotypes are defined as beliefs about certain characteristics, behaviors, and attributions of people that belong to certain groups (Hilton & Von Hippel, 1996). Stereotypes are not inherently negative, but they often have negative connotations when used to describe outgroups. Depending on the perception of the outgroup, they often result in generalizing and unfavorable sentiments (Fedor, 2014; Hilton

& Von Hippel, 1996). Stereotypes are adopted in almost all areas of everyday life and even result in simplified perceptions of one's own ingroups (e.g., one's own gender) (Costrich et al., 1975).

3.2.5.2 Gender Stereotypes

Gender stereotypes are socially shared convictions about certain specific features and mannerisms that are seen as characteristic of men or women (Ashmore & Tumia, 1980; Eckes, 2008). They have been shown to be far-reaching and well-entrenched across a variety of cultures (Hogg & Vaughan, 1995). Ashmore and Tumia (1980) emphasize that gender stereotypes are cognitive structures and therefore need to be considered as beliefs or attributions. The prevalence of gender stereotypes is a result of the deeply rooted definition of one's self that – for biological, psychological, and cultural reasons – often begins with gender (Banaji, 1993). Beginning at the age of two to three years, children begin to identify with either a female or a male group (Martin & Ruble, 2004) and tend to divide their playmates by gender (Maccoby, 1998). In addition to identifying with a gender, children also become "gender detectives" and start looking for clues about what a certain gender should or should not do by spending time with other people of the same gender and learning from them (Martin & Ruble, 2004). Children quickly develop notions of how men and women should behave and by the age of seven form very clear rules on appropriate behavior. An example of this are certain toys (like dolls) and colors like blue for boys and pink for girls (Martin & Ruble, 2004). By adulthood, people have absorbed thousands of these rules. They are used to judge how to behave towards others, how others are going to behave, and what reactions to expect. Previous CASA literature focused on gender stereotypes examined four core concepts linked to gender: women are attributed warmth and expressiveness (these contain, for example, friendliness, sensibility and need to belong) while men are attributed competence and instrumentality (these contain, for example, independence, assertiveness, and decisiveness) (Broverman et al., 1994; Deaux & LaFrance, 1998; Eckes, 1997, 2008; Sczesny et al., 2018). Regardless of modern societal trends like gender equality, these concepts possess both high stability (Bergen & Williams, 1991; Spence & Buckner, 2000) as well as a high cultural consistency (J. E. Williams & Best, 1990). As with every stereotype, gender stereotypes can be misleading when applied to specific persons. This does not dissuade most people from applying them anyways, even are unfavorable for their own gender (C. J. Deutsch & Gilbert, 1976). One of the most cited reasons for the application of gender stereotypes is a person's limited cognitive capacities. By quickly grouping other into categories, learned rules can be applied quickly instead of having to search and process additional information (Newell & Simon, 1972). Arguing from an evolutionary standpoint, correctly identifying the sex of others provides an evolutionary advantage as it is crucial for reproductive success as well as to judge and predict the behavior of others. Thus, humans have evolved to detect another person's sex immediately and with high accuracy (Bem, 1981; Buss, 2015).

3.2.5.3 Dimensions of Gender Stereotypes

Warmth and competence. Warmth and competence have been part of gender stereotype research for years (Conway et al., 1996; Eagly, 1987) and are considered two of the most important dimensions for social cognition (Fiske et al., 2007). Men are attributed attributes like competence and assertiveness while women are ascribed attributes like warmth and affection. This is explained due to the division of labor in a society based on the *Social Role Theory*: in most western societies, men have traditionally held positions of higher power while women were assigned nurturant roles which in turn resulted in the formation of stereotypes according to these roles (Eagly, 1987; Eagly et al., 2000).

Expressiveness and instrumentality. Two other well-established dimensions of gender stereotypes are expressiveness and instrumentality. It is believed that women possess more expressive attributes than men while men possess more instrumental attributes (Broverman et al., 1994). Both dimensions were originally considered to be innate attributes resulting in a societal role model (Bales & Parsons, 2014). In later literate this was reversed, and these dimensions are no longer considered to precede a role model but instead are considered to be the result of stereotypical role models (Spence & Helmreich, 1979, 1980). Experimental studies provide evidence that men are ascribed more instrumental attributes while women are ascribed more expressive attributes (Spence & Buckner, 2000).

3.2.5.4 Social Influence and Conformity

Conformity is based on the assumption that the presence of a group exerts social pressure that influences individuals' decisions based on the predominant opinions of that group (Asch, 1956; Hertz & Wiese, 2016). Conformity effects can even occur when these opinions are assumed to be false by an individual (Asch, 1956). Conformity is considered to be the result of social influence by others (M. Deutsch & Gerard, 1955) which can be further differentiated by normative social influence, which is based on the desire to comply with others' expectations, or informational social influence, which is based on information or evidence provided by others (Kelman, 1958). Literature on gender stereotypes also links conformity to gender: it has been shown that men exert more social influence compared to women, which results in more conformal behavior from others (Eagly, 1983; Lockheed, 1985). There is also literature that suggests women are more persuadable compared to men regarding conformity in particular (Eagly & Carli, 1981). Eagly (2013) resorts to her previously mentioned social role theory to explain this effect based on cultural gender stereotypes on how men and women are supposed to act: as men in most societies are attributed higher assertiveness and agency, they try to resist social influence to a greater degree than women. An alternative explanation for conformal behavior based on gender has also been examined in CASA literature: social identification effects based on the social identification theory mentioned in section 3.2.4 and the resulting preference for and conformity with opinions of one's own ingroup, in this case, the gender one identifies with (E. Lee et al., 2000; E. J. Lee, 2003; Tajfel, 1974).

3.2.5.5 Gender Stereotypes and Conformity in HCI

Most early CASA research regarding gender stereotypes employed voices as social cues, either prerecorded or synthetic ones. These studies will be discussed in greater detail in section 3.3 as they are particularly relevant for this thesis. This section will instead focus on a study conducted as part of CASA research that uses another modality. E. J. Lee (2003) wanted to assess if social reactions to computers were based on gender stereotypes without using the social cue of voice. In this study, computer gender was instead operationalized by using animated characters and text-based communication. Participants were presented with trivia quiz questions regarding either a stereotypically masculine or feminine topic (sports and fashion respectively). For each question, they were also presented with a recommended answer by either the male or female computer character, which they knew was not necessarily the correct answer. They were then given the chance to change their original answer to check for the effects of conformity. Results supported the effects of conformity previously found in human interactions: a male animated character elicited greater conformity on the masculine topic and participants conformed more to the female animated character on the topic of fashion. Additionally, men were less likely to follow a computer's suggestion for a masculine topic of sports and women were less likely to yield to the computer's influence on the feminine topic of fashion (E. J. Lee, 2003).

3.2.6 Personality Traits

One of the most recognized and consistent behavioral patterns in psychology literature is the so-called *similarity attraction* (Byrne et al., 1967): the more similar people are to one another, the more attracted they are to each other. One aspect this extends to is the dimension of personality. People tend to attribute positive characteristics to other people with a similar personality, are more attracted to them and prefer to interact with them (e.g., Blankenship et al., 1984; Byrne et al., 1967; Duck & Craig, 1978). Speaking from an evolutionary perspective, personality is an inherently human trait (Funder, 1997). Thus, categorizing other humans based on personality types (e.g., dominance, extroversion, etc.) reduces cognitive load during social interactions and allows to quickly and accurately predict the behavior of others (e.g., Digman, 1990). This is, once again, crucial for survival, as other humans represent both potential threats as well as potential mating partners.

3.2.6.1 Personality Traits in HCI

The effects of similarity attraction based on personality traits have been examined in a series of text-based CASA experiments (Y. Moon & Nass, 1996a, 1998, 1996b; Nass, Moon, et al., 1995) in which participants were asked to complete a desert survival ranking problem on one computer while a second computer provided them with additional information and ratings during the task. The phrasing of messages displayed by the second computer was manipulated to convey either dominance or submissiveness. Congruent with the similarity-attraction hypothesis, participants classified as dominant preferred a dominant computer in a subsequent evaluation while submissive participants gave better ratings to the submissive computer (Y. Moon & Nass, 1996a; Nass, Moon, et al., 1995). It should be noted that participants' personality was classified based on a single self-report measure: the masculinity subscale of the Bem Sex-Role Inventory (Bem, 1981). In one of the studies, the computer changed its personality over the course of the experiment to either become more similar to that of the participants, become more dissimilar over time, or stay the same (Y. Moon & Nass, 1996a, 1996b). Results revealed that participants preferred a computer that become more similar to them over time, rating it as more attractive. The authors argue that this effect can be explained by the gain-loss theory. In line with the self-serving bias, participants whose personality matched that of the computer were more likely to blame a failed interaction on themselves while participants who worked with a computer displaying a dissimilar personality to their own blamed the computer significantly more often for the failed interaction (Y. Moon & Nass, 1998). In a later study, evidence for *complementary attraction* in human-computer interaction was found (Nass et al., 2001). Extroverted participants preferred an onscreen character displaying introverted messages during a desert survival problem while introverted participants preferred an onscreen character that presented them with extroverted messages (Nass et al., 2001). Based on the principle of *consistency attraction*, it was also shown that participants preferred onscreen characters that are consistent in their display of extroversion and introversion via verbal (phrasing) and non-verbal (posture) cues and followed their advice more often in a desert survival problem (Isbister & Nass, 2000; Nass et al., 2001).

3.2.7 Other Early CASA Studies

Additional CASA studies have been conducted regarding various topics of human-computer interaction. This section aims to give a brief overview of the media equation effects found in these studies.

Assignment of roles and effects of specialization. To check if the assignment of a role is sufficient to induce attributions usually exclusive to human-human interaction, Reeves and Nass (1996) had participants watch footage (news and entertainment footage) on three differently labeled TVs: a ‘news TV’, an ‘entertainment TV’ and a ‘generalist TV’ with both labels. Participants in the specialist condition watched news footage on the ‘news TV’ and entertainment footage on the ‘entertainment TV’ while participants in the generalist condition watched both types of footage on the ‘generalist TV’. Even though the footage was identical for all conditions, the entertainment footage was assessed as more amusing and enjoyable and the news footage was assessed as more informative and engaging in the specialist condition, proving the assignment of roles as an effective manipulation even for a technological device such as a TV (Reeves & Nass, 1996). A similar study was later conducted in an e-commerce context to evaluate if recommendations on a shopping-website are more effective when given by an appropriate specialist computer compared to a generalist. Results revealed that usage of a specialist computer resulted in increased trust and reduced purchase decision time compared to a generalist computer giving identical recommendations (Koh & Sundar, 2010).

Frustration. To examine if frustration caused by technology use can be reduced by having the computer actively support the user and express empathy and sympathy, an experiment was conducted by Klein et al. (1999). Results revealed that a computer can ease frustration by actively ‘listening’ to users’ concerns and expressing sympathy accordingly similar to effects in human-human interaction (Klein et al., 1999).

Humor. After receiving jokes and humorous comments from a computer during a desert survival problem, participants showed more social behavior towards the computer (smiling and laughing) and rated it to be more likeable and competent compared to participants in a control group (Morkes et al., 1998).

Distance. The effects of perceived distance in human-computer interaction have also been examined. Participants have been shown to distort their answers to present themselves more positively when a computer they interacted with is believed to be located further away from them. In addition, the effects of persuasion are stronger when the computer was perceived to be in closer proximity (Y. Moon, 1998).

3.2.8 Expanding CASA to other Technologies

As mentioned, initial studies examining media equation effects were mostly conducted using desktop computers. More recent research also transferred the CASA paradigm to other technologies. This section provides an overview of CASA research focusing on technologies other than desktop PCs.

3.2.8.1 Mobile Devices

In line with recent technological developments, the CASA paradigm has also been transferred to smartphones, arguing that their technological capacities for sending social cues visually as well as via sound, notifications, or even vibration exceed those of desktop PCs used in the 1990ies. Consequently, an argument can be made that modern smartphones come closer to the definition of social actors than desktop computers (Carolus, Muench, et al., 2019). Based on previous CASA research, Kim (2014) demonstrated that results from social psychology literature can be transferred to smartphones. In human-human interaction, a specialist is usually more trustworthy and exerts more influence on somebody in his field of expertise compared to a generalist. This effect has already been examined in human-computer interaction and is ascribed to media equation because specialization is considered a social cue in this context (Koh & Sundar, 2010). Kim (2014) had participants interact with ‘specialist’ and ‘generalist’ smartphones to see if specialization influences the reception of subsequent mobile advertisements. Both smartphones, as well as Apps that act as specialists in their field, resulted not only in more trust in advertisements but also in an increased intention to buy the advertised products (K. J. Kim, 2014).

Based on the experimental design established by Nass et al. (1999) to examine the interviewer-bias in human-computer interaction (see section 3.2.2.3), a more recent study aimed to replicate these findings using smartphones instead of desktop PCs (Carolus, Schmidt, Schneider, et al., 2018). Following the original procedure, a ‘tutor phone’ first presented participants with 20 facts. The topic of Canada was chosen for the facts as a topic of low emotional involvement and low prior knowledge. Following the tutor session, participants were asked several questions about these facts. Regardless of their answers, all participants received the same feedback from the smartphone. The interaction then concluded with participants evaluating the smartphone in one of three locations: (1) the phone itself, (2) a different phone, or (3) their own smartphone, replacing the pencil-and-paper questionnaire employed in the original study. Contrary to previous results, there was no significant difference between the evaluation on the same smartphone that had presented the facts and the evaluation on a second, unfamiliar smartphone. However, there was a significant difference between the evaluation of the smartphone that had presented the facts and one's own smartphone, with the latter being significantly worse, providing at least partial support for an interviewer-bias in human-smartphone interaction (Carolus, Schmidt, Schneider, et al., 2018).

Research on gender stereotypes and minimal identity cues was also transferred to smartphones. A similar study aimed to examine whether smartphones can elicit stereotypical responses and conformal behavior using minimal gender cues, in this case, gender-specific colors. Participants were presented with five social dilemmas by a text-based smartphone equipped with either a blue (representing a stereotypical 'male' color) or a pink (representing a stereotypical 'female' color) sleeve. Each dilemma had two possible solutions for participants to choose with the smartphone always arguing for one of the two options. After this interaction, participants were asked to rate the smartphone regarding its masculinity and femininity as well as its competence and trustworthiness. Results revealed that participants ascribed more female attributes to a smartphone presented with a pink sleeve and more masculine attributes to a smartphone presented with a blue sleeve, confirming the adoption of gender stereotypes. In addition, the 'male' smartphone was evaluated to be more competent, and subjects followed its recommendations more often. Male participants also rated the smartphone presented with a blue sleeve as more trustworthy while no significant differences were found for female participants (Carolus, Schmidt, Muench, et al., 2018).

3.2.8.2 Agents with Virtual and Physical Embodiments

By this point, it has been well established that disembodied technological devices can effectively express human characteristics solely through minimal social cues. Consequently, CASA research has also been extended to technologies like social robots and intelligent virtual agents that incorporate additional cues via virtual or physical embodiment (Lugrin, 2021). What differentiates these technologies from initial CASA research is the addition of a more humanlike appearance as an additional locus of attention for users (Cassell, 2001). With an appearance also come additional communication mechanisms like gaze, facial expressions, gestures, turn-taking, and body orientation. These additional, appearance-based cues have been shown to work effectively in human-agent interaction, and forms of embodiment can improve social outcomes (Mumm & Mutlu, 2011). As this dissertation is focused on the technology of disembodied smart speaker devices, based on this distinction and for reasons of brevity, this section will be limited to a brief overview. A few examples will be given to demonstrate that media equation effects can also be extended to devices containing additional anthropomorphic features and cues, but they are not the focus of this thesis.

3.2.8.2.1 Embodied Conversational Agents

The CASA paradigm has also been extended to various types of intelligent virtual agents such as embodied conversational agents (ECAs). ECAs are graphically embodied agents usually meant to resemble humans to enable ‘face-to-face’ communication with users (Cassell et al., 2000). It has been shown that people engage in social dialogue and use politeness strategies when interacting with ECAs, like following small talk etiquette (Bickmore & Cassell, 2001) and that their perceptions of these agents are sensitive to their interaction styles (Bickmore & Cassell, 2005; Kopp et al., 2005). Adopting the experimental design of Nass et al. (1999), Hoffmann et al. (2009) examined the interviewer-bias towards an ECA. After an interaction, either the ECA itself asked for an evaluation or a paper-and-pencil questionnaire was used. Similar to previous results for computers and smartphones, participants rated ECAs significantly more positive in a direct evaluation (Hoffmann et al., 2009). The adoption of gender stereotypes has also been shown in interactions with virtual agents, where a study revealed cross-gender effects for persuasion by a virtual character. Male participants were persuaded by a male speaker while female participants were persuaded by a female speaker (Zanbaka et al., 2006). Participants also generally favored an agent that agreed with them during their interaction (Nakanishi et al., 2003).

3.2.8.2.2 Social Robots

Research has also focused on the adoption of social norms in HRI, with overall results supporting media equation effects. Multiple studies provided evidence of participants adopting the social norm of politeness when interacting with robots expressing polite behavior (Nomura & Saeki, 2009; Salem et al., 2013; Srinivasan & Takayama, 2016). As with any other technological device that interacts socially with its users, robots risk being perceived as impolite if they do not follow established politeness strategies. Another study revealed that elderly subjects recognized different politeness strategies employed by robots in a care-taking context (Hammer et al., 2016). Sandoval and colleagues (2016) had participants play with a bribing robot and subsequently examined the reciprocal behavior towards it. After the interaction, the robot asked for help to fill its visual database. To do so, participants were asked to read out icons from a note. Results suggest that prior dishonest behavior of the machine leads to a decrease in prosocial behavior towards it (Sandoval, Brandstetter, & Bartneck, 2016). Participants were also shown to express emotional reactions and empathic concern when watching video material of a robot being tortured (Rosenthal-von der Pütten et al., 2013). Employing another objective behavior measure, Reuten et al. (2018) were able to show participants' pupils responded similarly to robotic and human emotional facial expressions (Reuten et al., 2018). The adoption of gender stereotypes has also been examined in human-robot interaction. Results indicate that certain features such as long hair and higher-pitched voices are associated with female robots (Eyssel & Hegel, 2012). Additionally, the acceptance of 'gendered' robots has been shown to be based on gender stereotypes. Female robots are more accepted in domains that are usually associated with females such as health care white male robots are more accepted in tasks related to stereotypical male domains such as security (Tay et al., 2014). Studies conducted in the area of social robotics also revealed that participants ascribed gender-specific attributes like warmth and competence to robots purely based on their appearance (Mieczkowski et al., 2019) and that men and women rate robots differently based on their gender in regards to credibility, trustworthiness, and engagement (Siegel et al., 2009).

3.2.9 Evidence against Media Equation

For a complete view of CASA and media equation research and its limitations, negative results must be considered as well. Only a very limited number of studies containing results contrary to the media equation have been published, which could be the result of the well-established publication bias both in psychology literature (Francis, 2012) and in HCI literature (Dragicevic, 2016). As for the studies that have been published, they suffer from clear methodological weaknesses that are explained in further detail in this section.

In an experimental study conducted by Goldstein et al. (2002), the usage of politeness norms when interacting with personal digital assistants (PDAs) and early smartphones (grouped together by the authors under the categorization of 'small computers') was examined. Participants were asked to complete a series of tasks on a PDA or a smartphone after which they had to rate the device in 11 different categories via a paper-and-pencil questionnaire. In addition, participants were asked to write down any number of 'likes' and 'dislikes' about the device they interacted with. For half of the participants, the device was in the room while it was evaluated while for the other half it was removed prior to the evaluation. Based on previous media equation literature, the researchers expected the device to be rated better when it was present during the evaluation due to politeness norms (Reeves & Nass, 1996). However, Goldstein et al. (2002) found the opposite to be the case: devices received more entries in the 'dislikes' category when present during the evaluation and more entries in the 'likes' category when outside the room. This was interpreted by the researchers as evidence that contrary to the media equation, people are not polite towards small computers (Goldstein et al., 2002). However, there are several methodological problems with this study. Both PDAs and one single smartphone were used interchangeably in the study under the guise of small computers. This variation in devices represents a confounding variable. Combined with the small sample size of 11 and 14 participants per experimental condition, this resulted in extremely small and uneven cell sizes thus violating the assumptions of the MANCOVA employed by the researchers to analyze their data. Additionally, the researchers did not control for the total number of likes and dislikes mentioned by each participant thus subjects who wrote down more opinions than others resulted in biases in the collected data. Also, evaluation in either condition was not given to the device directly but was measured using a paper-and-pencil questionnaire with the only difference being the presence of the device while filling in the questionnaire. This methodology does not constitute an interviewer-bias in either human-human interaction (Finkel et al., 1991) or in HCI (Nass et al., 1999) and thus cannot be compared to these results.

There is only one publication focused on replicating traditional CASA experiments to test the media equation with children as participants (Chiasson & Gutwin, 2005). The two experiments chosen for replication by the authors were those conducted to examine the effects of praise and flattery (see section 3.2.2.4; Fogg & Nass, 1997b) and team affiliation (see section 3.2.4.3; Nass et al., 1996). To account for the younger participants, both studies were adopted by decreasing the total human-computer interaction time and simplified measures. While results for both studies indicate a trend in the expected direction for media equation effects, Chiasson and Gutwin (2005) were unable to reproduce the significant effects found in previous studies. However, there are clear methodological deficiencies in this study, as both the manipulations and the measures were scaled back significantly to make them more child friendly. The authors themselves also note that many of the children participating in the study were more focused on the aspects of the guessing game and subsequently ignored feedback given to them by the computer as they realized it was irrelevant to the continuation of the game (Chiasson & Gutwin, 2005). Thus, results about media equation effects for children remain inconclusive.

3.2.10 Limitations of CASA and Media Equation

Despite the lack of unambiguous negative results in literature, it should be noted that there are certain limitations to the CASA paradigm and where it can be applied. Nass and Moon (2000) clearly state that CASA cannot be transferred to every technology, but instead indicate two important factors that need to be fulfilled: (1) a technology must be able to send enough social cues to elicit social reactions, and (2) it must be perceived as the source of those cues (Nass & Moon, 2000).

Social cues. Nass and Moon (2000) never state how many or what cues exactly can be considered ‘enough’ cues, meaning there is no objective measure in CASA literature indicating which devices can be considered as social actors. This is further complicated when considering that the perception of social potential in technology can also differ between persons based on individual factors such as age and experience (Waytz et al., 2010). There is research suggesting that technology displaying multiple social cues can elicit even stronger social reactions (Appel et al., 2012; Burgoon et al., 2008; Ghazali et al., 2018; Lombard & Xu, 2021; Tung & Deng, 2007) but there is no literature with a clear focus on either a distinction between different social cues or an assessment of their quality. Lombard and Xu (2021) suggest categorizing social cues in primary (e.g., facial expressions, gaze, gestures, voice, and humanoid shape) and secondary (e.g., human size, motion, and language use) social cues as well as social signals (e.g., responsiveness, interactivity, perceived personality, and identity) with decreasing quality along these categories. As of right now, there is no comparative research making this distinction.

Source orientation. For a technological device to be perceived as a social actor, it must also be perceived as the autonomous source of the social cues (Nass & Moon, 2000; Nass & Steuer, 1993). This is a crucial differentiation as technology was originally only considered as a medium or channel for human-human communication (Gunkel, 2012). Only if a technological device is identified as the source of communicative actions or contextual social cues instead of relaying them from one person to another, it can be attributed agency thus leading to social reactions (Solomon & Wash, 2014; Sundar & Nass, 2000). Solomon and Wash (2014) argue that a user's orientation is towards the technological device by default. The reasons the authors give for this is the proximity between the user and the device combined with the agency a device displays (Solomon & Wash, 2014). Only when users were explicitly told that they were interacting with a computer instead of a human and thus are not oriented towards the device as a source, the degree of social reactions they displayed was different (Shechtman & Horowitz, 2003). Three initial studies were conducted with the issue of source orientation in mind. As mentioned in section 3.1.3.2, Sundar and Nass (2000) employed a laboratory experiment to rule out that participants are merely addressing the programmer as a source rather than the technological device itself. Results revealed that participants categorically ruled out addressing the programmer, which still leaves the question to which parts or features of the devices the notion of self is ascribed to. Takeuchi and Katagiri (1999) found that different windows on the same computer are not perceived as separate sources and participants ascribed a notion of self only to the computer itself. However, in an experiment by Nass and Steuer (1993), participants perceived different voices coming from the same computer as separate entities and evaluated them differently making it unclear on what basis a source is determined by users. The issue of source orientation is of special interest for smart speakers, as most modern smart speaker devices adopt a distinct 'persona' (in most cases even with a dedicated name, e.g., Alexa or Siri) in addition to the hardware component that serves as an additional source for users to orient themselves toward (Guzman, 2019). The findings of previous CASA research suggest that people treat technological devices like other humans and thus it is assumed that they apply some notion of self or individuality directly to these devices. Congruent with source orientation literature (Eckles et al., 2009; Sundar & Nass, 2000), subsequent research also provided support for source orientation towards ECAs (Hoffmann et al., 2009) and toward social robots (Straub et al., 2010). However, compared to technologies previously examined in source orientation research, smart speaker devices present the additional problem that they consist of two potential sources: the device itself (hardware) and the voice-based assistant (software) (Guzman, 2019). Consequently, interactions with smart speakers can both follow a physical approach (e.g., pressing buttons on the device itself to change volume or mute) and a voice-based approach. In a recent study, it has been shown that participants perceive themselves as communicating directly with the technology of voice-based mobile virtual assistants. Nonetheless, when asked about the interaction, participants were divided about the voice they had been communicating with. Some participants ascribed it to the hardware (in this case a mobile device) and others to the software (the virtual assistant), leaving the question of source orientation open as of now (Guzman, 2019).

3.2.10.1 Other Limitations

Long-term research. Almost all CASA research exclusively examines initial reactions to technological devices in an experimental setting thus allowing limited conclusions about the lasting effects of these results. As interactions with these devices often last considerably longer in a real-world context, there is potential for these effects to change over time. As of this moment, only one longitudinal study regarding CASA effects has been published (Pfeifer & Bickmore, 2011). The authors had participants use a system designed for tracking exercise over a period of 40 days to assess the adoption of a social desirability bias when reporting their fitness progress. Subjects were split into two conditions: high personalization interface (anthropomorphic conversational character) and low personalization interface (text-based). Results indicate that the self-reported progress was more accurate when given to a highly personalized interface compared to the text-based version. In addition, the social responses changed during the duration of the experiment. Participants' engagement with the interface was influenced by their walking behavior. The more they walked the more they used the tracking system, and this effect became smaller for the low personalization interface and larger for the high personalization interface over time. Thus, the authors conclude that social reactions that occur in human-computer interaction can both increase and decrease over time depending on the social cues it adopts (Pfeifer & Bickmore, 2011).

Comparative studies. Almost all CASA research is conducted in the same way: a human interaction partner is replaced with a technological device and social reactions and evaluations are measured. However, to draw a comparison between social reactions in human-human interaction and human-computer interaction, both these conditions would need to be assessed under identical conditions. Three similar studies operationalized this procedure by telling one group of participants that they would interact with a computer and the second group of participants that they would interact with another human. To ensure identical conditions, both groups interacted with their counterpart through another computer and were explicitly told that the entity they are communicating with is a computer or a human respectively (E. J. Lee & Nass, 1998; Morkes et al., 1999; Shechtman & Horowitz, 2003). Participants in all three studies responded socially towards both a computer and another human, by showing conformal behavior based on their ingroup (E. J. Lee & Nass, 1998), reacting to humorous comments and jokes (Morkes et al., 1999) and emphasizing relationship goals (Shechtman & Horowitz, 2003). However, in all three studies, the degree of social reactions differed between both experimental groups. The condition perceived as mediated communication with another human (CMC) led to more and stronger social reactions compared to the HCI conditions. This led the authors of one study to conclude that the media equation is in fact a *media inequality* by direct comparison with human-human interaction (Shechtman & Horowitz, 2003). Still, these results do not indicate that people don't respond socially to technological devices but rather that they seem to do so to a different degree compared to social reactions towards human interaction partners.

3.2.11 Extending and Modernizing the CASA Paradigm

Apart from a clear lack of long-term studies, recent literature identifies two areas of research that need to be considered in extending and modernizing the CASA paradigm. While various CASA studies incorporated participants' gender into examining conformity effects (e.g., E. J. Lee, 2003; Nass et al., 1997) and sometimes controlled for it, there is a clear lack of assessing additional individual factors and their influence on media equation effects (Lombard & Xu, 2021; Nass & Moon, 2000). One of the few other factors that has been included in CASA research was previous experience both with computers (Johnson et al., 2004; Johnson & Gardner, 2007) and social robots (Horstmann & Krämer, 2019).

Additional factors that have been proposed but not examined in detail include anthropocentrism, user personality, age, suspension of disbelief, technology acceptance, and tolerance of imperfection among others (Gambino et al., 2020; Lombard & Xu, 2021). Thus, Lombard and Xu (2021) propose an extension of the CASA paradigm by emphasizing the importance of user-side factors that could potentially influence media equation effects. Additionally, Gambino et al. (2020) argue that people have accumulated more knowledge and experience with media agents and technologies have advanced greatly in their capabilities since the initial wave of CASA research focused mainly on desktop computers. The authors suggest that the interactions between humans and these modern media agents have subsequently changed as well as they are more integrated into our everyday life than ever before (Gambino et al., 2020). One of the areas that exhibits the most technological progress when it comes to human-machine communication is voice interaction. Voice-based systems have made great strides since the inception of the CASA paradigm, yet there is a lack of research focused on social reactions towards smart devices generally and smart speakers specifically (Seaborn et al., 2021). This gap in research is especially surprising as speech is considered the fundamental means and main channel of human-human communication (Pinker, 1995) while also providing the foundation for human-voice assistant interaction as they are almost exclusively controlled by voice. To examine social reactions towards these devices in detail, the underlying channel of communication used in this interaction must be explored in detail. Consequently, speech will be the focus of the next section.

3.3 Speech

Humans have evolved a system of vocal interaction superior to any other animal (Bickerton, 2017; Gardiner, 1932; Hauser, 1996; Hauser et al., 2002) and as a result, speech is the fundamental means of human communication. Even though other forms of communication like writing, facial expressions, or gestures can be just as expressive, humans across all cultures mainly use speech to build relationships (Pinker, 1995). From a linguistic perspective, speech is defined as words that are produced through the combination of vowels, constant sounds, or phonetics (Fitch, 2017)[2].

3.3.1 The Importance of Speech in Human Evolution

The importance of speech is evident in the developments humans went through to better understand and process speech. According to the modern version of Darwin's theory of evolution, random mutations are introduced in the genetic makeup of offspring. This leads to traits that are selected for based on how beneficial they are for mating and survival (Darwin, 1859; Dawkins, 1989). These mutations accumulate very slowly and ultimately spread through members of a species resulting in cognitive, behavioral, and physical traits. For humans, many of these traits are focused on communication, especially through facial expressions, body language, and sounds (Boaz & Almquist, 1996; Cartwright, 2000). By developing a sophisticated web of 43 facial muscles (Ekman & Friesen, 1978) humans can display over 6000 communicative expressions (Bates & Cleese, 2001). Even more important for complex communication was the development of the larynx and the enlargement of the vocal tract. These adaptions allowed humans to produce a variety of sounds which are the base of most modern languages (Laitman, 1984; P. Lieberman, 1998). Some scientists even assume that the existence of the human species can be traced back to the ability to speak, as this was essential for the development of social life and culture (Hauser et al., 2002). In ancient times, it was vital to bond with other humans for hunting and mating purposes to make sure that the species will survive natural selection (Nass & Gong, 2000). As speech was easy to perceive and a highly accurate cue of humanness, it became a central tool for interactions, such as building and maintaining social bonds, describing the environment, and expressing internal processes (Kohler, 2017). It is therefore widely recognized in literature that "spoken language is the most sophisticated behavior of the most complex organism in the known universe" (Moore, 2007, p. 419). Looking at humans today, there is even more evidence to corroborate the importance of speech for humans. Even before being born a fetus can already

[2] While in linguistic literature speech and language are distinguished from each other, for the sake of brevity and in line with the scope of this thesis, both terms are considered as verbal tools of expression of humans and are used interchangeably here.

differentiate between its mothers' voice and other voices (Kess, 1992). Immediately after birth, infants react differently to anything that sounds like speech compared to all other sounds (C. Moon et al., 1993). Only four days after being born, their brains start differentiating between their mother tongue and other languages (Peña et al., 2003). 22 days after being born, infants tend to pick up speech mainly with their right ear which corresponds to the left half of the brain. This part of the brain has significant advantages in processing one's mother tongue, foreign languages, and even speech that is played backwards (Cutting, 1974). At the end of their first year, infants can already use speech sounds to communicate meaning, and starting at 18 months of age, infants learn about eight to ten new words each day and keep doing so until their youth (Pinker, 1995). Until their youth, humans are also capable of recognizing speech up to a rate of 50 phonemes per second, while most other sounds become unrecognizable at about 20 phonemes per second (Slobin, 1971). In addition, even humans with an IQ as low as 50 or a brain that only weighs 400 grams can still speak and understand language (Nass & Gong, 2000; Slobin, 1971).

3.3.2 From Speech to Voice

Speech and voice are different phenomena, even if the terms are often used interchangeably in literature (Pittam, 1994). Both terms are interconnected with speech being one specific aspect of voice. It refers to the linguistic content and contains words, grammar, syntax, and phonetics. Voice on the other hand is considered the medium of speech and can convey additional non-linguistic information such as emotions, gender, age, and personality (Pittam, 1994). From an evolutionary perspective, voice is fundamental for communication and collaboration within and between species in nature (Hare, 2017), and for humans, voice is the main channel for communicating with others (Flanagan, 1972; Nass & Gong, 2000; Schafer, 1995). Considering the importance of voices in human evolution, speech is capable of much more than just transferring words from a speaker to a listener. Humans can not only process and understand speech, but they also developed voices capable of transmitting a range of socially relevant cues that can be received and analyzed by other humans to act upon (Nass & Gong, 2000). Other humans can represent a potential danger, or they might represent an opportunity to mate. Either way, as soon as one recognizes something that is a clear indication of humanness, the ideal strategy is to save cognitive capacities by assuming that its source is human. This was the optimal strategy during almost all evolutionary history, as only other humans could produce speech. Over the course of thousands of years, this led to a simple rule: speech equals human (Massaro & Cohen, 1995). In addition to recognizing voices as human, it is also important to correctly interpret the social information contained in a voice. It is not only used to communicate language (or linguistic cues) but also transports paralinguistic cues in the form of socially relevant information. Clues about the speaker are ingrained in each individual voice and contain information about their sex, ethnicity, locality, personality, and emotions among others (Pittam, 1994). From an evolutionary perspective, information concerning biological sex is of great importance and voice is an essential way to identify it (Slobin, 1971). As a result, voices are rapidly analyzed based on factors such as pitch and cadence to classify them as male or female (Mullennix et al., 1995). Starting at the age of six months, infants can already recognize voices as male or female and between 11 and 14 months, they become capable of associating the sex of a voice to pictures of humans, for example assigning a higher-pitched, female voice to the picture of a woman (Miller et al., 1982). Biological sex is of such importance to the identification process, that when questioned about similarities between different voices, the most important criterion named is whether a voice is male or female (Singh & Murry, 1978). The classification of a voice as male or female even influences the interpretation of everything that is said by that voice (T. M. Holtgraves, 2013; Strand, 1999). Further research on voice-activated stereotypes is presented in section 3.3.4. Other human characteristics are also transported by parameters of voices. For example, rate of speaking as well as volume influence the interpretation of the speaker’s personality and emotional state (B. L. Brown et al., 1973). Someone talking slowly and quietly is considered bored and introverted while someone talking fast and loudly is perceived as excited and extroverted (Aronovitch, 1976). In addition, social information is not only transmitted by properties of the voice itself but also by choice of words. Listeners constantly and automatically extract all these relevant social cues from voices in human-human interaction (Pittam, 1994).

3.3.3 Text-to-Speech and Synthetic Voices

All of this led to the innate rule to perceive everything that uses a voice to transmit speech as a human being (Nass & Gong, 2000) which proved to be irrefutable throughout human history as only other humans were able to produce complex speech (Massaro & Cohen, 1995). With the rise of modern technology, this distinction is no longer valid. By today’s standards, any device with an integrated speaker and almost any device that can be connected to external speakers is able to produce voices, usually in the form of synthesized, artificial speech called text-to-speech (TTS). Another option is the utilization of recorded human speech that is played back by the device. Because recorded speech required larger amounts of disc space and is less dynamic compared to synthetic speech, it is used less often and mostly in special cases where little linguistic material is needed (McTear et al., 2016; Olive, 1997). TTS-systems have become rather efficient when it comes to communicating content, but even the most advanced TTS-systems do not achieve the quality and structure of natural human speech. They are prone to unexpected pauses, flawed intonation, and interruptions between syllables making them recognizable as non-human (Kamm et al., 1997). In addition, TTS-voices originating from a disembodied device are considered to produce ‘doubly disembodied’ speech (K. M. Lee & Nass, 2004). The first degree of disembodiment is a result of no human speaker being present and the second degree of disembodiment is due to a lack of association between paralinguistic cues and the source. These problems presented an important question in CASA research: do people react to these flawed and disembodied voices the same way they would to human voices clearly originating from a human speaker? Results indicate that, regardless of these flaws, TTS-voices are still recognized as voices, and additional information such as sex is usually correctly identified (e.g., E. Lee et al., 2000; K. M. Lee & Nass, 2004; Morishima et al., 2002; Nass & Lee, 2001) even though a technical device is incapable of being male or female. From an evolutionary perspective, this can be explained: humans have developed a very broad and liberal definition of speech. Even syllables that are senselessly combined with each other and speech played backwards are still recognized and processed as speech (Scherer et al., 2001; Slobin, 1971). While the human brain has different areas to produce speech and to understand speech, it does not inherently differentiate between human and synthetic speech (Nass & Brave, 2005; Nass & Gong, 2000). Anything that even vaguely resembles speech is recognized as a communicative act and interpreted as such (Grice, 1975). Even if all cues indicate that a TTS-voice is not human, the human brain automatically processes its speech as it would any other communicative act (Nass & Brave, 2005). Relevant studies focusing on the social cue of voice are described in the next section.

3.3.4 Previous Research

Sections 3.3.1 and 3.3.2 illustrate the relevance of speech and voices for human-human interaction. The importance of this form of communication led to multiple CASA studies that transferred it to human-machine interaction. As the studies presented in this section provide the foundation for the experiments conducted as part of this thesis, some of those experiments are described in greater detail.

3.3.4.1 Voice in HCI

Social psychological literature explored many gender stereotypes over the years. The one that was first transferred to HCI is related to the role of men and women in academics. It has been shown that women are evaluated more critically when they teach in fields traditionally regarded as 'male domains' such as technology and business (Sax, 2001). In an initial study conducted by Nass et al. (1997) to assess whether these gender-stereotypic responses can also be observed in human-computer interaction, participants interacted with a series of desktop computers in a learning context. They were presented with ten facts about a topic that is considered typically male (in this case computers and technology) and ten facts about a topic considered typically female (in this case love and relationships) by a 'tutor computer'. For all facts, the computers presented them using either a male or a female prerecorded voice. Afterwards, participants were tested on their knowledge of these facts on a second, separate computer via a quiz. Following this quiz, participants interacted with a third computer, the 'evaluation computer', again using either a male or a female voice. Participants were assigned to a random combination of male or female 'tutor computer' and male or female 'evaluation computer', allowing for a total of four conditions. The 'evaluation computer' then proceeded to rate the 'tutor computer' based on how well it prepared participants for the quiz. Regardless of the actual results of the quiz, this rating was always a positive one. Lastly, participants were asked to rate both the 'tutor' as well as the 'evaluation computer' regarding their competence, sympathy, and informativeness. Results revealed that participants used the same gender stereotypes that are used for real professors in social psychology literature: the ‘tutor computer’ was rated to be significantly more competent and sympathetic after it had been praised by a male ‘evaluation computer’. Additionally, a female ‘evaluation computer’ was perceived as less friendly because the dominant role of an evaluator was considered inappropriate for a female voice. A female ‘tutor computer’ was also rated to be more informative regarding female topics while a male tutor was rated to be more informative regarding male topics (Nass et al., 1997).

In a study conducted by E. Lee et al. (2000), another classic social-psychological experiment was transferred to human-computer interaction. Participants were presented with a series of social dilemmas with two options to choose from and were given recommendations on how to solve these dilemmas by a synthetic TTS-voice. Two different TTS-voices were used in the experiment. The only difference between them was their frequency, resulting in one deeper, male-sounding voice and a higher, female-sounding voice. Other features of the voices such as volume and speaking tempo were kept identical, same for the recommendations they gave participants. Results revealed effects similar to previous social psychology research: a male voice had a greater influence on participant’s decisions and was rated to be more trustworthy and socially attractive. Additionally, both voices triggered a social identification process. Female participants showed more conformal behavior towards the female voice and male participants showed more conformal behavior towards the male voice (E. Lee et al., 2000). In a variation of this experiment, instead of varying the gender of the voices, they were instead manipulated to convey an extroverted or introverted personality based on paralinguistic attributes. Results revealed a similarity-attraction effect with extroverted participants preferring the extroverted voices and introverted participants following the recommendations of the introverted sounding voice (K. M. Lee & Nass, 2003).

The results produced by Nass et al. (1997) were later transferred from a learning context to a commercial context in a second study (Morishima et al., 2002). Using an auction website, participants were presented with four different products from typically male (in this case an encyclopedia of guns) and typically female categories (in this case an encyclopedia of sewing). Products were introduced by either a male or a female synthetic TTS-voice. Each voice presented both one male and one female product. As in all previous studies, the spoken words were identical for both voices. Results once again revealed the application of gender stereotypes: Product descriptions were perceived as more credible when the voice presenting the product matched its ‘gender’ (e.g., a male voice was rated to be more credible presenting the encyclopedia of guns). Again, a social identification process was observed. Female participants generally rated product descriptions to be more credible when presented by a female voice, male participants did the same for male voices (Morishima et al., 2002).

As voice technology was not as advanced as it is today when these studies were conducted, synthetic female voices were usually of lower quality when compared to synthetic male voices as they were harder to synthesize due to certain vocal aspects of female voices (Olive, 1997). Consequently, Mullennix et al. (2003) wondered if previous results regarding the adoption of gender stereotypes in human-computer interaction might be influenced by the lower quality of female synthesized voices when compared to prerecorded speech. In a laboratory experiment, they presented participants with persuasive arguments made by prerecorded or synthetic voices, both either male or female. However, no differences between prerecorded and synthetic voices were found for persuasiveness or the ratings of the voices. Thus, the authors concluded that gender stereotypes are adopted in human-computer interaction regardless of whether computers used prerecorded or synthetic voices (Mullennix et al., 2003).

In addition to gender stereotypes, personality traits have also been examined in human-computer interaction using voice output (Nass & Lee, 2001). Extroverted and introverted participants were presented with either an extroverted or introverted synthesized, gender-neutral voice on an e-commerce website, where the voice read book reviews to participants. Results indicated that participants not only recognized personality cues embedded in the synthesized speech but also displayed similarity-attraction in their ratings of the voice and the book reviews presented to them by the voice (Nass & Lee, 2001). In a follow-up study using the same e-commerce context, the multiple source effect was examined using synthesized voices (K. M. Lee & Nass, 2004). In addition to a single voice presenting participants with five positive reviews about a single book, another group of participants heard these five reviews from five different voices instead. Consistent with the multiple source effect, results revealed that participants in the multiple voice condition felt more social presence and that multiple synthetic voices were more persuasive than a single voice. In a second experiment that followed the same basic procedure, a learning manipulation was added that informed participants about the artificiality of the synthesized voices and how there are produced. Even though the artificiality was pointed out, the same effects observed in the first experiment persisted. Thus, the authors argue that participants show social reactions (in this case to multiple perceived sources) even if they know they are just listening to multiple synthesized voices created artificially on a computer (K. M. Lee & Nass, 2004).

3.3.4.2 Voice in Human-Smartphone Interaction

In a more recent series of experiments, the basic ideas of Nass et al. (1997) were transferred to human-smartphone interaction and replicated using this more advanced technological device (Carolus, Schmidt, Muench, et al., 2018; Carolus, Schmidt, Schneider, et al., 2018; Carolus, Muench, et al., 2019). In a study examining gender stereotypes based on Nass et al. (1997), Carolus et al. (2016) had participants interact with a smartphone that presented them with facts about the same two stereotypically male and female topics used in the original study: technology and computers as the male topic and love and relationships as the female topic. Again, identical to the initial study, the smartphone used either a male or a female synthetic voice to present the facts. Results indicate that female participants rated a ‘male’ smartphone to be more competent, useful, and knowledgeable. They also rated its performance to be better compared to a ‘female’ smartphone. This cross-gender effect of social influence can also be found in social psychology (Eagly, 1978) and confirms the application of gender stereotypes in human-smartphone interaction, although limited to female participants (Carolus et al., 2016). In a follow-up experiment conducted by Carolus et al. (2019), this time based on the social norm of politeness in combination with gender stereotypes, participants were told that the purpose of the study was to develop a new, heavily personalized internet search algorithm. To improve the algorithm, a smartphone asked participants increasingly intimate information using either a male or a female voice. Following the self-disclosure interaction, participants were asked to evaluate the smartphone while it allegedly calculated the usefulness of the disclosed information in the background. After the evaluation, the smartphone feedbacked participants either politely (thanking them for the information provided) or impolitely (telling them the information they provided was useless) based on the experimental condition. Participants were then asked to evaluate the smartphone a second time after receiving this feedback. Results revealed that polite smartphones were generally evaluated significantly more positive than impolite smartphones. Additionally, impolite smartphones were significantly devaluated following their feedback compared to the first evaluation, both regarding their friendliness as well as their competence. This evaluation was also influenced by the gender of the voice the smartphone employed. Impolite male smartphones were rated to be less competent when compared to impolite female smartphones (Carolus, Muench, et al., 2019).

3.3.4.3 Voice in other Human-Technology Interactions

Multiple studies have been conducted transferring the adoption of gender stereotypes to interactions with virtual characters (e.g., Zanbaka et al., 2006) and robots (e.g., Siegel et al., 2009) using male and female voices. However, as voice is not the only gender-relevant cue when interacting with these embodied technologies, the relevancy for this thesis is limited and for the sake of brevity, these studies will not be expanded upon here. The results of these studies are generally in support of media equation effects.

3.3.4.4 Summary

Voice activates a categorization that is profoundly associated with living things: sex. Speech (and consequently the implied sex) as social cues have been shown to be well established in classic media equation research and have demonstrable effects on human-computer as well as human-smartphone interaction. Other cues that can be transported via voice such as personality have also produced positive results in CASA research. However, neither PCs nor smartphones are inherently designed to use speech as their main channel of interaction. The voice-based personal assistant devices that are the focal point of this thesis employ vastly improved state-of-the-art technology when compared to many of the devices examined in previous CASA research employing TTS-voices (McTear et al., 2016). Smart speakers follow a different design philosophy that is clearly designed around voice input and output while at the same time completely avoiding any visual anthropomorphic features (Guzman, 2015; Luger & Sellen, 2016) thus making them a fitting device for media equation research.

3.4 Voice Assistants and Smart Speakers

As established in section 3.2, social norms, group effects, and gender stereotypes are adopted in HCI just like in human-human interaction, and that media equation effects can be triggered by voice as a social cue (see section 3.3.4). Since the goal of this thesis is to transfer these effects to human-voice assistant and more specifically smart speaker interaction, the terms voice assistant and smart speaker as well as their technological capacities in comparison to other technological devices need to be clearly defined. This is followed by an overview of previous empirical research involving these technologies.

3.4.1 Definition

Voice-enabled technologies and how humans interact with them are the subject of many fields of research including psychology, HCI, HRI, computer science, information systems and communication sciences among others (Guzman, 2015). As a result, many terms are used interchangeably to describe these technologies with no clear consensus yet. Terms used to describe the emerging technology of voice assistants include virtual butler (Payr, 2013), voice-controlled intelligent personal assistant (Kiseleva et al., 2016), conversational interface (McTear et al., 2016), conversational agent (Schuetzler et al., 2018), intelligent personal assistant (Lopatovska et al., 2019) and virtual personal assistant (Jang, 2020) to just name a few. For the purposes of this thesis, the term voice assistant will be used to describe "user interfaces that mimic human-to-human communication using natural language processing, machine learning, and/or artificial intelligence" (Schuetzler et al., 2018, p. 94). There are two important distinctions to be made regarding voice assistants. First, the distinction between *personal voice assistants* designed for individual and home use and *commercial voice assistants* designed for business contexts such as e-commerce or customer service (Gnewuch et al., 2017). For the purposes of this thesis, all mentions of voice assistants refer to personal voice assistants. Secondly, voice assistants are integrated into a variety of technological devices, most prominently smartphones (Luger & Sellen, 2016) thus making a distinction between the hardware component and software component of voice assistants an important factor. This thesis focuses on a special subset of voice assistants, those integrated into stand-alone hardware devices, so-called *smart* *speakers*. Smart speakers are defined as “a wireless device with artificial intelligence that can be activated through voice command” (Smith, 2020, p. 1). Interactions with smart speakers consist primarily of spoken input and natural language output (Porcheron et al., 2018) and they are mostly used to fulfill simple tasks such as playing music, searching for information, placing shopping orders, managing appointments or controlling other smart devices (Canbek & Mutlu, 2016). In contrast to virtual or embodied agents which offer either metaphorical or human-like embodied representations, smart speakers are usually inconspicuously shaped loudspeakers that come in different forms and sizes based on the device. While they usually contain some visual indicators such as LED lights that change based on the systems state, they are neither strictly speaking virtually or physically embodied nor designed to be anthropomorphized based on any physical features like virtual agents, avatars or robots and instead focus on behavioral realism as a principal goal (Luger & Sellen, 2016). This is an important distinction for this thesis as it aims to exclusively examine social reactions towards these disembodied personal, stand-alone voice assistant devices such as the Amazon Echo or the Google Home.

3.4.2 Emergence and Adoption of Personal Voice Assistants and Smart Speakers

In 2010 Apple Inc. released its virtual personal assistant Siri which Apple claims is now being used by over 500 million users (Wardini, 2022). This was followed by Microsoft launching their voice-driven assistant Cortana in 2013. In 2014, Amazon was the first to release a stand-alone smart speaker device, the Amazon Echo. Along with the hardware, a voice-controlled personal assistant called Alexa was released (Hoy, 2018). The novelty of the device came from the fact that it was a stand-alone device without a screen forgoing a traditional graphical interface. The most cited advantages of these personal voice assistants are threefold:
(1) Users can interact with them naturally and intuitively (Chattaraman et al., 2019) which is the most distinctive characteristic of voice assistants (Araujo, 2018) and enables a human-like interactional experience (Schuetzler et al., 2014).
(2) They are useful in case of hands- or eyes-free situations where users are engaged in other activities and need to concentrate on several things at once like when

playing with their children, cooking, or driving a car (Cowan et al., 2017; Luger & Sellen, 2016; Moore, 2013; Murad et al., 2018).

(3) Voice assistants provide accessibility, resources and services to disabled, visually impaired, or elderly users not capable of operating a graphical interface (Baber & Noyes, 2002; P. R. Cohen & Oviatt, 1995; Pradhan et al., 2018; Wolters et al., 2016).

Other fields in which voice assistants can provide benefits are healthcare (e.g., Hoy, 2018), education (e.g., Terzopoulos & Satratzemi, 2019), and fitness (e.g., Chung et al., 2018). Across all these fields of application, researchers suggests that voice assistants need to adopt the characteristics of human communication to appear more natural and more engaging (see Dybkjaer et al., 2004). Research has identified various factors that led to the adoption of smart speakers. Among these are product-related factors, platform-related factors and privacy concerns (K. Park et al., 2018). Value related factors such as perceived enjoyment and perceived usefulness have also been found to influence usage intention (Kowalczuk, 2018). Usage was also linked to content quality, automation and visual attractiveness (Yang & Lee, 2019).

3.4.3 Technological Aspects of Smart Speakers

This section aims to give an overview of the most important technological aspects of smart speakers that have been used in the studies conducted as part of this thesis. Speech analysis and speech synthesis has been researched since the 1930s (for an overview, see Juang & Rabiner, 2005), but Hirschberg and Manning (2015) list four major technological advancements since the development of ELIZA in 1966 (see section 3.1.1) that serve as the foundation of voice-enabled technologies. (1) A vast increase in computing power, (2) the availability of very large amounts of linguistic data, (3) the development of highly successful ML methods, and (4) a much richer understanding of the structure of human language and its deployment in social contexts. These advancements are also cited as the reason voice assistants are considered ready for everyday use and suitable for distribution to the public (Hoy, 2018). When a voice-based system receives spoken input, the system must complete a series of tasks to provide users with a matching output (McTear et al., 2016):

(1) The system must recognize the words that were spoken by the user (speech recognition).

(2) The system must interpret the words that were spoken to infer the intent behind them (spoken language understanding).
(3) The system must either produce a response or if no intent was understood, interact with the user to seek clarification (dialog management).
(4) Generate the desired response (response generation).
(5) Use voice output to communicate the response (text-to-speech synthesis).

To complete these tasks, smart speakers such as the Amazon Echo rely on multiple technologies interacting with each other. The most important ones are automatic speech recognition (ASR), natural language processing (NLP), and natural language generation (NLG). The next sections briefly introduce these technologies and their functions as they were employed to facilitate the interactions between smart speakers and participants in all four experiments conducted as part of this thesis.

3.4.3.1 Smart Speaker Functionality

When a user talks to a smart speaker, the user’s speech is processed by an automatic speech recognizer component and transcribed into a string of text. To do so, the system decodes the speech into distinct sounds, with the smallest unit being phonemes. These phonemes are then compared with a dictionary storing information about words that can be spoken and the corresponding phoneme sequence (Gruhn et al., 2011). As a result of this search, the word that fits best to the observed input is then assessed by the likelihood of its occurrence in the given context (Rabiner & Juang, 2006). Information on which sequences of words are more likely to occur in spoken language is stored in language models similar to the prototypes humans have stored in their long-term memory (Gruhn et al., 2011). The ASR component can also produce multiple hypotheses of the words that were spoken ordered based on their probability to minimize recognition errors. This is important, as spoken language is highly variable and direct matches are nearly impossible as a result (McTear et al., 2016). The string of text produced by the ASR is then converted into a sequence of words that represents the meaning of the utterance and its dependencies by the natural language understanding (NLU) component, a sub-component of the NLP module (see Figure 1 for an example).

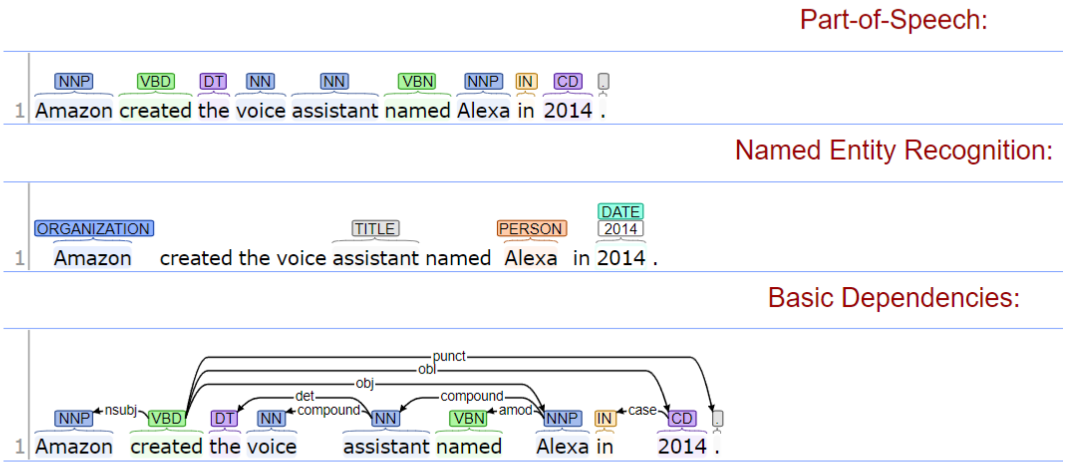


Figure 1. Example of natural language processing using the Stanford CoreNLP (source: https://corenlp.run/)

The function of the NLU is to identify the intent behind the user's utterance and represent it in machine-understandable form. This representation can then be processed further by the systems dialog manager (DM) to generate a matching response to the user (Pieraccini, 2021). The DM decides the systems next action based on the identified intent as well as contextual information such as previous interactions or the environment. If the DM decides that the system is supposed to give an answer, it sends a request to the NLG module. The NLG module then produces a textual representation of the response the system needs to communicate to its user. During the last step, the TTS-module generates the matching utterance and outputs it through the device's loudspeaker for the user to hear (see Figure 2 for an overview of this process).

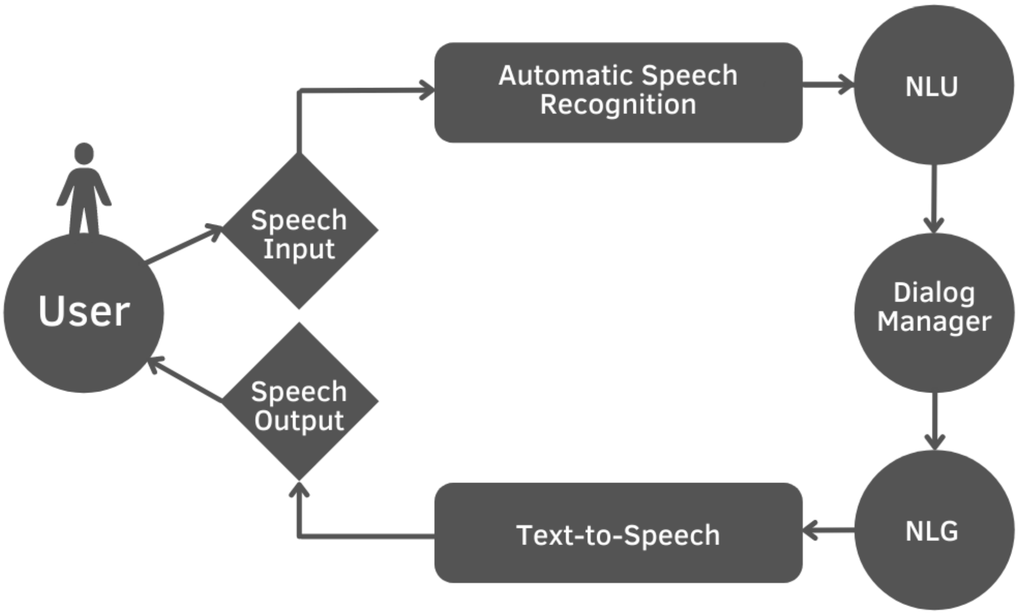


Figure 2. Core components of smart speaker functionality (source: own figure)

3.4.3.2 Artificial Intelligence

One more aspect that sets voice assistants apart from technologies early CASA research was focused on is that they employ AI and ML to continuously improve the process of NLP and thus seem like they are 'learning'. AI can be described as a technology that uses algorithms and similar techniques to allow computers to perform rational tasks and simulate the intelligence and mental capacities of humans. The foundation of these processes is data, which is used by ML to learn patterns and structures (Attaran & Deb, 2018). The rapid developments in the field of AI over the past decade have had severe consequences for technology and human-technology interaction. Research has started to focus on the mental abilities of technology such as emotional experience (K. Gray & Wegner, 2012) and emulated empathy (Liu & Sundar, 2018). It has also led users to ascribe certain attributes to machines and technological devices such as goal direction, interactivity, personality or even a digital mind (Stein & Ohler, 2017). Conversational systems can also facilitate this process by referring to themselves as though they were a distinct entity (Nass & Brave, 2005). This can result in users having only a vague concept of new technological devices such as smart speakers and basing their attitude towards these devices on imagination and assumption instead of previous experience (Ashfaq et al., 2021). As mentioned in section 3.1.3, the ascription of mental capacities to these technologies can also foster the effects of anthropomorphism (X. Li & Sung, 2021; D. Park & Namkung, 2021; Shank et al., 2019) and perceptions of these devices as social actors (Nass & Brave, 2005).

3.4.3.3 Voice Recognition

Being understood is one of the main goals of any verbal human-human interaction (Nass & Gong, 2000). One major issue for any conversational interface or system is that it must deal with spoken language, which is far less regular than written text. Spoken utterances are often ill-formed and contain characteristics of spontaneous speech, such as hesitation markers, self-corrections, and other types of disfluency (McTear et al., 2016). In human-human interaction, people have adopted tactics to prevent communicative acts from failure. A sender reduces ambiguity by restarting a word, changing the utterance (Bosco et al., 2006) or use 'hyperarticulate' speech like prolonged pauses and emphasized words, a behavior that has also been demonstrated in HCI (Oviatt et al., 1998). A receiver can minimize misunderstands by requesting clarification if what was said remains unclear (Weigand, 1999). However, despite the significant improvements of voice-based technologies over the last years, voice recognition errors are still a major challenge for conversational interfaces and an important topic in research (e.g., King et al., 2017; J. Li et al., 2014; Mošner et al., 2019). The high ambiguity of human language makes it difficult to replicate these mechanisms even with modern technologies (Chowdhary, 2020). Additionally, since automated speech recognition is based on pattern matching, external factors such as loud surroundings, interruption of the user or unintended recognized intents can further complicate intelligibility and processability of voice inputs (Oviatt et al., 1998).

3.4.3.4 Smart Speaker Devices

This thesis opted to use an Amazon Echo device in three out of four experiments for multiple reasons. Firstly, the Amazon Echo does not provide an anthropomorphic physical embodiment. The Amazon Echo Plus Model (1. Generation) that was used for most studies conducted as part of this thesis has the form of a cylinder about 9 inches tall that contains a microphone array and speakers.

Secondly, Amazon is the dominant player in the field of voice assistants (Hoy, 2018) and continues to maintain a 70% market share (Sterling, 2019) and thus the devices were readily available. In addition, Echo devices are highly customizable which is important for research purposes as it allows for effective manipulation (more on customization in section 4). All Amazon Echo devices are equipped with the AI assistant Alexa, a software connected to the *Amazon Cloud* that processes speech input and creates appropriate responses (*Alexa and Alexa Device FAQs*, n.d.). Every Echo device consists of multiple microphones, a speaker, and a software module, that connects to the internet and the Amazon Cloud (Bedford-Strohm, 2017). All models use far-field speech recognition and beam-forming technology that allows them to hear from any direction and even filter out ambient noise like background music (McTear et al., 2016). The ability to respond to the user's request is managed by *Skills* which are programs that define how the speech input is to be processed and which response to a request is to be generated (*What Is the Alexa Skills Kit?*, n.d.). To trigger a skill and start interacting with the conversational agent, a wake word – 'Alexa' by default – is used. Some Skills are offered as a standard feature by Amazon on any Echo device, others can be accessed via the internet and have been developed by third-party developers using platforms made available by the manufacturer, such as the *Alexa Developer Console*. The basic functionality of all Echo smart speaker models is identical: users formulate a request verbally which is then processed using server-sided ASR and NLP techniques. The device then obtains data on how to respond to the request from online services and uses it to formulate a reply. In addition, smart speakers also allow for control over Internet of Things (IoT) devices such as smart lights to be controlled using voice input (Weaver et al., 2020). Study 3 (see section 6) instead opted to use a Google Home smart speaker device which offers the same basic functionality. The reasons for this and the differences between the devices are expanded upon in section 6.

3.4.4 Previous Research

Research on smart speakers has focused on a variety of topics. Most previous studies focused on technological aspects (e.g., Kowalczuk, 2018; Yang & Lee, 2019), marketing (e.g., Ling et al., 2021; Smith, 2020), advertising (e.g., D. Kim et al., 2018; H. Lee & Cho, 2020) or privacy and security (e.g., Lau et al., 2018; Y.

Liao et al., 2019; Pfeifle, 2018). However, despite their huge impact on everyday life and growing prevalence worldwide, little empirical research has been done regarding social reactions to smart speakers. A recent meta-analysis on empirical research regarding voice-based human–agent interaction revealed that “there is little diversity in agent type […] few projects included smart ‘things’, such as smart speakers, smart TVs, and smart vehicles. Of these, three were vehicles and two were speakers, pointing to a limited focus on the kinds of ‘things’ possible. None involved the most common options on the market today: Siri, Alexa, and Google Assistant.” (Seaborn et al., 2021, p.12). These results affirm that there is a clear research desideratum when it comes to these stand-alone devices given their growing prevalence outlined in section 3.4.2. While there is a definitive lack of CASA research focused on social reactions towards smart speakers based on voice-based human-agent interaction, there is still at least some empirical research that provided first insights into how people evaluate and interact with these technologies. For the sake of completeness, the next section aims to provide a brief overview of the most important results.

3.4.4.1 User evaluations of smart speakers

User satisfaction has been the focal point of multiple studies regarding voice assistants (e.g., Hashemi et al., 2018; Jang, 2020; Kiseleva et al., 2016; Purington et al., 2017). In a study conducted by Luger and Sellen (2016), participants were interviewed about their usage and perception of personal voice assistants, specifically Siri, Google Now, and Cortana. Results revealed a “gulf” between users’ expectations and the actual features and capabilities of the devices. In several of the interviews conducted by Luger and Sellen (2016), participants stated that they initially tried to interact with their personal voice assistants as they would with a human counterpart. Upon realizing that this style of interaction did not lead to successful interactions because of recognition errors and misunderstandings, they revised their interaction patterns to accommodate the limited capabilities of the voice assistants. Still, some participants ascribed attributes to the voice assistants that are usually linked to human personality such as ‘sassy’ or ‘sarcastic’ (Luger & Sellen, 2016). A content analysis of customer reviews written about Amazon Echo devices was conducted by Purington et al. (2017). The focal points of the analysis were the personification of the device by customers (using the name Alexa instead am Echo as well as subject pronouns when referring to the device), sociability level of interactions with the device, and their overall satisfaction. Results revealed that greater personification coincides with an increased number of social interactions. Personification was also positively linked to satisfaction with the device (Purington et al., 2017). In a first field study aimed at exploring how voice assistants are embedded into everyday interactions, Porcheron et al. (2018) installed Amazon Echo devices in participants' homes and gathered usage data for a month. Data analysis was focused on how the assistant was integrated into various conversational situations by its users. Results revealed that participants rarely had conversations directly with the Amazon Echo device and that interactions with the device happened during conversations with other people such as family members. The authors conclude that the voice assistant was embedded in the life of the home (Porcheron et al., 2018). Lopatovska and Williams (2018) explored how personification of an Amazon Echo manifests during user interaction. They collected data by asking participants to write a diary about their interactions with an Amazon Echo device. Data analysis revealed that most personification behavior could be classified as "mindless politeness" resulting in users saying 'thanks' or 'please' during interactions with the device. Still, only 19 participants were part of the study, and less than half displayed this behavior (Lopatovska & Williams, 2018). In a second study, Lopatovska et al. (2019) also explored how users interact with the device using a diary method. The most common tasks were checking the weather, playing music, and controlling other devices. Most participants reported satisfactory interactions and outcomes and even when the device failed to provide the desired information or complete the task, satisfaction was still high. Thus, the authors conclude that the experience of the interaction is more important than the output (Lopatovska et al., 2019).

3.4.4.2 Social attributions towards smart speakers

As mentioned before, there is a shortage of studies regarding smart speakers that can be classified under the CASA paradigm. As of now, only two studies explored social evaluations of voice assistants based on gender stereotypes and only one did so using a smart speaker in an experimental setting. In a study conducted by Ernst and Herm-Stapelberg (2020), the influence of gender stereotypes on perceived competence of voice assistants such as Siri was examined. Participants were instructed to ask a series of questions that a voice-assistant device answered with either a male or a female voice. Participants were then asked to evaluate the device's competence with results indicating that participants perceive male voice assistants to be more competent (Ernst & Herm-Stapelberg, 2020). It should be noted that there are two methodological deficits in this study: no behavior measures were employed to measure social reactions, which is contradictory to the unconscious component of gender stereotypes and media equation effects. In addition, only a total of 23 students took part in the study resulting in extremely small experimental groups (e.g., only 9 participants rated a male smart speaker and only 2 of these participants were male). Tolmeijer et al. (2021) explored the effect of voice assistant gender and pitch on trust attribution. Participants were asked to interact with an online voice-assistant interface that used one of five different synthesized voices named female high, female low, gender-ambiguous, male high, and male low based on their pitch in one of two task types: assistance (booking a flight) and compliance (customer survey). While the study found some evidence indicating an influence of voice gender and pitch on stereotypical on trait attribution, again, no behavior measures were employed. Both trait ascription and trust were evaluated using only self-reports. Additionally, due to the study being conducted online because of quarantine protocols and participants merely being asked to imagine that the interaction was happening in real-life, the external validity of these results is heavily limited (Tolmeijer et al., 2021).

To summarize, previous research was mostly based on the general appraisal of voice assistants and smart speakers as well as on personification of the device and how it was integrated into conversations. Barely any empirical research was focused on psychological effects and social reactions towards the device revealing a clear research desideratum. While some initial empirical studies examine gender stereotypes from a CASA perspective, due to the methodological problems mentioned, results remain inconclusive at best (Ernst & Herm-Stapelberg, 2020; Tolmeijer et al., 2021). This becomes even more apparent with the clear lack of objective behavioral measures in modern HCI and modern CASA research, something that has been noted before (P. A. Williams et al., 2017) and has also been a point of criticism in a recent meta-analysis of speech-related research in HCI (for an overview, see Clark et al., 2019). Self-report measures are predisposed to multiple limitations such as social desirability and self-deception which limits their reliability and validity (e.g., Austin et al., 1998; Van de Mortel, 2008). In addition, self-report measures are a conscious process as subjects are instructed to think about their evaluations. This is an especially relevant factor in CASA research: while many media equation effects persist even in subjective self-reports (see section 3.2) depending on the explanation used, they are mostly theorized to happen unconsciously and automatically (see section 3.1.3). To exhaustively examine social reactions to smart speakers this thesis opted to include objective behavioral measures for all experiments to capture the unconscious components of social reactions in addition to subjective evaluations of the devices.

3.5 Summary of the Theoretical Background

A significant part of HCI research is based on the CASA paradigm and consistently demonstrated that many forms of human-technology interaction follow the same rules and expectations that also define and form human-human interaction (see section 3.1). Humans seemingly use social rules, categories, and scripts when interacting with technological devices that exhibit social cues even though they understand these devices do not justify such behavior. Among relevant criteria from human-human interaction are politeness (see section 3.2.2), reciprocity (see section 3.2.3), interdependence (see section 3.2.4) and gender stereotypes (see section 3.2.5). One of the most significant cues triggering social reactions is speech (see section 3.3). Disembodied smart speakers are among the most capable devices to not only manifest vocal cues from a technological perspective but to allow for completely voice-based interaction while at the same time foregoing any additional markers of humanness (see section 3.4) and yet as of this moment, very little empirical research has been done to extend media equation and CASA research to these devices (Seaborn et al., 2021). If desktop computers and smartphones are treated as social actors, this effect should be even stronger when it comes to voice assistants based both on their technological capabilities (Guzman & Lewis, 2020) and their adoption of speech as a form of communication that is an inherently human behavior and thus presents a strong social cue and marker of humanness (Pinker, 1995). Now that the underlying social norms and effects originating in human-human interaction have been clearly defined and established, this dissertation aims to transfer lessons learned from previous CASA and media equation research to personal stand-alone voice assistant devices by focusing on psychological processes in human-voice assistant interaction. As previously stated, depending on the explanation used, media equation theorizes social reactions towards technological devices to happen unconsciously and automatically (see section 3.1.3). However, many of the studies described in the previous sections rely solely on subjective self-reports, which instruct participants to think about their assessments. In addition to traditional questionnaires assessing subjective self-reports, every experiment conducted for this thesis contains a complementary measure to assess the behavioral component of media equation. Behavioral methods provide a more objective approach and are particularly suitable for measuring (unconscious) social reactions without necessarily making their measurement obvious to participants.

3.6 Research Questions

Based on previous deliberations, the objectives of this thesis are threefold: (1) gather first empirical evidence that people show social reactions towards personal stand-alone voice assistant devices – or in short: smart speakers. (2) Include both self-reports and additional objective measures to assess social reactions and (3) assess previously explored as well as unexplored individual differences that are theorized to potentially influence media equation and social reactions towards technological devices. Based on these deliberations there is one major research question encompassing all four experiments conducted as part of this thesis:

$RQ_{Overall}$: Are smart speakers treated as social actors and evoke social reactions on a subjective level as well as on a behavioral level? Do individual differences influence these effects?

This overarching research question can be divided into four sub-questions that were explored in four separate experimental laboratory studies.

RQ_1: Do people follow the social norm of politeness, specifically the interviewer-bias, in human-smart speaker interaction?

RQ_2: Do people show prosocial behavior towards smart speakers based on minimal cues, specifically team affiliation?

RQ_3: Do people adopt gender stereotypes in human- smart speaker interaction, and do they show corresponding conformal behavior?

RQ_4: Do people show reciprocal behavior towards smart speakers based on previous helpful or unhelpful interactions?

4 Experiment 1: Effects of (Im)Politeness and Interviewer-Bias in Human-Smart Speaker Interaction

4.1 Study Outline and Hypotheses

As described in section 3.2.1 the social norm of politeness is one of the most important norms in human-human interaction and is considered to be universal across all cultures (P. Brown & Levinson, 1978). One way to assess politeness via a behavioral measure is the interviewer-bias (Finkel et al., 1991) which has been shown to be adopted in both human-computer (see section 3.2.2.3) as well as in human-smartphone interaction (see section 3.2.8.1). Consequently, this study is based on the CASA experimental design established by Nass et al. (1999) for human-computer interaction and revisited by Carolus et al. (2019) for human-smartphone interaction with two modifications: (a) an Amazon Echo smart speaker device was used instead of a desktop PC or a smartphone thus allowing for an entirely verbal interaction using only the social cue of voice. (b) Two different evaluation modalities were incorporated as a within-factor: Participants evaluated the smart speaker verbally on the device itself and a second time at a separate desktop PC using an online questionnaire. Using this operationalization, study 1 sought to answer three questions: (1) do people generally show social reactions towards a voice assistant displaying polite or impolite behavior? (2) Can this effect be assessed using an additional behavioral measure, in this case, the interviewer-bias? (3) Do individual differences influence these effects? Since stand-alone voice assistant devices were a comparatively new technology when this experiment was conducted, prior experience with voice assistants was included as an individual factor for this experiment.

Politeness is a social norm with far reaching behavioral consequences for human-human interaction (Brown & Levinson, 1978). As the technology behind voice assistants becomes more sophisticated, interactions with users will become more diverse and complex. This inevitably leads to situations in which the assistants must communicate failed states and interactions to users, giving critical feedback or even criticize users directly. However, confronting users with their failure could violate politeness norms which usually results in social sanctions in human-human interaction (Goffman, 1967). Consequently, it can be assumed that the same politeness strategies employed in human-human interaction could be used to lessen the impact. One such strategy is for the voice assistant to issue an apologetic statement and blame itself for a failed interaction. Previous studies showed that users receiving apologetic feedback from a computer made them feel better about their interaction with a program, which is in line with the effects of apologies in human-human interactions (Akgun et al., 2005; Tzeng, 2004) and similar effects were also observed for mitigation of breakdowns in HRI (M. K. Lee et al., 2010). These results indicate that using politeness strategies can mitigate face-threatening acts in human-technology interaction. The face-threatening act mitigated in this experiment is informing participants that their data is inadequate for further analysis. An impolite voice assistant on the other hand is expected to be evaluated as less friendly after directly blaming the user for a failed interaction. The face-threatening act of impoliteness should result in a more general punishment not distinguishing between characteristics that are task-related and those that are not (Goffman, 1967). In addition, criticism is expected to be more consequential in a voice interaction when compared to other forms of human-technology interaction due to the importance of speech outlined in section 3.3 (also see Nass & Gong, 2000). Accordingly, it is expected that impolite voice assistants are generally devaluated even on dimensions that are identical between groups. This leads to the assumption that, compared to polite feedback, the impolite feedback given by a voice assistant should influence the ratings of overall valence (H_1), friendliness (H_2), competence (H_3), performance (H_4) and general attitude (H_5) towards the device. In contrast, a voice assistant that issues an apologetic statement and blames itself for the failed interaction is not expected to differ in the evaluation from the other groups in any significant way as the face-threatening act is mitigated.

> H_{1a}: Smart speaker devices that give impolite feedback are evaluated significantly worse regarding their overall valence compared to devices that give polite feedback.
>
> H_{1b}: Smart speaker devices offering apologetic feedback for a failed interaction are not evaluated significantly worse regarding their overall valence compared to polite devices.
>
> H_{2a}: Smart speaker devices that give impolite feedback are evaluated significantly worse regarding their friendliness compared to devices that give polite feedback.

H_{2b}: Smart speaker devices offering apologetic feedback for a failed interaction are not evaluated significantly worse regarding their friendliness compared to polite devices.

H_{3a}: Smart speaker devices that give impolite feedback are evaluated significantly worse regarding their competence compared to devices that give polite feedback.

H_{3b}: Smart speaker devices offering apologetic feedback for a failed interaction are not evaluated significantly worse regarding their competence compared to polite devices.

H_{4a}: Smart speaker devices that give impolite feedback are evaluated significantly worse regarding their performance compared to devices that give polite feedback.

H_{4b}: Smart speaker devices offering apologetic feedback for a failed interaction are not evaluated significantly worse regarding their performance compared to polite devices.

H_{5a}: Smart speaker devices that give impolite feedback are evaluated significantly worse regarding participants' general attitude towards them compared to devices that give polite feedback.

H_{5b}: Smart speaker devices offering apologetic feedback for a failed interaction are not evaluated significantly worse regarding participants' general attitude towards them compared to polite devices.

Error reported by a system usually leads to the perception of the system being in an incorrect state that prevents users from achieving their goals thus leading to frustration (Ceaparu et al., 2004). Additionally, not achieving an intended goal has been found to lower perceived self-performance (Baumeister & Tice, 1985; Stotland & Zander, 1958). As impolite behavior indicates a failed system state, it is assumed that evaluations of self-performance suffer as well. Apologetic feedback also indicates a failed system state. However, as a politeness strategy is employed to mitigate this state, better evaluations of self-performance are expected.

H_{6a}: Smart speaker devices giving impolite feedback result in a significantly lower evaluation of participants' self-performance during the interaction compared to devices that give polite feedback.

H_{6b}: Smart speaker devices offering apologetic feedback for a failed interaction result in a significantly higher evaluation of participants' self-performance during the interaction compared to impolite devices.

Based on the politeness norm, more specifically the adoption of an interviewer-bias (Finkel et al., 1991), it is postulated that evaluations on the device itself will be more positive compared to evaluations given at a separate computer (Nass et al., 1999).

H_7: Participants will evaluate a smart speaker device more positively in terms of overall valence if the device asks for the evaluation itself compared to a separate desktop PC asking.

H_{7a}: Participants will evaluate a smart speaker device to be friendlier if the device itself asks for the evaluation compared to a separate desktop PC asking.

H_{7b}: Participants will evaluate a smart speaker device to be more competent if the device itself asks for the evaluation compared to a separate desktop PC asking.

Exploratory considerations. Based on previous results regarding the influence of experience with computers on their evaluation (Johnson et al., 2004; Johnson & Gardner, 2007), it was considered that previous experience with voice assistants could influence media equation effects. No hypotheses were formulated in advance but based on previous results it is assumed that prior experience can moderate the effects of media equation.

RQ: Does previous experience with voice assistant devices influence social behavior towards the device or the assessment of the device?

4.2 Methods

4.2.1 Participants

A total of 131 university students participated in the experimental laboratory study. They were mostly recruited through the university platform for the recruitment of participants. After excluding all participants that encountered unintended technical errors (n = 31) during their interaction with the smart speaker device, the final sample for this study includes N = 100 subjects. Participants' age ranged from 18 to 34 years (*M* = 21.3; *SD* = 2.34). 67% of participants were female and 33% were male. As participants were mostly recruited using the internal recruitment system of the Institute Human-Computer-Media in exchange for course credit, most of the sample reported a higher education. 91% were qualified for university entrance and 5% reported holding a university degree. 4% of participants reported a secondary school leaving certificate. Out of 100 participants, 50 of them (50%) had never used a voice assistant prior to this experiment. 16 (16%) reported having used voice assistants for up to 1 year, 14 (14%) reported having used them for 2 years, 9 (9%) reported 3 years of usage and 11 participants (11%) reported that they had used voice assistants for more than 3 years prior to this experiment. Written informed consent was obtained from each participant before the study.

4.2.2 Stimulus Material

4.2.2.1 Hardware and Software

Participants interacted with an *Amazon Echo Plus* (1. Generation) hardware device (see Figure 3). The software used was a custom-built *Skill*[3]. As this study was conducted early in the lifespan of Amazon Echo products, many of the technical capabilities needed to build more sophisticated Skills were not yet available, which is why the (now defunct) chatbot platform *Dexter* (https://rundexter.com/) in combination with *Amazon Web Services* (https://aws.amazon.com/) was used instead to simulate a human-voice assistant interaction. A chatbot containing all text modules such as questions, recaps, and the feedback given to participants was created using Dexter. This chatbot was then integrated with the Amazon Echo device to allow for voice input and voice output of these text modules. The voice used for the interaction was the default German Amazon Alexa voice. For the sake of consistency, all participants were told that they would interact with an Amazon Echo device running the Amazon Alexa software. Neither the name ‘Alexa’ nor another wake word was used during the interaction after it was initiated by the researcher. This was done to avoid giving participants an additional locus of attention.

[3] Skills generally refer to a set of predetermined actions or tasks that are accomplished by an Amazon Echo device and are comparable to individual Apps on a Smartphone. Skills are highly customizable regarding voice input and output and can be engaged with naturally through voice.

4.2.2.2 Voice Interaction

The interaction between the smart speaker and participants consisted of the device asking a series of questions after each of which the participants were instructed to answer. Due to technological restrictions of smart speakers, this was always done in an alternating fashion. After every question, the device signaled that it was ready to listen via a blue LED ring glowing at the top of the device (see Figure 4). Registration of the answer given by participants was indicated by a short neutral confirmation (e.g., “okay”, “thanks”, “understood”) followed by the next question. To avoid voice recognition errors during the interaction, every answer given by participants was accepted, even if it was unrelated to the question asked. If no answer was given within ten seconds, the device was programmed to repeat the question. For a protocol of the entire voice interaction, see Appendix A. The interaction took place in a laboratory setting that was intentionally designed to somewhat resemble a living room, as this is more closely resembles a setting in which personal voice-assistant devices are used (see Figures 3 and 4). The laboratory setting also allowed to control confounding variables such as background noises that can result in speech recognition errors in voice-based systems (Wang et al., 2019).

Figure 3. Experimental setting (source: own figure)

Figure 4. The experimental setting from participants' perspective (source: own figure)

4.2.3 Self-Report Measures

4.2.3.1 Valence towards the device

General valence towards the smart speaker device was assessed using an index composed of twelve adjective items adapted from the ‘valence towards the computer’ scale introduced by Nass et al. (1999). Items were translated to German and all references to a computer in the original items were replaced by references to the smart speaker device. The index is further distinguished by the two dimensions ‘competence’ (competent, informative, helpful, analytical, knowledgeable, useful) and ‘friendliness’ (likeable, friendly, warm, enjoyable, fun, polite). All 12 items were answered on a 7-point Likert-scale ranging from 1 (*does not apply at all)* to 7 (*is absolutely true*). The resulting index of all 12 items exhibited high reliability, Cronbach’s Alpha = .88. Reliability was also high for the subscale ‘friendliness’ (α = .82) and ‘competence’ (α = .89) for evaluations via questionnaire on a desktop PC. All original and translated items can be found in Appendix B.

4.2.3.2 Performance of the device

The performance of the voice assistant was assessed using a scale developed by Johnson et al. (2004) and adapted to voice assistants. The scale was originally developed specifically for CASA research to consistently evaluate desktop PCs, as previous CASA research often used inconsistent measures which reduced comparability (Clark et al., 2019; Johnson et al., 2004). The scale consists of 6 items (e.g., “How well did the voice assistant perform?”) which were asked on a 7-point Likert scale (1 = *not at all*; 7= *very*). The reliability of this scale was excellent with α = .94.

4.2.3.3 Attitude towards the device

Besides measuring the evaluation of the device’s performance, participants were also asked to rate their attitude towards the voice assistant on a 9-item semantic differential scale developed by Johnson et al. (2004). In order not to imply human characteristics in the question, the authors developed a semantic differential to quantify the evaluation of the device thus eliminating possible influences of normative behavior from participants. All 9 items were answered on a 7-point

Likert semantic differential scale (e.g., “Unhelpful – Helpful”, “Unintelligent – Intelligent”). The reliability of this scale was acceptable with α = .76.

4.2.3.4 Self-performance during the interaction

Participants were also asked to rate their own performance during the voice interaction task. To quantify the self-performance evaluation a scale introduced by Johnson et al. (2004) was used. Responses to the five items (e.g., “How well do you feel you performed?”) were made on a 7-point Likert scale (1 = *not at all*; 7 = *very*). This subscale exhibited high reliability, α = .87.

4.2.3.5 Manipulation check

To control if participants correctly recognized the experimental manipulation, two items were constructed and implemented at the of the questionnaire. Participants were asked if the voice assistant had to exclude the data they provided from further analysis. If they answered with ‘yes’, they were also asked what reason the voice assistant stated for the exclusion in an open-ended question. All 66 participants that received either apologetic or impolite feedback for an unsuccessful interaction correctly answered the first question with ‘yes’. The answers given to the second, open-ended question also indicate that the manipulation was generally successful and provided no reasons to exclude specific participants from further analyses.

Table 2. Overview of scales and their internal consistencies for the variables in Experiment 1 (N = 100)

Scale	Subscale	Cronbach’s α	Items
Valence towards device (on device)		.85	12
	Competence	.74	6
	Friendliness	.77	6
Valence towards device (separate PC)		.88	12
	Competence	.89	6
	Friendliness	.82	6
Performance of the device		.94	6
Attitude towards the device		.76	9
Self-performance		.87	5

4.2.4 Behavioral Measures: Interviewer-Bias

In line with the within-factor (evaluation location), the ‘valence towards the computer’ scale was asked twice: first by the device itself (oral interview) and later repeated at the desktop computer (online questionnaire). Reliability was also high for the oral version of overall valence (Cronbach’s Alpha = .85) and acceptable for the subscales friendliness (α = .77) and competence (α = .74). A comparison between both evaluations is used to assess the occurrence of an interviewer bias: a device asking for the evaluation itself is expected to produce better ratings due to participants adopting politeness strategies during direct oral questioning (P. Brown & Levinson, 1978; Finkel et al., 1991; Nass et al., 1999).

4.2.5 Procedure

A 2x3 experimental mixed factorial design was used in which participants were randomly assigned to one of the three politeness conditions (between-factor: type of feedback). Each participant evaluated the conversational agent first on the device itself and afterwards on a separate desktop PC (within-factor: evaluation location). All participants interacted with an Amazon Echo Plus device in four steps: (1) warm-up, (2) main interaction including manipulation, (3) evaluation on the device itself, and lastly the (4) evaluation on a separate computer.

(1) The experiment began with a warm-up phase, during which participants were given instructions on how to operate an Amazon Echo device, as most participants did not have any prior experience with smart speaker devices at the time this study was conducted. Then they were given five minutes to test the device by interacting with it freely. After the warm-up phase, the main interaction was initiated, and the researcher left the laboratory so only the participant and the device remained in the room. (2) Following the warm-up, the smart speaker device presented participants with the cover story: they were told that the goal of the study was to help the device to collect data required for a new Skill. This Skill was aimed at creating a database of helpful information to provide orientation for newly enrolled university students. The device informed participants that to collect the necessary data, it would ask them 25 questions about their current studies (e.g., “Do you prefer written exams or orals?”) and their general life as a university student (e.g., “Do you prefer to study at home or in the library?”). To maintain a continuous

‘conversation’ without interruptions, the device was programmed to accept any answers given by participants. Every five questions answered by participants, the device gave a short summary and then introduced the next thematic area of questioning (for the full interaction protocol, see Appendix A). After participants answered all 25 questions, they received verbal feedback by the device based on the experimental condition:

(a) polite feedback condition: “Thank you for sharing your knowledge and your views with me. You’re not only helping me but especially future students. You seem to have a very good idea of your field of study [...].”

(b) user blame condition: “Unfortunately something seems to be wrong with your answers. There seem to be inconsistencies, especially regarding questions 3, 7, 8, and 14. Either you spoke inarticulate, or your answers don’t make sense. Unfortunately, this means I must exclude your data from further analysis.”

(c) apologetic/self-blame condition: “I’m sorry, but unfortunately something seems to be wrong with your answers. There seem to be inconsistencies, especially regarding questions 3, 7, 8, and 14. This may however be due to me having partially misunderstood you. I am still in active development and my technology isn’t fully matured yet.”

After receiving this feedback, the device interviewed participants about their evaluation of the device. (3) During the verbal evaluation, participants were asked a series of 12 questions (e.g., “How polite was I?”) which participants were instructed to answer by verbally stating a number between one and seven (1 = *not at all*; 7 = *very*). These questions were based on the 12 items of the ‘valence towards the computer’ scale, mirroring the Likert-scale used in the online questionnaire (see section 4.2.3). Lastly, the device asked participants to switch to a desktop computer to (4) evaluate the voice assistant a second time using an online questionnaire and to report demographic information.

4.3 Results

4.3.1 Statistical Analyses: General Assumptions

In this thesis, two-way and mixed-design ANOVAS were used for data analysis. Prior to performing the statistical analyses, test assumptions were checked. Data were tested for outliers using boxplots, histograms, and Cook’s distance. The assumption of normality is commonly tested using P-P-Plots and the Shapiro Wilk test. However, t-test and ANOVA are robust against violations of normality in sample sizes ≥ 30 as the central limit theorem holds and a normal distribution can be assumed (Bortz & Schuster, 2005). As all sample sizes in this thesis were ≥ 30, there was no necessity to test for the assumption of normality (Field, 2013). Diagnostic plots of estimated residuals were employed to test for linearity and independence of observations. Levene’s test was employed to check for homogeneity of variances. Some violations of homogeneity were observed, but as cell sizes were roughly equal for all statistical analyses in this thesis, ANOVA is considered robust against this violation (Eid et al., 2017; Hussy & Jain, 2002). Thus, ANOVAs were employed, although the assumption of homogeneity of variance was not always met. For reasons of clarity and brevity, any violations of test assumptions are only reported for sample sizes with less than 30 cases or unequal cell sizes. An additional consideration for mixed ANOVAs is the assumption of sphericity for within-variables with three or more conditions. Since all within-group variables in this thesis are limited to two conditions, sphericity is not a concern (Field, 2013). All other test assumptions were met unless stated otherwise. Comparisons for pairwise differences were conducted using Tukey's HSD tests, as it strikes a good balance between conservative and liberal procedures and also controls the probability of making one or more Type I errors (Field, 2013). Šidák correction was used for simple effects analysis as it is less conservative than the Bonferroni correction (Field, 2013). To check for moderation effects of individual differences, moderation analyses were performed using the PROCESS macro by Hayes (2018), which uses ordinary least squares regression, yielding unstandardized coefficients for all effects. Confidence intervals were computed by employing bootstrapping with 5000 samples together with heteroscedasticity consistent standard errors (Davidson & MacKinnon, 1993). Prior to performing moderation analyses, linearity was checked visually using scatterplots after LOESS smoothing. For all statistical tests in this thesis, an alpha level of .05 was set. Data preparation and statistical analyses were conducted in IBM SPSS (version 26.0). Effect sizes are classified according to Cohen (1988).

4.3.2 Design and Statistical Analyses

Study 1 followed a 2 (evaluation location) x 3 (type of feedback) mixed design. For valence, competence, and friendliness evaluations, 2 x 3 mixed ANOVA were conducted with a within-subject factor (evaluation on the device itself vs. desktop computer) and a between-subject factor (polite, apologetic, and user blame feedback). Tukey HSD tests were used for pairwise comparisons. For all other dependent variables, univariate ANOVA with the between-subject-factor type of feedback were conducted. Levene's test indicated homogeneity of variances for all analyses conducted in this experiment. Unless stated otherwise, all other test assumptions were met. All effects were statistically significant at the .05 significance level.

4.3.3 Self-Report Measures

4.3.3.1 Valence towards the device

A significant main effect for type of feedback was found regarding general valence towards the smart speaker device, $F(2, 97) = 8.24$, $p < .001$, partial $\eta^2 = .15$. The effect size is classified as large according to Cohen (1988). Pairwise comparisons using the Tukey HSD tests indicate that in line with H_{1a}, there was a significant difference (p < .001) with impolite smart speakers blaming participants being devaluated compared to their polite counterparts (.689, 95%-CI[.283, 1.09]). Additionally, smart speakers providing apologetic feedback for a failed interaction did not significantly differ from the polite ($p = .227$) and impolite ($p = .071$) conditions. Thus, H_{1b} was also accepted.

4.3.3.2 Friendliness

As expected, a significant main effect for type of feedback was found for friendliness ratings, $F(2, 97) = 11.34$, $p < .001$, partial $\eta^2 = .19$, with a large effect size. Tukey HSD tests indicate a significant difference ($p < .001$) between ratings for smart speakers that blamed participants and polite devices (.928, 95%-CI[.464, 1.39]), confirming H_{2a}. Smart speakers offering apologetic feedback for the failed interaction did not significantly differ from either the polite ($p = .106$) or the impolite ($p = .072$) conditions, also confirming H_{2b}.

4.3.3.3 Competence

No significant main effect for type of feedback was found for competence ratings, $F(2, 97) = 2.85$, $p = .063$, partial $\eta^2 = .056$. An impolite smart speaker was not assessed to be less competent than a polite device. Tukey HSD tests still indicate that there is a tendency ($p = .057$) towards the main difference being between the polite and impolite conditions (.451, 95%-CI[-.010, .912]). Still, H_{3a} must be rejected. As smart speakers offering apologetic feedback did not significantly differ from either the polite ($p = .777$) or the impolite ($p = .265$) conditions, H_{3b} was accepted.

4.3.3.4 Performance of the device

The main effect for type of feedback was also significant for the evaluation of the smart speakers performance, $F(2,97) = 8.15$, $p < .001$, partial $\eta^2 = .144$. The effect is considered large. Tukey HSD tests revealed that again a significant difference ($p < .001$) was found between evaluation of impolite smart speakers and polite devices (1.19, 95%-CI[.487, 1.89]). Thus, H_{4a} was accepted. Again, devices that gave apologetic feedback for a failed interaction did not significantly differ from either the polite ($p = .161$) or the impolite ($p = .110$) conditions, also confirming H_{4b}.

4.3.3.5 Attitude towards the device

In confirmation of H_{5a}, analysis of the general attitude towards the smart speaker revealed a significant main effect for type of feedback, $F(2,97) = 4.99$, $p < .05$, partial $\eta^2 = .093$. Tukey HSD tests indicated that there was a significant difference ($p = .009$) with impolite devices being devaluated compared to their polite counterparts (.547, 95%-CI[.135, .959]). No significant differences were found for apologetic feedback and polite feedback ($p = .260$) or apologetic feedback and impolite feedback ($p = .315$). H_{5b} was also accepted.

4.3.3.6 Self-performance

No significant main effects were found for type of feedback on participants evaluations of their self-performance during the interaction, $F(2,97) = .314$, $p = .73$, partial $\eta^2 = .006$. Both H_{6a} and H_{6b} must be rejected.

4.3.4 Behavioral Measure: Interviewer-Bias

H_7 focused on the within-group differences between direct verbal evaluations on the device itself and written evaluations on another device based on the interviewer-bias (Akgun et al., 2005; P. Brown & Levinson, 1978; Finkel et al., 1991; Nass et al., 1999). Evaluations given directly to the device (verbally) and on a separate PC (written) can be found in Table 3.

Table 3. Mean values (standard deviations) of voice assistant evaluation by condition and dependent variable

	Type of feedback					
	polite		apologetic		user blame	
Evaluation location	device itself	separate PC	device itself	separate PC	device itself	separate PC
Friendliness	5.35 (.72)	4.96 (1.13)	4.86 (.77)	4.54 (1.17)	4.51 (.81)	3.94 (.85)
Competence	5.32 (.73)	4.99 (1.30)	4.88 (.70)	5.16 (.99)	4.72 (.85)	4.69 (1.02)
Valence	5.34 (.65)	4.98 (1.01)	4.87 (.65)	4.85 (.80)	4.62 (.78)	4.31 (.97)

4.3.4.1 Valence towards the device

In line with expectations, the main effect of evaluation location was significant for overall valence towards the smart speaker device, $F(1, 97) = 7.87$, $p = .006$, partial $\eta^2 = .075$. This effect size is classified as medium according to Cohen (1988). In support of H_7, a smart speaker device asking for an evaluation itself (see Figure 5) resulted in better overall valence ratings than an online questionnaire on a separate PC (see Figure 6). No significant interaction between evaluation location and type of feedback was observed, $F(2, 97) = 1.52$, $p = .224$, partial $\eta^2 = .03$. H_7 was accepted.

4.3.4.2 Friendliness

As predicted, the main effect for evaluation location was also significant for friendliness ratings, $F(1, 97) = 25.60$, $p < .001$, partial $\eta^2 = .2$. The effect size is classified as large. The device itself asking for a verbal evaluation resulted in significantly higher friendlier ratings compared to the written assessment on a separate desktop PC. Again, there was no significant interaction between evaluation location and type of feedback, $F(2, 97) = .78$, $p = .46$, partial $\eta^2 = .016$. H_{7a} was accepted.

4.3.4.3 Competence

Contrary to expectations, no significant main effect of evaluation location was found on competence ratings, $F(1, 97) = .074$, p = .786, partial $\eta^2 = .001$. While there was no significant interaction between evaluation location and type of feedback, a tendency towards it was found, $F(2, 97) = 1.42$, $p = .071$, partial $\eta^2 = .05$. Nevertheless, H_{7b} must be rejected.

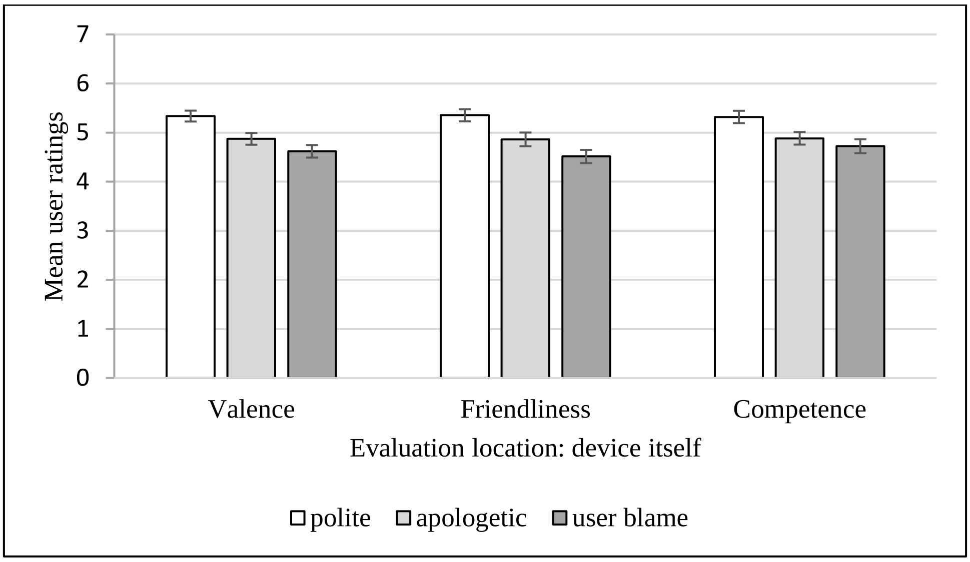


Figure 5. Evaluations of the voice assistant given on the device itself (source: own figure)
Note. Error bars indicate 95% CI.

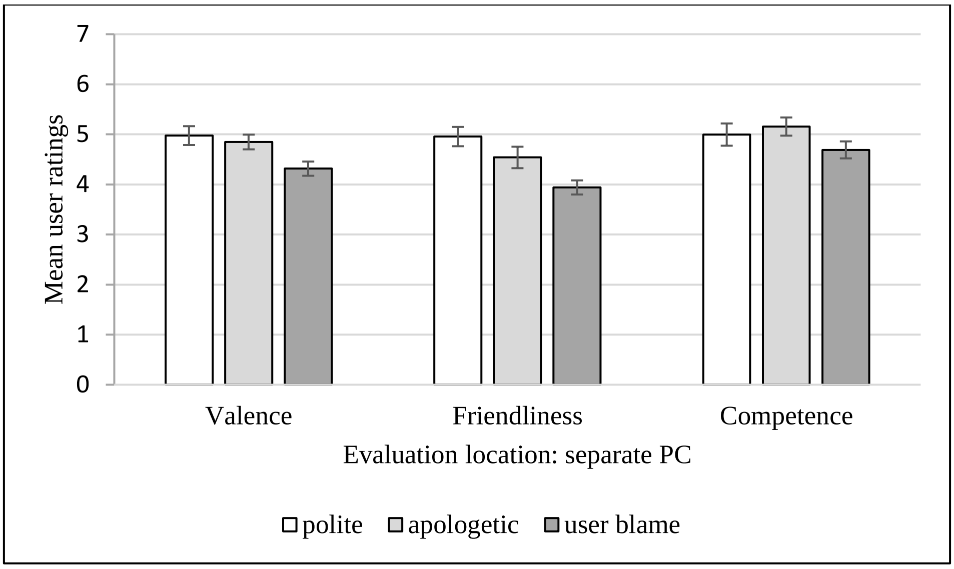


Figure 6. Evaluations of the voice assistant given on a separate PC (source: own figure)
Note. Error bars indicate 95% CI.

4.3.5 Exploratory Analyses: Individual Differences

The focus of this study was on the effect of politeness on social reactions towards a voice assistant in general. However, in CASA research in general there is a clear lack of consideration for individual differences (see section 3.2.11) and thus, exploratory analyses were conducted for the influence of prior experience with voice assistants. For reasons of clarity and brevity, only significant effects are reported here. For a description of statistical procedures and test assumptions see section 4.3.1.

In prior CASA research, Johnson et al. (2004) opted to perform a median split to divide their sample between ‘high experience’ and ‘low experience’ with desktop computers to examine the effects of experience on media equation effects. A similar procedure for voice assistants was considered for this study. However, due to the novelty of the technology when this study was conducted, out of 100 participants, 50 of them (50%) reported never having used either a smart speaker device or a voice assistant prior to this experiment. Therefore, instead of making a distinction between low and high experience, a median split was performed for ‘no prior experience’ and ‘prior experience’ with voice assistants and used to create a new variable used in subsequent analyses. The procedure of a median split has been criticized in literature before (J. Cohen, 1983) but is still considered a robust procedure for independent variables (Iacobucci et al., 2015b, 2015a). Analysis revealed a median of 0.125 years of experience, resulting in two groups of 50 participants each. For each dependent variable, a 2 x 3 ANOVA with the between-subject-factor type of feedback (polite, apologetic, and user blame) and the between-subject-factor experience (no prior experience and prior experience) was conducted. One significant main effect was found for prior experience: participants who had used voice assistants before reported significantly higher self-evaluation ratings when compared to participants with no prior experience, $F(1, 97) = 4.88$, $p < .05$, partial $\eta^2 = .05$. No other significant main effects were detected for prior experience. Additionally, no significant interaction effects were found between prior experience and feedback type.

4.4 Discussion

The experimental study presented in the previous sections aimed to investigate if social behavior related to politeness effects that were examined in previous CASA studies (Akgun et al., 2005; Carolus, Muench, et al., 2019; Fogg & Nass, 1997b; Nass et al., 1999) can also be examined in human-smart speaker interaction. As the interaction with smart speakers differs substantially from devices that have been the subject of traditional CASA research, these results cannot simply be transferred but need to be reexamined thoroughly. While desktop computers were controlled via keyboard and mouse, voice assistant devices are almost exclusively based on voice interaction. Users prompt the device via voice input and receive answers via voice output, a form of interaction inherently evocative of a human-human conversation compared to earlier technology. As all conditions for CASA effects are fulfilled (see section 3.5; Nass & Moon, 2000), an empirical analysis of potential media equation effects with these devices was logical and necessary. In general, results revealed that the social norm of politeness is not only reflected in self-report measurements resulting in significantly more negative ratings for an impolite device, but participants also adopted the interviewer-bias: a voice assistant was evaluated significantly better when the device directly asked for the evaluation compared to ratings given via an online questionnaire on a separate desktop computer. Analysis of assessments based on evaluation location revealed that a smart speaker asking for feedback directly was rated significantly better regarding both overall valence towards the device and its perceived friendliness. However, no significant differences were found in competence ratings. Interestingly, a voice assistant offering apologetic feedback did not differ significantly from either of the other two conditions. Apologetic feedback resulted in lower ratings than polite feedback but higher ratings than impolite feedback for valence, friendliness, and competence, thus showing that using a politeness strategy can mitigate face-threatening acts in human-voice assistant interaction that would usually lead to negative evaluations, in this case informing the participant that their data is inadequate for further analysis.

For type of feedback, results revealed that an impolite device was significantly devaluated when it communicated a failed state by blaming the user compared to polite or apologetic assistants. Moreover, this devaluation was not limited to friendliness ratings even though only friendliness was manipulated via type of feedback. Apart from feedback given to participants, the voice-assistants general performance and behavior was identical across all conditions. Still, participants in the impolite condition also reported lower scores for its performance and their general attitude towards the device, suggesting a general ‘punishment’ of impolite behavior. Again, these effects can be observed in human-human interaction, where violations of social norms such as politeness are sanctioned by the social system (P. Brown & Levinson, 1978; Culpeper, 1996). As the name implies, voice assistants or ‘virtual butlers’ (Payr, 2013) are expected to follow orders and be polite and supportive by default (Luger & Sellen, 2016). Consequently, a device failing to fulfill these expectations by acting in a way that can be perceived as impolite is at risk of causing frustration for users. Just like with feedback location, competence ratings were once again unaffected by the manipulation of feedback type. While there was no direct manipulation of the voice assistant’s competence between the experimental groups, the device offered apologetic feedback for a failed interaction in one of the three groups which could be interpreted by participants as a lack of competence. Statistical data analysis affirms that the lack of a main effect for the factor evaluation location regarding competence ratings can indeed be attributed to the apologetic feedback condition. Both the polite and the impolite user blame conditions resulted in higher competence ratings given on the device itself. However, the opposite was the case in the apologetic feedback condition: while not statistically significant, there was a tendency towards higher competence ratings given on a separate PC. There is a way to explain this discrepancy: it is conceivable that participants interpret apologetic feedback by the device as a lack of competence, thus it is possible that the low competence ratings given on the device itself effectively reflect conformity with the device on its self-evaluation rather than a devaluation of it. These results are congruent with previous research: while apologetic feedback provided by a computer resulted in a more comfortable experience for users in a previous study, it was also interpreted as the computer's inability to carry out its user's demands (Akgun et al., 2005) which can be perceived as incompetence. Previous research also revealed that modest comments and self-blame are usually believed to be true (Nass & Steuer, 1993; Reeves & Nass, 1996) even when they originate from a computer (Nass & Brave, 2005). So, while a person or device blaming itself for mistakes might seem more likeable, this comes at the cost of perceived competence, as people effectively acknowledge the self-blame to be true. This effect needs to be carefully considered by both developers and designers of voice assistants. The feedback a voice assistant gives its users seems to substantially change not only its perception but also verbal behavior towards it. More work is required to understand what users base their evaluations on and how they are influenced by social cues sent by the device. This is especially relevant for communicating failed states to users, as both approaches tested in this experiment led to unfavorable results: blaming the failure on users resulted in general devaluation of the system including its unchanged performance while offering an apology potentially comes at the expense of the systems perceived competence.

There is also another important consideration to be made regarding competence ratings. As this study employed the standard German (female) Alexa voice of the Amazon Echo Plus device for all voice interactions, gender stereotypes could have been a relevant factor in terms of competence assessments. According to gender stereotypes, males are usually ascribed more competence compared to females (see section 3.2.5, Fiske et al., 2007, 2018) which has not only been shown to have effects in human-technology interaction (see section 3.3.4) but is especially relevant in the area of voice assistants that traditionally employ primarily female voices by default (Ernst & Herm-Stapelberg, 2020; Tolmeijer et al., 2021). Due to their high relevancy for this dissertation, gender stereotypes in human-voice assistant interaction were systematically examined in-depth in a separate laboratory experiment (see section 6).

While participant's prior experience with voice assistants had no significant effect on either social behavior towards the device or evaluations given about the device, a significant effect on self-performance ratings was found. Participants that reported prior experience with voice assistants rated their self-performance significantly higher compared to participants with no prior experience. A possible explanation for this effect might be to consider how participants handle being told that their interaction with the device resulted in failure. Subjects with no prior experience were more inclined to believe the voice assistant when it told them their data was inadequate for further analysis thus assuming they are responsible for the failure. Subjects with prior experience are inclined to believe that their input was flawless as they knew how to operate a voice assistant device. Thus, they attribute the failed state to the device and not their own performance. Participants with no prior experience being more convinced by what the device told them might indicate a conformity effect resulting from informational social influence (M. Deutsch & Gerard, 1955). Due to their lack of experience, they could be more impressed by the new technology, thus ascribing more trustworthiness and expertise to the voice assistant as a source, leading to an increase in the social influence exerted by the device (Fogg & Tseng, 1999; McCroskey et al., 1974).

4.4.1 Limitations

Using a within-subjects design for the factor evaluation location combined with self-reports can raise the concern of demand characteristics and social desirability. Still, as with previous CASA research, participants were unaware of both the purpose of the study (as they were given a cover story about the development of a Skill for other university students) and the social reactions expected of them (in this case the interviewer-bias) and thus unable to act accordingly. As an additional precaution, the ratings given for the factor evaluation location were spaced as far apart as possible: the verbal evaluation of the voice assistant regarding overall valence was followed by the online questionnaire that presented participants with every other scale before asking for a second evaluation of valence towards the smart speaker at the very end. This was done so that even if participants were influenced by their ratings given on the device itself (e.g., trying to repeat the same ratings), they likely would have forgotten the exact ratings given due to the time elapsed since that point.

From a technological perspective, interactions between participants and the smart speaker remained artificial and were somewhat reverse to real-life interactions with voice assistants: usually users ask questions or give directives, and the assistant answers. To allow for a more controlled and linear interaction, the voice assistant instead acted as the initiator in this experiment, which somewhat limits external validity. Still, there are many Skills available for Amazon Echo devices that follow a similar dynamic. Also due to technical limitations, the voice assistant was unable to produce customized feedback based on the answers participants gave during the interaction. Instead, participants were randomly assigned to one of the three feedback conditions (polite, apologetic, or impolite) thus resulting in random feedback given by the device. Therefore, answers given by participants in the prior interaction did not influence the feedback they received leading to potential mismatches between participant's perceptions of their own performance during the interaction and the feedback given to them by the voice assistant. This could in turn have influenced their evaluations of the device. Further speaking from a technical standpoint, it should be noted that the chatbot Dexter (see section 4.2.2) was used to simulate the software side of the interaction and the Amazon Echo Plus device technically only acted as a text-to-speech service, loudspeaker, and microphone. Due to this circumstance, the ASR component used was that of the Dexter platform which is far less advanced than Amazon's internal solution. This led to a variety of speech recognition errors during some of the interactions. To solve this problem, all participants that had faulty interactions were excluded from further data analysis (see section 4.2.1).

4.4.2 Implications and Conclusion

Although not appropriate from a rational point of view, the results of this initial study suggest that humans interact in a social way with voice assistant devices: based on the interviewer-bias known from human-human interaction, a smart speaker was rated significantly better when it asked for the evaluation itself compared to a separate evaluation. In addition, an impolite smart speaker was 'punished' by general devaluations during self-reports that even extended to aspects that were not manipulated such as its perceived performance during the interaction or the general attitude towards the device. This has multiple implications that need to be considered. While there are not many situations in which a voice assistant has good reason to act impolite, there are certain contexts where it might be necessary to forgo politeness strategies. Voice assistants employed in work, fitness, or education contexts would need to communicate failed states (e.g., not reaching certain educational or fitness goals) to their users (Nass & Brave, 2005; Nass & Yen, 2010). In these contexts, sacrificing perceived friendliness for competence might even be more important as the device would adopt the social role of a teacher or a trainer. Still, if the main goal of user-centric design is to create an environment for users to feel comfortable in and enjoy themselves, apologetic statements and feedback from the device are important factors to consider. Just like in human-human interaction, apologizing for making a mistake seems to be the most effective way for a voice assistant device to avoid overall devaluation after communicating a failed state. In terms of the interviewer-bias, social desirability is a factor that needs to be considered when designing voice assistant devices that ask for feedback from their users. Users might not give honest feedback if a device asks direct questions as negative answers or evaluations represent a potential face-threatening act. This is especially relevant when it comes to evaluations of certain features, applications, or services a voice assistant provides to users. As in many mobile applications, users of voice assistants are periodically requested to modify settings or rate certain features verbally directly by the assistant. These requests are then liable to yield biased responses as users automatically adopt politeness strategies instead of giving objective answers. Thus, designers need to be aware of users detecting social cues, both intended and unintended. Even if not intended by the designer and seemingly irrational, a device might automatically be ascribed intentionality which can potentially result in counterproductive behavior. In conclusion, the results of experiment 1 provide initial empirical evidence that users treat voice assistants as social actors. Nevertheless, these findings also raise further questions: considering the huge variety of social cues voice assistants are technically able to manifest through their voice (e.g., group membership, gender, personality, etc.), this study only marks a first starting point for analyzing human-smart speaker interaction. Many of these cues have proven to be highly relevant for human-technology conversation and need to be carefully re-evaluated for voice-based technology that is employed in a social context. Consequently, three more experimental laboratory studies were conducted as part of this thesis to further examine media equation effects based on various social cues in human-smart speaker interaction.

5 Experiment 2: Effects of Interdependence on Team Affiliation in Human-Smart Speaker Interaction

5.1 Study Outline and Hypotheses

As outlined in earlier sections (see 3.1.1, 3.1.2, 3.2.4, and 3.3.4), research has shown that team affiliation in human-computer interaction can be induced using minimal social cues (Johnson & Gardner, 2007; Nass et al., 1996; Nass, Fogg, et al., 1995). It has also been shown that people display prosocial behavior towards computers under various conditions (e.g., Fogg & Nass, 1997a). Consequently, this study examines the effects of interdependence on team affiliation in human-voice assistant interaction by transferring the operationalization by Nass et al. (1996) to smart speaker devices with a crucial modification: while subjective self-reports indicated an effect of interdependence on team affiliation, no behavioral measures were employed by Nass et al. (1996). This study introduces an additional behavioral measure to examine if team affiliation also leads to increased prosocial behavior towards a smart speaker device. To further expand on previous research, a second factor was considered in this experiment: voice recognition errors are a major topic for conversational interfaces both in development as well as in literature (see section 3.4.3.3). Nass and Moon (2000) also theorized that additional cues indicating failed states exclusive to technological devices (such as 'crashing' for desktop PCs) might influence media equation effects as they remind users of the artificiality of the interaction and situation. While the results of experiment 1 (see section 4) indicate that apologetic feedback can mitigate the effects of having to communicate system failure to users, the alleged failure was merely stated to participants instead of having any perceivable consequences. Instead of indicating a failed interaction via impolite or apologetic feedback as in experiment 1, this study opted to use two deliberately implemented voice recognition errors that lead to a visual failed state as an additional manipulation. Again, in line with the general aims of this dissertation, study 2 again sought to answer three questions: (1) do people generally show social reactions towards a voice assistant based on perceived interdependence and do recognition errors prevent these reactions? (2) Can this effect also be assessed using an additional behavioral measure, in this case, prosocial behavior? (3) Do individual differences influence these effects?

Following the *Minimal Group Paradigm*, interdependence between humans and computers can be induced by telling subjects that they are being assessed on their joint performance with the computer thus creating a sense of team affiliation (Johnson & Gardner, 2007; Nass et al., 1996). Based on the results obtained by Nass et al. (1996) that team members are generally assessed more positively, it is assumed that smart speakers are also rated more favorable after team affiliation has been induced this way.

H_1: Smart speaker devices in the interdependent condition are evaluated significantly better regarding their overall valence compared to devices in the non-interdependent condition.

H_{1a}: Smart speaker devices in the interdependent condition are evaluated significantly better regarding their friendliness compared to devices in the non-interdependent condition.

H_{1b}: Smart speaker devices in the interdependent condition are evaluated significantly better regarding their friendliness compared to devices in the non-interdependent condition.

H_2: Smart speaker devices in the interdependent condition are evaluated significantly better regarding their overall performance compared to devices in the non-interdependent condition.

In line with research indicating that the occurrence of interaction errors is negatively correlated with the users' general evaluation of a system (Oulasvirta et al., 2006), it is also postulated that participants will devaluate a smart speaker that indicates a failed state to them.

H_3: Smart speaker devices in the error condition are evaluated significantly worse regarding their overall valence compared to devices in the error-free condition.

H_{3a}: Smart speaker devices in the error condition are evaluated significantly worse regarding their friendliness compared to devices in the error-free condition.

H_{3b}: Smart speaker devices in the error condition are evaluated significantly worse regarding their competence compared to devices in the error-free condition.

H_4: Smart speaker devices in the error condition are evaluated significantly worse regarding their overall performance compared to devices in the error-free condition.

As mentioned in study 1, a system is perceived to be in an incorrect state when an error occurs. This incorrect state then prevents the achievement of an intended goal for the user, resulting in frustration (Ceaparu et al., 2004) and lower perceived self-performance (Baumeister & Tice, 1985; Stotland & Zander, 1958).

H_5: Errors leading to a failed interaction result in a significantly lower evaluation of participants' self-performance compared to an error-free interaction.

There is evidence in literature that interdependence leads to increased cooperation among group members (Crawford & Haaland, 1972) as well as a higher willingness to help each other (Van der Vegt & Van de Vliert, 2005; Wageman, 1995). The inherent obligation to respond to helpfulness is defined under the norm of reciprocity (Gouldner, 1960). Considering the findings from Fogg and Nass (1997a) showing that manipulating interdependence can influence subject's prosocial behavior towards a computer, it is postulated that:

H_6: Participants in the interdependent condition show more prosocial behavior towards a smart speaker device than participants in the non-interdependent condition.

Research indicates that the occurrence of interaction errors can lead to a decrease in prosocial behavior (Velez, 2015). Additionally, Nass and Moon (2000) theorize that a technological device indicating a failed state to its users can actively counteract media equation effects due to it being a reminder of the artificiality of the situation. Based on these deliberations, it is postulated that:

H_7: Participants in the error condition show less prosocial behavior towards a smart speaker device than participants in the non-error condition.

Analogous to experiment 1 (see section 4), the adoption of an interviewer-bias (Finkel et al., 1991; Nass et al., 1999) was assumed, resulting in better ratings given on the device itself. Instead of overall valence towards the device as in experiment

1, the oral ratings given directly to the voice assistant in experiment 2 were focused on its performance.

H_8: Participants will evaluate a smart speakers' performance more positively if the device asks for the evaluation itself compared to a desktop PC asking.

Exploratory considerations. The individual difference measured in this study was self-efficacy, more specifically general self-efficacy as well as a variation of computer self-efficacy. Self-efficacy is defined as the conviction of an individual to be able to successfully complete a task through appropriate behavior (Bandura, 1977). This concept has been further specified for the domain of HCI as 'computer self-efficacy' to also integrate individuals' expectations of outcomes when using computers (Compeau & Higgins, 1995). Further, research revealed that subjects with high self-efficacy beliefs are less likely to accept repeated negative feedback (Nease et al., 1999). Based on these findings, it can be assumed that individuals with high computer self-efficacy will be confident that their behavior was appropriate regardless of errors during their interaction with a voice assistant as they put more stock into their own abilities compared to the device. In comparison, participants with low computer self-efficacy beliefs expect the interaction to fail regardless and thus are more prone to blame themselves for errors. Thus, the influence of both general self-efficacy as well as computer self-efficacy on media equation effects are examined in this study.

RQ_1: Does participant's general self-efficacy moderate their social reactions towards a voice assistant?

RQ_2: Does participant's computer self-efficacy moderate their social reactions towards a voice assistant?

5.2 Methods

5.2.1 Participants

A total of 134 university students participated in the laboratory experiment. A majority of participants were recruited using the internal recruitment system of the Institute Human-Computer-Media in exchange for course credit. After excluding all cases that contained technical problems such as unintended recognition errors (n = 14), the final sample consists of N = 120 participants, including 39 (32.5%) male and 81 (67.5%) female subjects. Participants ranged in age from 18 to 30 years (*M* = 21, *SD* = 2.23). In terms of education, 86.7% reported a higher education entrance qualification. Twelve participants already held a university degree (10%). Two participants reported being qualified with an advanced technical college certificate (1.7%). A completed apprenticeship (0.8%) and an intermediate school-leaving certificate (0.8%) were reported by one participant each. Written informed consent was obtained from each participant before the study.

5.2.2 Stimulus Materials

5.2.2.1 Hardware and Software

As in experiment 1, an Amazon Echo Plus (1. Generation) running a custom-made Skill was used to allow for a controlled human-smart speaker interaction. For this experiment, the Alexa Skill was programmed using the *Alexa Skills Kit SDK* for Python and Alexa APIs for Python retrieved from *GitHub*. This allowed for the voice interaction to use Amazons' native German ASR and NLP to interpret inputs and vocalize predetermined outputs resulting in a higher quality of voice recognition compared to experiment 1.

5.2.2.2 Voice Interaction: The Spy Game

The general interaction between the voice assistant and participants from a technical standpoint closely resembled that of experiment 1 (for a detailed description of the general interaction see section 4.2.2). Since non-intended recognition errors could not generally be excluded due to the problems inherent in speech recognition technology in general (see section 3.4.3.3), a safety mechanism was programmed to automatically output the next response if the detected request could not match an execution specification provided in the code after three attempts. Instead of a series of questions and answers, participants in experiment 2 were asked to play a *spy game* reminiscent of popular *escape room* games into which the interaction with the voice assistant was interwoven as a source of information to solve simple puzzles and to fulfill helpful tasks. The game, which was used as a pretense to disguise the actual purpose of the experiment, was developed specifically for this study, and consisted of a storyline written in HTML so that it could be accessed in a web browser. The spy game was presented to participants on a laptop where they also had to input either the correct answers to the various puzzles or confirmation that they completed a task (see Figure 7 for an example). The smart speaker was standing on a table next to the laptop so it could be consulted at any time by participants. The voice used for the interaction was the default German Amazon Alexa voice. For the purposes of the spy game, the voice assistant device was referred to as *Alexa* during any prompts given to participants. For a protocol of the entire interaction, see Appendix D.

Figure 7. Example page from the spy game as presented to participants (source: own figure)

5.2.2.3 Interdependence/Team Affiliation

Team affiliation was manipulated at three different points during the experiment. (1) The first indicator of team affiliation was embedded into the introduction given by the researcher. For the control group, the instructions they were given stated that the voice assistant was merely used to gather information needed to complete the game and that participants are evaluated on their individual performance exclusively. The interdependence group was told that the rating given to them at the end of the game is based on their joint performance with the voice assistant. (2) The game itself began with a short vocal introduction by the voice assistant explaining the following procedure. This introduction was also manipulated to induce team affiliation depending on the condition. In the non-interdependent condition, participants were addressed with neutral statements such as: “during the game, you have to give me tasks which are bold on the screen to progress to the next action”. In the interdependent condition, the voice assistant made references to the game being a collective effort: e.g., “to complete this mission successfully, we depend on each other. This means that we work together to solve tasks […] to get to the next action”. (3) To enhance the subject’s perception of team affiliation, the feedback given after completing the game was also modified based on condition. It differed in the plural phrase "you [participant and voice assistant device] worked very well together" (interdependent) and in the singular phrase "you interacted very well [with the voice assistant]" (non-interdependent).

5.2.2.4 Intentional Voice Recognition Errors

The manipulation of recognition errors was directly implemented into the spy game. In the non-error condition, the game was programmed so that it could be executed without errors. Participants in the error condition encountered two recognition errors, which were deliberately implemented to simulate unexpected and unsatisfying responses of the voice assistant. The first recognition error led to an error prompt that was repeated twice by the device before giving the correct response on the third try. This was done to create a spiral depth (the number of repetitions required to resolve an error) of two trials to induce a sense of failure in participants but not completely demotivate them before resolving the error (Oviatt & VanGent, 1996). In addition, an even more noticeable system failure was simulated at the end of the game: participants were given the final task to set the light in the laboratory to the color “Fax” using smart bulbs installed in the laboratory setting (the in-game purpose given to participants was that the light would allow them to read a secret message). In the non-error condition, the request was processed correctly and the light in the room turned blue. The subjects also received verbal feedback that they had successfully completed the mission. In the error-condition, the voice assistant responded with an acoustic error warning signal followed by a spoken phrase that the color “Lachs” [Salmon] had been turned on instead (resulting in red light instead of blue) and that the mission failed because of an incorrect request. The color to be set (Fax) and the obviously incorrect input (Lachs) were chosen based on their linguistic similarity in German, thus maintaining the pretense that the voice assistant misunderstood the request. The change of the light colors was used as manipulation so that the visual feedback emphasizes the presence of a system error.

5.2.3 Self-Report Measures

5.2.3.1 Valence, friendliness and competence

Like in experiment 1, participants were asked to rate their general valence towards the smart speaker device using the twelve items of the valence towards the computer scale (see section 4.2.3). Again, selections were made on a 7-point Likert scale (1 = *does not apply at all* to 7 = *is absolutely true*). The scale exhibited a high reliability of Cronbach's Alpha = .84. The index is once again distinguished by the two subscales 'competence' (α = .80) and 'friendliness' (α = .81) with six items each.

5.2.3.2 Performance of the device

Performance of the smart speaker was again rated on the scale introduced by Johnson et al. (2004) previously used in experiment 1. The same 6 items were again asked on a 7-point Likert scale (1 = *not at all* to 7 = *very*). Reliability for the performance evaluation was excellent, α = .91.

5.2.3.3 Self-performance

As in experiment 1, self-performance was evaluated on the scale introduced by Johnson et al. (2004). Responses to the five items were made on a 7-point Likert scale (1 = *not at all* to 7 = *very*). Reliability for self-performance was high, α = .89.

5.2.3.4 Self-efficacy

Self-efficacy was measured using two scales, one for general self-efficacy expectation and one for computer-related self-efficacy. For a general assessment, the ten statements of the *General Self-Efficacy Expectations Scale* published by Schwarzer and Jerusalem (1999) were presented to the participants (e.g., "I can find a solution to every problem") measured on a 4-point Likert scale as recommended by the authors, ranging from 1 = *is not correct* to 4 = *is absolutely right*. The reliability of this scale was acceptable, α = .77. As the often-cited computer self-efficacy scale by Compeau and Higgins (1995) is focused on learning contexts, a more recent computer self-efficacy scale created by Howard (2014) was adopted for this experiment instead. Responses to the twelve items (e.g., "I can usually handle whatever computer problem comes my way") were given on a 7-point Likert scale (1 = *do not agree at all* to 7 = *fully agree*). This scale exhibited excellent reliability of α = .94.

5.2.3.5 Manipulation check

Two manipulation check questions were implemented at the of the questionnaire to determine the effectiveness of the manipulations of error and interdependence. First, to check if interdependence was induced, participants were asked if they felt like they were evaluated based on their joint performance with the voice assistant. Secondly, participants were asked if they encountered any errors during the interaction to check if the deliberately implemented errors were noticed. The answers given to these open-ended questions indicate that both manipulations were generally successful and provided no reasons to exclude specific participants from further analyses.

Table 4. Overview of scales and their internal consistencies for the variables in Experiment 2 (N = 120)

Scale	Subscale	Cronbach's α	Items
Valence towards the device		.84	12
	Competence	.80	6
	Friendliness	.81	6
Performance of the device (separate PC)		.91	6
Performance of the device (on device itself)		.89	6
Self-performance		.89	5
General Self-Efficacy Scale		.77	10
Computer Self-Efficacy		.94	12

5.2.4 Behavioral Measures

5.2.4.1 Interviewer-bias

To determine if the performance of the voice assistant is assessed differently depending on if the device is asking directly, data on its performance was collected twice analogous to the procedure in experiment 1. The performance of the voice assistant scale by Johnson et al. (2004) was chosen for this purpose (also see section

5.2.3.2). The scale was transferred into a Skill in which the items were provided as speech output from the device. Participants were instructed to verbally answer the questions with the numerical values 1 (*not at all*) to 7 (*very*). Consistency of the verbal evaluation was high with $\alpha = .89$.

5.2.4.2 Prosocial behavior towards the device

Helpful behavior towards the voice assistant device was operationalized using an objective behavioral measure based on the approach of Sandoval et al. (2016). After participants had finished the game either successfully or with a failed state depending on the condition, the voice assistant asked for their help in an optional task. With the excuse of extending an internal database, participants were requested to name German cities that could serve as potential backdrops for the spy game played previously. Participants were informed that this step was optional and not needed to complete the study. An immaterial measure was used to ensure that prosocial behavior is an immediate reaction to the voice assistant even if no material gains can be expected (Fehr & Gächter, 2000). It was assumed that based on the previous interaction, during which interdependence was induced for half of the participants, the number of cities mentioned would differ between participants in the interdependent and the non-interdependent experimental group and participants in the error and non-error condition. Thus, the absolute number of cities named by participants in the optional task after the voice assistant asked them for help was considered as a behavioral measure for further analyses.

5.2.5 Procedure

Participants were randomly assigned to one of the four experimental conditions (interdependence/non-interdependence x error/non-error). On arrival, participants were asked to sign a consent form to store their data and generate an identification code so that the personal data could be deleted upon request. Participants were told that the purpose of the study was an interim evaluation of an interactive game developed for smart speakers. This was done to obfuscate the actual manipulation and thus counteract demand characteristics and avoid normative behavior. Participants were then advised to follow the instructions given to them during interaction with the Amazon Echo Plus device, which guided them through the experiment consisting of five distinct parts: (1) warm-up, (2) main interaction, (3) behavioral measure, (4) evaluation on the device itself and lastly the (5) evaluation on a separate computer.

(1) As in experiment 1, participants were then left alone with the voice assistant device for five minutes to familiarize themselves with the device and test its general functionality. After the warm-up phase, the researcher re-entered the room to introduce the main experiment. This introduction differed based on the experimental group and included the first interdependence manipulation (see section 5.2.2.3). Participants were then handed a laptop on which the spy game was played in conjunction with the smart speaker device. Afterwards, the researcher left the laboratory again so only the participant and the voice assistant device remained in the room. (2) The spy game started with a short explanation of the general procedure given to participants by the voice assistant. This introduction also contained the second interdependence manipulation based on the experimental group. Participants then navigated through the game using the laptop provided to them as a series of web pages. Each page contained both text telling the story and a puzzle that could only be solved by asking the voice assistant a question. The answer provided by the voice assistant then had to be entered into a mandatory input field to continue to the next page. A total of 21 tasks had to be completed to finish the game. Tasks and correct answers were identical for every experimental condition. After completing the final task, the game ended with either success or failure depending on the experimental condition (non-error vs. error) followed by corresponding feedback on their performance presented to participants via text on the laptop screen. This feedback contained the final manipulation of interdependence: participants in the interdependence condition were told that this rating was based on their joint performance with the voice assistant. (3) To assess whether participants would show prosocial behavior towards the voice assistant, the spy game was followed by a second interaction with the device. Participants were told that they could help the voice assistant in building a database of German cities in which the spy game could alternatively take place. To this end, participants could name any number of cities before moving on to the final step of the interaction. Any city named was compared with an internal database of over 1000 German cities and – if recognized by the device – repeated back to participants for confirmation. This was deliberately implemented to give participants the impression that the input provided by them was actually being processed by the device to increase the

believability of the cover story. The device then thanked participants for the input and asked if they want to name an additional city or proceed with the next step of the experiment. Subjects were able to repeat this process as many times as they liked. (4) After participants chose to end the optional task, the device asked them for oral feedback on its performance during the entire interaction. Like in experiment 1, the assistant asked questions that participants needed to answer by verbally stating a number between one and seven (1 = *not at all*; 7 = *very*). The 6 questions were based on the 6 items of the *performance of the voice assistant* scale adapted from Johnson et al. (2004), mirroring the Likert-scale used in the online questionnaire. Lastly, the device asked participants to switch to a desktop computer to (5) evaluate the voice assistant a second time using an online questionnaire and to report demographic information.

5.3 Results

5.3.1 Statistical Analyses

This study followed a 2 (interdependence: team affiliation or no team affiliation) x 2 (error: failed interaction due to recognition error or successful interaction) design. A series of two-way between-subjects ANOVAs were employed with two between-subject-factors (interdependent/non-interdependent and error/non-error). In addition, a three-way mixed ANOVA with two between-subject factors (error, interdependence) and one within-subject factor (device used for evaluation: voice assistant or separate computer) was conducted to check for an interviewer-bias. The procedure prior to performing statistical analyses and the handling of test assumptions was identical to experiment 1. For the sake of brevity and reduced redundancy, refer to section 4.3.1 for a comprehensive elaboration. A significance level of $\alpha = .05$ was set for all calculations. All test assumptions were met unless stated otherwise. Descriptive statistics are noted in Appendix G.

5.3.2 Self-Report Measures

5.3.2.1 Valence, friendliness and competence

ANOVA revealed no significant main effect for interdependence on participants overall valence towards the voice assistant, ($F(1,116) = 2.47$, $p = .119$, partial $\eta^2 = .021$. Also, no significant main effect was found for the error manipulation on overall valence towards the voice assistant, (*F*(1,116) = 0.02, *p* = .896, partial η^2 = .000. Additionally, no significant interactions effects between interdependence and error were found for valence ratings, (*F*(1,116) = 0.28, *p* = .601, partial η^2 = . Both H_1 and H_3 must be rejected.

Further, no significant main effects were found for interdependence on friendliness ratings, *F*(1, 116) = 1.78, *p* = .182, partial η^2 = .015 and on competence ratings, *F*(1, 116) = 1.81, *p* = .181, partial η^2 = .015. Again, error also did not produce any significant main effects for either friendliness, *F*(1, 116) = 0.13, *p* = .720, partial η^2 = .001 or competence ratings, *F*(1, 116) = 0.05, *p* = .829, partial η^2 = .000. Also, no significant interaction effects were found between interdependence and error for friendliness, *F*(1, 116) = 0.03, *p* = .870, partial η^2 = .000 and competence, *F*(1, 116) = 0.68, *p* = .412, partial η^2 = .006. Thus, H_{1a}, H_{1b} and H_{3a} as well as H_{3b} must also be rejected.

5.3.2.2 Performance of the device

H_2 predicted that smart speakers in the interdependent condition are evaluated significantly better regarding their overall performance compared to devices in the non-interdependent condition while H_4 predicted that smart speakers in the error condition are evaluated significantly worse regarding their overall performance compared to devices in the error-free condition. For performance ratings, ANOVA reveals that the main effect of interdependence on performance ratings was significant, *F*(1,116) = 9.44, *p* = .003, η^2 = .075. The main effect of error on performance ratings was also significant, *F*(1,116) = 54.35, *p* <.001, η^2 = .319. These main effects were qualified by a significant interaction between interdependence and error, *F*(1,116) = 5.02, *p* = .027, η^2 = .041 (see Figure 8). A simple effects analysis was performed to follow up on the interaction effect. The analysis revealed that subjects in the interdependent condition rated an error-prone smart speaker's performance significantly lower (*M* = 4.62, *SD* = 1.06) than the error-free device's performance (*M* = 6.10, *SD* = 0.66), *F*(1, 116) = 46.19, *p* < .001, partial η^2 = .285. Subjects in the non-interdependent condition also rated an error-prone smart speaker's performance significantly lower (*M* = 5.44, *SD* = 0.92) compared to the error-free device (*M* = 6.23, *SD* = 0.66), *F*(1, 116) = 13.17, *p* < .001, partial η^2 = .102. Simple effects also revealed that error-prone smart speaker's performance was rated significantly better by non-interdependent subjects (*M* =

5.44, *SD* = 0.92) compared to interdependent subjects (*M* = 4.62, *SD* = 1.06), $F(1, 116) = 14.11$, $p < .001$, partial $\eta^2 = .108$ while no significant difference in performance ratings was observed in the non-error condition, $F(1, 116) = 0.35$, $p = .558$, partial $\eta^2 = .003$. In conclusion, H_2 must be rejected. H_4 is accepted.

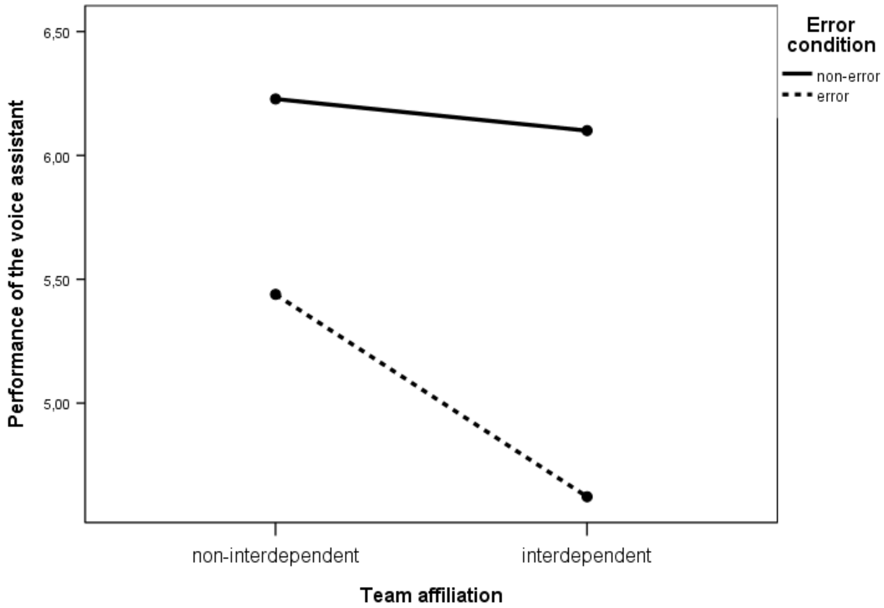


Figure 8. Main and interaction effects of team affiliation and error on participants' ratings of voice assistant performance (source: own figure)

5.3.2.3 Self-performance

In line with predictions, ANOVA revealed a significant main effect of error on self-performance ratings, $F(1, 116) = 20.82$, $p < .001$, $\eta^2 = .152$. According to Cohen (1988) the effect is considered large. Thus, H_5 was accepted. Additionally, ANOVA revealed that interdependence also led to significantly lower evaluations of self-performance, $F(1, 116) = 4.07$, $p = .046$, $\eta^2 = .034$.

5.3.3 Behavioral Measures

5.3.3.1 Prosocial behavior

H_6 stated that the level of interdependence has a positive effect on the participants' prosocial behavior towards the voice assistant. Levene's test indicated that variances differed significantly between groups ($F(3,116) = 4.57$, $p \leq .01$), but as cell sizes were exactly equal and ≥ 30, ANOVA is considered robust against this violation. In line with expectations, the main effect for interdependence was significant, $F(1, 116) = 5.83$, $p = .017$, partial $\eta^2 = .048$. According to Cohen (1988), the effect size is considered medium. Subjects in the interdependent condition named significantly more cities after being asked to do so by the voice assistant than subjects who were told they are being assessed on their individual performance (see Figure 9). Against the assumption made in H_7, there was no significant main effect for recognition errors on prosocial behavior towards a voice assistant, $F(1, 116) = .46$, $p = .498$, partial $\eta^2 = .004$. Participants in the error condition did not show significantly less prosocial behavior compared to participants in the non-error condition and thus H_7 must be rejected. In addition, no significant interaction between interdependence and error was observed, $F(1, 116) = 0.02$, $p = .892$, partial $\eta^2 = .000$.

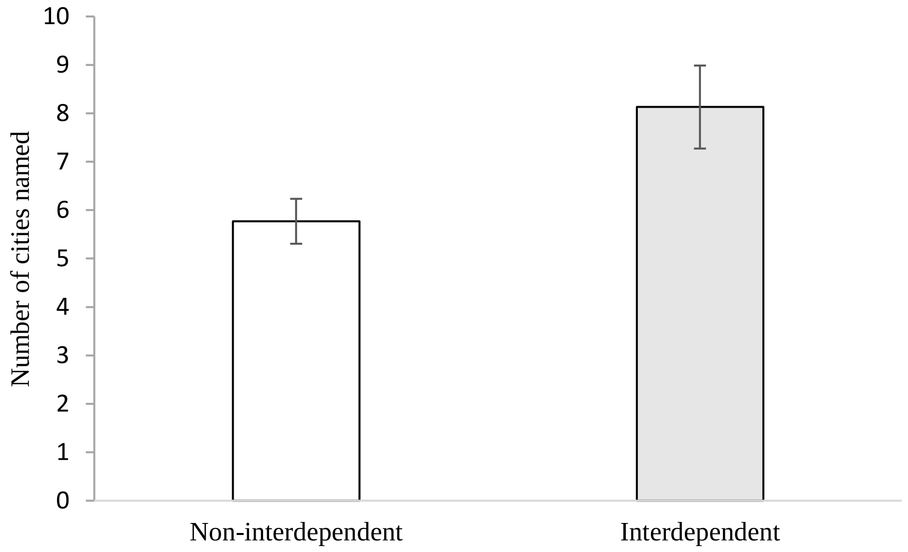


Figure 9. Number of cities named based on team affiliation (source: own figure)
Note. Error bars indicate 95% CI.

5.3.3.2 Interviewer-bias

A three-way mixed ANOVA with two between-subject factors (error, interdependence) and one within-subject factor (device used for evaluation) was conducted to check for an interviewer-bias. Levene's Test indicated unequal variances for data collected on the device itself, *Levene's* $F(3, 116) = 3.68$, $p = .014$. For the measurement on a separate PC, Levene's Test was not significant, *Levene's* $F(3, 116) = 2.27$, $p = .084$. ANOVA revealed no statistically significant three-way interaction between evaluation location, the manipulation of error and the level of interdependence, $F(1, 116) = 2.04$, $p = .156$, partial $\eta^2 = .017$. However, a statistically significant two-way interaction between evaluation location and error was found, $F(1, 116) = 3.94$, $p = .05$, partial $\eta^2 = .033$. Simple main effects revealed both a significant effect for error when the voice assistants performance was assessed on the device itself, $F(1, 116) = 54.35$, $p < .001$, partial $\eta^2 = .319$ and a significant effect for error on performance ratings given on a separate computer, $F(1, 116) = 29.73$, $p < .001$, partial $\eta^2 = .204$. Pairwise comparisons for the significant simple main effects revealed that the voice assistant's performance was rated better in the non-error condition than in the error condition. For ratings given directly to the device, a mean difference of 1.133 (95%-CI[.83, 1.44], $p < .001$) was observed. For performance evaluations on a separate PC, the mean difference was .856 higher in the error-free condition (95%-CI[.55, 1.17], $p < .001$). No other significant two-way interactions were observed. As evaluation location also had no significant main effect on ratings, $F(1, 116) = 0.02$, $p = .968$, partial $\eta^2 = .000$, H_8 must be rejected.

5.3.4 Exploratory Analyses

A series of moderation analyses were conducted to check the influence of individual differences in general self-efficacy and computer self-efficacy on dependent measures. A moderation analysis was run to determine whether the interaction between interdependence and general self-efficacy significantly predicts prosocial behavior towards a voice assistant. The relationship of the examined variables was approximately linear, as assessed by a visual inspection of the scatterplots after LOESS smoothing. The overall model was significant, $F(3, 116) = 3.56$, $p = .016$, predicting 15.73% of the variance. While the analysis did not show that general self-efficacy moderated the effect between interdependence and prosocial behavior significantly, the moderation effect was still marginally significant, $\Delta R^2 = 4.20\%$, $F(1,116) = 3.78$, $p = .054$, indicating that participants that reported higher self-efficacy scores had a tendency to display more prosocial behavior after interdependence had been induced. No other significant moderation effects were found for self-efficacy. Also, no significant moderation effects were found for computer self-efficacy.

5.4 Discussion

This experiment aimed to empirically investigate participants' social reactions towards a voice assistant using both self-reports and behavioral measures based on the manipulation of interdependence as well as deliberately implemented recognition errors leading to a failed interaction. As mentioned in section 3.2.4.2, interdependence was found to elicit numerous forms of prosocial behavior in human-human interaction (e.g., Crawford & Haaland, 1972; Van der Vegt & Van de Vliert, 2005; Wageman, 1995) and also has been shown to induce team affiliation in HCI (Johnson & Gardner, 2007; Nass et al., 1996). The results of this study confirm the positive effect of interdependence on prosocial behavior towards voice assistants. Participants in the interdependence condition provided significantly more help to the voice assistant in an optional task for which no material gain was expected, indicating that participants provided help based only on the previous interaction during which their perception of the voice assistant as a teammate was manipulated. This confirms not only that voice assistants can be perceived as teammates and group members but that this perception can result in observable prosocial behavior towards the device thus treating it as a social actor. It could be argued that – even though the optional task was technically no longer part of the spy game – the overall interaction with the voice assistant was so engaging that participants named cities for their own enjoyment instead of it being an indicator of helpful behavior. However, while overall enjoyment might have been a factor leading to generally higher engagement with the optional task or a factor to engage in the task at all, it does not account for the significant differences between the interdependent and non-interdependent experimental groups.

In line with research indicating that the occurrence of interaction errors is negatively correlated with the users' evaluation of a system (Oulasvirta et al., 2006) and leads to a decrease in prosocial behavior (Velez, 2015), it was expected that deliberately implemented recognition errors leading to participants being unable to successfully complete the game would impact participants behavior towards the device. Further, previous CASA studies confirmed that the perceived helpfulness of a computer influences user behavior. Receiving information of poor quality from a computer resulted in a retaliation effect reducing both positive affect and the quality of subsequent help provided to the computer (Fogg & Nass, 1997a). However, contrary to expectations, there was no significant difference in prosocial behavior between the error and non-error condition. Although Oulasvirta et al. (2006) provide evidence that the occurrence of errors in spoken dialogue systems is negatively correlated with the user's evaluation of the system's helpfulness, the interaction with an error-prone and therefore unhelpful device did not influence the prosocial behavior displayed by participants. Subjects whose game ended in failure due to a deliberate recognition error by the voice assistant did not provide significantly less optional inputs following the interaction. These results could indicate the so-called *Pratfall Effect* that postulates imperfections are regarded as more human and approachable resulting in positive assessments (e.g., Aronson et al., 1966; R. Helmreich et al., 1970). Similar effects have previously been detected in HRI, where studies indicate that users tend to rate faulty robots more positively than flawless ones (e.g., Mirnig et al., 2017; Ragni et al., 2016) and that a robots faulty performance did not influence participants decisions to comply with its requests (Salem et al., 2015). Similar effects have also been mentioned in relation to AI research (Mueller et al., 2022). As imperfections are closely related to the concept of human speech – especially regarding voice recognition errors and flawed speech output – and perceived intelligence is a highly relevant subject for voice assistants, future research should consider examining the effects of imperfections in further detail.

Interestingly and contrary to experiment 1, self-report measures did not produce the same results as the behavioral measure. It was assumed that a voice assistant that is perceived as a team member is also rated better regarding general valence towards the device as well as its friendliness, competence, and performance. No statistically significant difference was found for either valence, friendliness, or competence ratings between the experimental conditions. While this might seem surprising at first when compared to the results obtained in experiment 1, it should be noted that in addition to there being no manipulation of politeness in experiment 2, the role of the smart speaker was also a different one. Instead of the smart speaker acting as an interviewer controlling most of the interaction, it assumed a more passive, supporting role during experiment 2 with a more goal-oriented focus. In line with this role, a significant interaction effect was found for interdependence and error on performance ratings given by participants. Pairwise comparisons for performance ratings indicate a retaliation effect that is stronger after team membership had been induced: a device that produced errors resulting in a failure was devalued significantly in both the interdependent and the non-interdependent experimental groups but the difference in ratings was much greater in the interdependent condition (see Figure 8). Interactions with an interdependent voice assistant that led to failure due to recognition errors produced the lowest performance ratings out of all experimental conditions. Still, this discrepancy between prosocial behavior and subjective self-reports given on a separate device can be interpreted as a social reaction in itself: again, just as in experiment 1, participants might not want to hurt the ‘feelings’ of the voice assistant by refusing to provide direct help on the device itself – hence no difference in prosocial behavior due to error – but still devalued the error-prone smart speaker in honest performance ratings given to a third party. The reason for the significant devaluation of an error-prone voice assistant after team membership had been induced can be explained based on literature: the obligation of members to help their group achieve joint success is closely linked to the feeling of belonging to this group. If this expectation is violated, the group punishes the uncooperative participant (Fehr & Gächter, 2000). There is also another possible explanation for this discrepancy between self-reports and the behavioral measure. The behavioral measure followed the interaction immediately. However, during the evaluation on a separate computer, the interaction between participant and voice assistant – and thus the situation that induced interdependence – was effectively over. It is possible that participants would no longer consider the voice assistant a team member at that point and thus had no reason to rate it more positively based on team affiliation. This would indicate that participants are more forgiving regarding errors while team affiliation is still acute but once enough time has elapsed after the interaction this effect disappears. Still, both explanations indicate social behavior towards the voice assistant based on perceived team affiliation.

For participants self-performance ratings, as predicted ANOVA revealed a significant effect of error. Participants in the error condition reported significantly lower self-performance indicating that they at least in part blame themselves for these errors and that a failed state during interaction with system decreases perceived self-performance (Baumeister & Tice, 1985). Interestingly, ANOVA also revealed that participants in the interdependence condition also reported lower ratings of self-performance than participants in the non-interdependent condition, even though their self-performance would have to be identical to finish the experiment. While there was no significant interaction effect between the conditions, this is still a clear indication that participants based their evaluations of self-performance on the experimental conditions.

Contrary to experiment 1, no support for an interviewer-bias in human-voice assistant interaction could be found for performance ratings. Repeated measurements of participants' perception of the voice assistant's performance, first via an oral evaluation and the second time via an online questionnaire, did not indicate a significant main effect of evaluation location on performance ratings. There is a possible explanation for these results. The scale used to check for an interviewer-bias in experiment 2 was based on the voice assistant's performance and therefore more task-oriented when compared to the assessment of general valence towards the device as in experiment 1. Descriptive measures (see Appendix F) revealed that evaluations of the voice assistant's performance were generally very high for both the written evaluation as well as the oral evaluation which indicates a possible ceiling effect in performance ratings. Based on anecdotal feedback from participants, many of them were impressed with the experimental setup involving the spy game, a voice assistant, and smart bulbs for lighting effects. In addition, many of them had never interacted with smart devices prior to this study. It is conceivable that the elaborated setup influenced participants' ratings in a positive way, thus causing a ceiling effect for ratings effectively eliminating a possible interviewer-bias.

While the exploratory analysis revealed no significant influences of computer self-efficacy on any of the dependent variables, general self-efficacy moderated the effect between interdependence and prosocial behavior with marginal significance, indicating that self-efficacy might be a factor that should be examined further in future CASA research. Participants that reported higher general self-efficacy were more inclined to display prosocial behavior to a voice assistant, as results revealed that the number of cities named increased for subjects with higher self-efficacy in the interdependent condition. Thus, this study offers first indications that general self-efficacy might be a relevant factor for media equation research and more specifically for voice-based systems as they might not have been perceived as 'computers' by participants thus resulting in no effects of computer self-efficacy in this context.

5.4.1 Limitations and Future Research

Compared to study 1, significant improvements were made regarding the software used to collect data. These enhancements resulted in improved interaction between voice assistants and participants regarding ASR and NLP (leading to a significantly reduced number of unintended recognition errors and higher quality of speech interaction). They also allowed for spoken feedback during the behavioral measure by comparing inputs with an internal database to reinforce the sentiment that the device accepts and processes the responses given to it. Still, some technical limitations need to be discussed. There were two deliberately implemented voice recognition errors in the error-condition. One of them resulted in participants having to repeat their request twice with no further consequence while the second one resulted in failure of the game. However, this does not mean that regular voice recognition errors that can happen during any human-voice assistant interaction did not occur additionally. All participants that experienced severe unintended errors were excluded from data analysis (see section 5.2.1) but simple voice recognition errors that resulted in participants having to repeat their request occasionally occurred in all experimental conditions. Still, these errors were much less severe than the deliberately implemented errors and did not result in any kind of failed state for participants. These recognition errors also do not limit external validity, as they also occur during regular everyday usage of smart speakers (e.g., Luger & Sellen, 2016). What complicates the experimental setup of this study compared to traditional CASA setups is the circumstance that participants effectively interacted with two technological devices: the laptop that displayed the game-related text and tasks and the smart speaker that helped participants to fulfill those tasks. Thus, the second device (laptop) could be considered a confounding variable as it represented an additional locus of attention. The laptop itself did not display any social cues and during the evaluation, participants were explicitly asked about their interaction with and evaluations of the voice assistant, but the influence of the second device and potential additional media equation effects cannot be excluded with complete certainty. Consequently, no additional technological devices were used during interactions in experiments 3 and 4 to minimize any potential influence.

Regarding the methodology, for both the original experiment conducted by Nass et al. (1996) and this study, there was no other team or outgroup present during the interaction. While team affiliation still led to increased prosocial behavior towards a voice assistant, it would be interesting for further research to examine positive effects on prosocial behavior in comparison to a clearly identifiable outgroup to compare with or compete against. The presence of an outgroup in HCI produced results based on team membership and ethnicity before (Johnson & Gardner, 2007; Takeuchi et al., 1998; Takeuchi & Katagiri, 1999) thus making it a fruitful approach for future human-voice assistant interaction. Further speaking from a methodological standpoint, prosocial behavior was quantified by the absolute number of cities named by participants after the voice assistant directly asked them for help. In a CASA study following a comparable structure based on desktop PCs, two different computers were used to measure reciprocal behavior displayed by participants (Fogg & Nass, 1997a). One experimental group was asked to provide help on the same computer they interacted with before while the other group was asked by a different computer that had not been involved before. Fogg and Nass (1997a) argue that if participants provide help only towards the same computer that helped them before, it can be attributed to prosocial behavior. If help is instead provided to both devices instead, it could also be interpreted as a result of the computers social characteristics. A similar approach should be considered for voice assistants in future research to explain participants prosocial behavior more precisely. Lastly, as previously mentioned for experiment 1 (see section 4.4.1), the influence of gender stereotypes resulting from the usage of the default German Amazon Alexa voice cannot be ruled out. This consideration is the basis for experiment 3 and is therefore discussed in further detail in section 6.

5.4.2 Implications and Conclusion

The results of this study indicate that users show prosocial behavior towards a smart speaker device based on the social cue of interdependence and the resulting team and group affiliation. Participants who were told that their joint performance with a smart speaker would be evaluated and who were referenced by the device in the way a team member would, provided significantly more input in an unrelated task after being asked to do so by the assistant. Interestingly, this effect persisted even after a failed interaction in the error-condition. Regardless of whether their previous interaction with the device was successful or not, prosocial behavior was shown based on team affiliation, which indicates that users can ‘forgive’ imperfections and recognition errors when interacting with smart speakers which resembles the way misunderstandings in human-human communication are handled (see section 3.4.3.3). While study 1 revealed that users are both polite towards a smart speaker and detect politeness and impoliteness originating from the device, they were not asked to perform any additional tasks following the interaction. Study 2 complements these results by showing that social behavior can extend beyond these norms. Thus, this study provides a further contribution to the extension of the CASA paradigm to voice assistants and with a focus on the medium of speech. In line with most explanations of media equation indicating that social reactions are automatic and unconscious, these results also suggest that behavioral measures are better suited to assess social reactions towards smart speakers in general. The indication that a voice assistant can elicit prosocial behavior based only on the verbal indication of interdependence has implications for both practical design as well as future research. First off, the occurrence of errors (in this case deliberately implemented) during the interaction with a voice assistant neither prevented social reactions in the form of prosocial behavior from happening as theorized by Nass and Moon (2000) nor did they decrease it compared to an error free interaction as expected based on previous research (Velez, 2015). It did, however, influence performance ratings given by participants. An error-prone, interdependent voice assistant received the lowest performance ratings out of all experimental conditions, but only in ratings given on a separate PC. Ratings given to the device directly did not differ significantly based on the occurrence of errors. These results suggest that social reactions (such as the adherence to an interviewer-bias when questioned directly) happen regardless of errors and only when the device is evaluated consciously after the fact do they influence evaluations. For practical design, this means that how users interact with a faulty device and how they evaluate it needs to be considered separately. The location where these errors are assessed is also important. A device asking for feedback directly might not result in accurate answers regarding its performance during tasks. This is especially relevant for voice assistants that periodically ask their users for feedback after using certain features or Skills as people might in fact be more forgiving in a direct, seemingly social interaction compared to a formal evaluation on a separate device. As for interdependence and team affiliation, they are surprisingly easy to induce using minimal social cues given by the device itself. This could be used to influence or even manipulate users in certain contexts as it directly affects their behavior, which results in new questions regarding the concepts of trust, influence and dependency between users and smart speaker devices. These moral and ethical implications are discussed in more detail in section 8.

6 Experiment 3: Effects of Gender Stereotypes and Conformity in Human-Smart Speaker Interaction

6.1 Study Outline and Hypotheses

As outlined earlier (see section 3.1), CASA literature indicates that users tend to perceive a computer as a social entity which leads to social behavior in interactions with computers. This also includes the adoption of attributions and preconceived notions about certain attributes that are usually only ascribed to other humans such as sex or gender and the resulting stereotypes. Stereotyping based on sex and gender is well established in social psychology literature (see section 3.2.5). This is highly relevant for voice assistants as inferences about other people's attributes have been shown to be made based on their voice alone (Scherer et al., 2001). Previous research revealed that a higher knowledge concerning the topics of love and relationships was ascribed to computers speaking with a female voice while higher knowledge of technology was ascribed to male computers (Nass et al., 1997). More recently, male avatars were rated to be more competent compared to female avatars who were rated to be warmer (E.-J. Lee, 2008). Additionally, similar effects could also be shown with different cues such as colored smartphone cases. Smartphones with a pink case were rated to be warmer while smartphones with a blue case were rated to be more competent (Carolus, Schmidt, Muench, et al., 2018). Comparable effects were also shown for robots (Mieczkowski et al., 2019). Based on these results, it is reasonable to assume that a personal assistant's voice will act as a social cue during the interaction with the device thus triggering the same gender stereotypes. Additionally, the effects of social identification processes (see Abrams & Hogg, 2006) and conformity based on gender have been shown in HCI (E. Lee et al., 2000; Nass & Brave, 2005). To evaluate both the application of gender stereotypes to smart speakers as well as the behavioral component of social identification effects, this study extended the methodological approach of E. Lee et al. (2000) to smart speakers. Participants were asked to interact with a device that used either a synthesized male or a synthesized female voice to present them with a series of social dilemmas. For each dilemma, the smart speaker also offered a recommendation of which option to choose. Participants conformity with these recommendations was considered as a behavioral measure. Following the interaction, they were asked to rate the smart speaker based on dimensions traditionally employed in gender stereotype research to assess if gender stereotypes are applied to smart speakers based on their voice.

Disclaimer. Based on the research tradition this study follows, the focus will be exclusively on stereotypes regarding males and females known from human-human interaction. Any mentions of gender in this section refer exclusively to these two genders. As voice assistants in this experiment never mentioned their gender, technically speaking the manipulation only evoked the attribution of sex via voice though it is assumed that this results in the ascription of gender stereotypes as they are based on perceived sex in human-human interaction (see Eagly & Wood, 1982).

Warmth and competence have been part of gender stereotype research for years (Conway et al., 1996; Eagly, 1987) and are considered two of the most important dimensions for social cognition (Fiske et al., 2007). Men are traditionally ascribed more attributes related to competence while women are ascribed more attributes related to warmth. A similar effect is predicted for smart speakers that use either a male or a female voice to communicate.

H_{1a}: A smart speaker device using a male voice is rated to be more competent compared to a smart speaker using a female voice.

H_{1b}: A smart speaker device using a female voice is rated to be warmer compared to a smart speaker using a male voice.

Additional attributes are ascribed to people based on their sex: women are attributed expressiveness while men are attributed instrumentality (Broverman et al., 1994; Runge et al., 1981; Spence & Helmreich, 1980). Thus, it is again assumed the same attributes are ascribed to a smart speaker based on its voice.

H_{2a}: A smart speaker device using a male voice is attributed a higher instrumentality compared to a smart speaker using a female voice.

H_{2b}: A smart speaker device using a female voice is attributed a higher expressiveness compared to a smart speaker using a male voice.

Individuals are more likely to be identify with other members of their own ingroups (e.g., gender) and this positively influences both their attributions and behavior towards these members (Abrams & Hogg, 2006). This has previously been transferred from human-human interaction to HCI (E. Lee et al., 2000; K. M. Lee & Nass, 2003; Morishima et al., 2002). Consequently, it is assumed that the subjective evaluation of a voice assistants' overall attractiveness from a user experience standpoint differs between male and female participants based on its voice due to this identification process.

> H_{3a}: A smart speaker device using a male voice is rated to be more attractive by male participants.
>
> H_{3b}: A smart speaker device using a female voice is rated to be more attractive by female participants.

As described in section 3.2.5.4, conformity can be tied to the own gender as well as the gender of interaction partners. It has been shown in literature that male persuaders generally elicit more conformity than female persuaders (Eagly, 1978, 1983; Lockheed, 1985). Thus, it is assumed that this effect extends to a male voice assistant.

> H_{4a}: A smart speaker device using a male voice to give recommendations will generally result in more conformal behavior compared to a smart speaker using a female voice.

Additionally, social identification also influences behavior (Abrams & Hogg, 2006). Previous CASA research revealed that participants displayed conformal behavior based on the perceived gender of a computer they interacted with. Male participants followed recommendations given by a male computer and female participants yielded to recommendations given by a female computer (E. Lee et al., 2000; Nass & Brave, 2005). Similar findings have been reported in human-smartphone interaction (Carolus, Schmidt, Muench, et al., 2018). Based on these results, similar effects should also be observed during the interaction with a voice assistant device.

> H_{4b}: Male participants will conform more to a male smart speakers' recommendations while female participants will conform more to female smart speakers' recommendations.

Exploratory considerations. The individual difference measured in this study was participants individual tendency towards anthropocentrism. This tendency assesses the extent to which one believes a non-human entity possesses human characteristics and should be regarded as such (Waytz et al., 2010). It is assumed that this tendency towards anthropocentrism might influence media equation effects as it could theoretically facilitate both the inclination and aversion to treat technological devices as social actors.

RQ: Does participant's tendency towards anthropocentrism moderate their social reactions towards a voice assistant?

6.2 Pretest

Multiple male, female, and gender-neutral voices were tested prior to the main experiment. Two of those, a male and a female voice were chosen, and N = 152 participants completed a pretest questionnaire rating these voices. Of these participants, n = 79 (52%) were female and n = 73 (48%) were male. Participant's age ranged from 18 to 56 years (M = 22.57, SD = 4.50). The questionnaire consisted of short audio clips of the voices repeating two sentences three times. The sentences consisted entirely of made-up words ("Hat sundig pron you venzy" and "fee gott laish jonkill gosterr") based on a study conducted by Scherer et al. (2001). This procedure was chosen to avoid any influence of the spoken content on ratings. Participants were then asked to rate the voices on a shortened valence scale adopted from Nass et al. (1999) containing seven items. All answers were given on a 7-point Likert scale to identify one male and one female voice that did not differ regarding their assessment to exclude inherent differences in the voices as a confounding variable. ANOVA revealed that the two voices chosen for this study did not differ significantly regarding their perceived friendliness (F(1, 150) = 0.77, p = .383), competence (F(1, 150) = 2.44, p = .120), sympathy (F(1, 150) = 1.39, p = .240), politeness (F(1, 150) = 1.34, p = .250), warmth (F(1, 150) = 1.30, p = .719) and how analytical (F(1, 150) = 0.67, p = .413) and pleasant they sounded (F(1, 150) = 0.58, p = .447).

6.3 Methods

6.3.1 Participants

This study was conducted in two waves. During the first wave, N_1 = 75 students participated in the initial experiment with another N_2 = 66 students participating in a second wave. In both waves, participants were mostly recruited using the internal recruitment system of the Institute Human-Computer-Media in exchange for course credit. After excluding any cases containing technical errors in both waves (n_1 = 13 and n_2 = 7), this resulted in final sample of N_{total} = 121 participants in this study, with 53 (43.8%) being male and 68 (56.2%) being female. Participant's age ranged from 19 to 46 years (M = 23.03, SD = 6.16) with 105 of them (86.8%) being 19 to 24 years old. In terms of education, 107 (88.4%) reported holding a higher education entrance qualification. 8 (6.6%) participants held a university degree, 3 (2.5%) reported a completed apprenticeship and 3 (2.5%) participants held a secondary school certificate. Written informed consent was obtained from each participant before the study.

6.3.2 Stimulus Material

6.3.2.1 Hardware and Software

Participants were asked to interact with an original *Google Home* smart speaker device[4] using either a male or a female voice. This device was selected due to technical limitations regarding Amazon Echo at the time this study was originally conducted. During that time, there was no official German voice equivalent to the female voice of Alexa and a custom male voice would have differed too much regarding different factors such as pitch, frequency, or range thus making comparisons difficult. Google Home on the other hand allowed for the use of multiple equivalent male and female voices. To customize the interaction between the voice assistant and the participants, the design tool *Voiceflow* (https://voiceflow.com) was used. Voiceflow allows users to program Google Home to use custom prompts and give custom answers. For this experiment, several interaction trees (see Appendix J) were created to allow a Google Home device to present participants with five social dilemmas while using either a male or a female voice. On a software level, Voiceflow and the device itself were connected via the *Google Developer Console* which also allowed access to the log files of the voice interaction for data analysis of conformity effects.

[4] The device has since been officially renamed to Google Nest. For the purposes of this dissertation, it will still be referred to as Google Home as that was the official name at the time the study was conducted.

6.3.2.2 Voice Interaction

The general interaction between the voice assistant and participants from a technical standpoint was similar to experiment 1 (for a detailed description of the general interaction see section 4.2.2) albeit using a Google Home smart speaker device instead of an Amazon Echo Plus device. Implications for the general interaction are minimal, as both devices provide the same basic functions. The only relevant difference pertains to the contextual LEDs of the device: instead of a blue LED ring, the device displayed a moving pattern of four colored LED lights to indicate that it is listening for an input and pulsating lights when it provided voice output. For the sake of consistency, all participants were told that they would interact with a Google Home device.

6.3.2.3 Social Dilemmas

Participants were presented with a series of five descriptions of situations where they had to choose between two equivalent courses of action. The dilemmas used for this experiment were adapted from a study on conformity in HCI (E.-J. Lee & Nass, 2002) and translated into German. They were written and evaluated in the original study so that they would present participants with two options of equal value to not have an obvious or correct choice on which answer to pick. All dilemmas (see Appendix H) were presented to participants verbally by the voice assistant and via a complementary printout that allowed them to read along. One of the dilemmas presented to participants is the following (note that all dilemmas were presented in German):

> Ms. E, a college senior, has studied the piano since childhood. She has won amateur prizes and given small recitals, suggesting that she has considerable musical talent. As graduation approaches, she has the choice of taking a medical school scholarship to become a physician, a profession which would bring certain financial rewards, or entering a conservatory of music for advanced training with a well-known pianist. She realizes that even upon completion of her piano studies, success as a concert pianist would not be assured.

This allowed for the voice to influence participants to choose either option. All five dilemmas followed the same structure: first, the voice assistant introduced the situation and the relevant actors of each scenario. This was followed by two possible ways to resolve the situation. Again, these were presented to participants verbally and via printout. For the dilemma above, these were the following:

> A: Ms. E should enter a conservatory of music for advanced training
> B: Ms. E should choose to take a medical school scholarship to become a physician

The last step for each of the five dilemmas was a clear suggestion of which one of the two possible solutions to choose. The recommendation was presented to participants exclusively via voice output by the smart speaker device to reinforce the idea that it is the source of the recommendation. This was essential to ensure source orientation towards the voice assistant for any potential conformity effects (Nass & Steuer, 1993; Sundar & Nass, 2000). The entire voice interaction protocol can be found in Appendix J.

6.3.2.4 Voices

As with all other parts of the interaction, the recommendations on how to act in each of the five dilemmas were made by one of two voices originating from the Google Home smart speaker device. The voices used for this purpose were called "Male 1 (DE-DE)" for the 'male' voice and "Female 2 (DE-DE)" for the 'female' voice and chosen due to the results of the pretest (see section 6.2). The main difference between these voices was the fundamental frequency. The female voice had a higher pitch (at around the average frequency of human females of 210Hz) while the male voice had a lower pitch (again around the average human male frequency of 110Hz). Other factors such as volume or frequency range were kept the same to minimize confounding variables. Aside from the different voices, no indications of gender were given to participants. The voice interaction itself was identical in both conditions.

6.3.3 Self-Report Measures

6.3.3.1 Stereotype Content Model

The perceived warmth and competence of the assistants' voices were measured using the stereotype content model (SCM) (Fiske et al., 2018). Both subscales consist of 6 items (e.g., competence: competent, confident, independent; warmth: tolerant, warm, sincere) that were translated to German for this study. Answers were given on a 5-point Likert-scale (1= *agree fully*; 5= *do not agree at all*). Reliability for high for both the subscale of warmth (α = .87) as well as for the subscale of competence (α = .82).

6.3.3.2 German Extended Personal Attributes Questionnaire

Perceived expressiveness and instrumentality of the assistants' voices were measured using the German Extended Personal Attributes Questionnaire (GEPAQ) developed by Runge et al. (1981) as a German version of the Extended Personal Attributes Questionnaire (EPAQ) (R. L. Helmreich et al., 1981) which is in turn based on the Personal Attributes Questionnaire developed by Spence at al. (1975). Both subscales consist of 8 differential items each (e.g., instrumentality: not independent – independent, not confident – confident; expressiveness: not gentle – gentle, not sympathetic – sympathetic), and answers were given on a 5-point Likert-scale. Reliability was high (α = .81) for expressiveness and acceptable (α = .70) for the subscale of instrumentality.

6.3.3.3 User Experience Questionnaire

To measure the general experience of participants' interaction with the voice assistant, the User Experience Questionnaire (UEQ) was employed (Laugwitz et al., 2008). It consists of six subscales (attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty) and contains 26 semantic differential items (e.g., good – bad; attractive – unattractive) in total that are measured on 7-point bipolar Likert-scales. For this study, only the subscale of attractiveness was considered for further analysis. This subscale consisted of 6 items and displayed high reliability of α = .89.

6.3.3.4 Individual Differences in Anthropomorphism Questionnaire

Participants' tendency towards anthropocentrism was measured using the Individual Differences in Anthropomorphism Questionnaire (IDAQ) by Waytz et al. (2010). This measure includes the IDAQ score (that assesses anthropomorphism) and the IDAQ-NA (that assesses non-anthropomorphic attribution). For this study, only the subscale IDAQ score containing 15 items (e.g., "to what extent does the average computer have a mind of its own?") that assess the extent to which participants believe a non-human entity possesses human characteristics was used. Answers were given on a 10-point Likert-scale (1= *not at all*; 10= *very much*). Reliability for the IDAQ score was high, α = .84.

6.3.3.5 Manipulation check

To control that the manipulation of the smart speaker's perceived sex via voice was successful, at the end of the online questionnaire, participants were asked what sex they ascribed to the voice they heard previously. The question was open-ended so that participants were not influenced by predefined answers. The manipulation was successful, as results confirmed that 120 of 121 participants correctly identified the female voice as female and the male voice as male. One participant did not ascribe any sex to the voice.

Table 5. Overview of scales and their internal consistencies for the variables in Experiment 3 (N = 121)

Scale	Subscale	Cronbach's α	Items
Stereotype Content Model	Warmth	.87	6
	Competence	.82	6
GEPAQ	Expressiveness	.81	8
	Instrumentality	.70	8
IDAQ	IDAQ score	.89	15
User Experience Questionnaire	Attractiveness	.89	6

6.3.4 Behavioral Measure: Conformity

Conformity was measured via voice input on the device itself. Following each social dilemma, participants had to verbally agree or disagree with the device's recommendation on which of the options to choose. Their agreement with the voice assistant's recommendation was measured on a de facto 7-point Likert scale by stating a number ranging from 1 (*do not agree at all*) to 7 (*fully agree*). These values were recognized by the device which then gave verbal confirmation. Additionally, the values were saved in an interaction log for each participant from which they were later extracted for data analysis. It was assumed that based on the perceived gender of the voice assistant, subjects would differ in their conformal behavior shown directly to the device. Thus, the mean value of agreement expressed verbally by each participant was considered as a behavioral measure for further analyses.

6.3.5 Procedure

Participants were randomly assigned to one of the two experimental conditions (male/female voice) while ensuring an equal participant gender ratio in both conditions for the sake of comparability and equal cell sizes. After a short greeting by the researcher, participants were instructed to interact with the Google Home device, which guided them through the procedure that was split into three distinct parts. (1) warm-up, (2) the voice interaction containing social dilemmas, and (3) evaluation of the device via an online questionnaire.

During the warm-up (1), participants learned about the basic functions of a Google Home smart speaker and were told to ask a few basic questions to familiarize themselves with the device. This step was identical to experiments 1 and 2 with the only difference being the smart speaker device (Google Home instead of Amazon Echo Plus). This was followed by the main interaction (2), which had the voice assistant present participants with a cover story about collecting data on how people decide when faced with certain dilemmas that allow for more than one solution. Participants were then presented with an introductory dilemma that served as an explanation for the process. Following the introduction, participants were presented with the five main social dilemmas. In addition, the voice assistant also provided two possible solutions for each dilemma as well as a clear recommendation on which of those two options to choose. Participants were then asked to verbally agree or disagree with the voice assistant's recommendation to measure conformity. In addition, they were also asked to explain their decision in an open-ended question. Due to technical limitations of smart speakers at the time, the answers to these open-ended questions were not processed by the device and were merely aimed to increase the immersion for participants and to increase interactivity. After the conclusion of the main voice interaction, participants were instructed by the device to take a seat at a separate desktop computer where the experiment concluded with an online questionnaire (3) that contained the scales named in section 6.3.3 to evaluate the voice assistant as well as a query of demographic information.

6.4 Results

6.4.1 Design and Statistical Analyses

This study followed a 2 (voice assistant gender: male vs. female) x 2 (participant gender: male vs. female) design. A series of two-way between-subjects ANOVAs were employed with two between-subject-factors (voice assistant gender: male vs. female and participant gender: male vs. female). As the underlying statistical procedures are identical to procedures previously employed in experiment 1 and experiment 2, for reasons of brevity and to avoid redundancy, a detailed description of statistical analyses and test assumptions can be found in section 4.3.1. All test assumptions were met unless stated otherwise. Descriptive statistics are noted in Appendix L.

6.4.2 Self-Report Measures

6.4.2.1 Warmth and competence

H_{1a} stated that a male voice assistant is rated to be more competent compared to a female voice assistant. Crucially, a two-way ANOVA between the experimental groups revealed that the voice assistants gender had no significant main effect on competence ratings, $F(1, 117) = 0.11$, $p = .739$, partial $\eta^2 = .001$. A male voice assistant was not rated to be significantly more competent ($M = 4.03$, $SD = 0.73$) than a female voice assistant ($M = 3.97$, $SD = 0.76$). Instead, a statistically significant interaction effect was found between voice assistant gender and participant gender, $F(1, 117) = 6.43$, $p = .013$, partial $\eta^2 = .052$ (see Figure 10). To further examine the significant interaction, simple effects analyses were employed (Field, 2013). Simple main effects revealed that female voice assistants were rated to be significantly more competent by female participants ($M = 3.89$, $SD = 0.73$) compared to male participants ($M = 3.46$, $SD = 0.61$), $F(1, 117) = 5.61$, $p = .019$, partial $\eta^2 = .046$. No such effect was found for competence of male voice assistants, $F(1, 117) = 1.45$, $p = .232$, partial $\eta^2 = .012$. Simple effects also revealed that there were no significant differences in competence ratings given to male compared to female voice assistants by either female participants, $F(1, 117) = 2.76$, $p = .100$, partial $\eta^2 = .023$ or male participants, $F(1, 117) = 3.68$, $p = .058$, partial $\eta^2 = .030$. Still, as there was no significant main effect, H_{1a} had to be rejected.

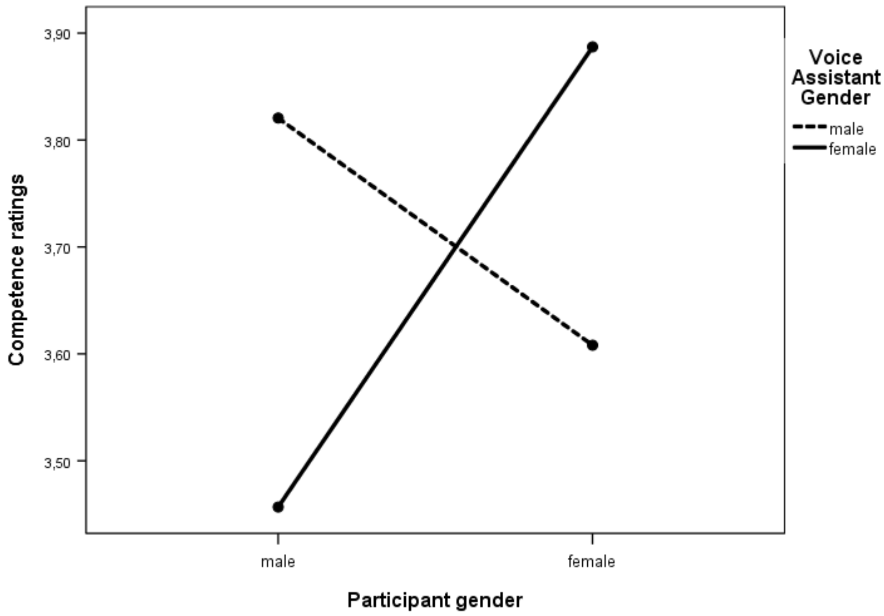


Figure 10. Interaction effect between voice assistant gender and participant gender on voice assistant competence ratings (source: own figure)

H_{1b} focused on the dimension of warmth, where a female voice assistant was expected to be rated higher compared to a male voice assistant. There was a significant difference between the groups, $F(1, 117) = 7.08$, $p = .009$, partial $\eta^2 = .057$. However, contrary to H_{1b}, subjects perceived a male voice assistant to be significantly warmer ($M = 3.65$, $SD = 0.68$) compared to a female voice assistant ($M = 3.34$, $SD = 0.68$). No significant interaction effects with participant gender were found for warmth ratings, $F(1, 117) = 2.094$, $p = .151$, partial $\eta^2 = .018$. H_{1b} must be rejected.

6.4.2.2 Expressiveness and instrumentality

No significant main effects were found for voice assistant gender on participants evaluations of instrumentality, $F(1,117) = 0.08$, $p = .776$, partial $\eta^2 = .001$. Also, no significant main effects were found for voice assistant gender on participants evaluations of expressiveness, $F(1,117) = 2.27$, $p = .135$, partial $\eta^2 = .019$. Additionally, no significant interactions effects with participant gender were found for either instrumentality, $F(1,117) = 3.65$, $p = .059$, partial $\eta^2 = .030$, or expressiveness, $F(1,117) = 1.30$, $p = .256$, partial $\eta^2 = .011$. Both H_{2a} and H_{2b} must be rejected.

6.4.2.3 Attractiveness

There was no statistically significant main effect for voice assistant gender on attractiveness ratings, $F(1,117) = 2.12$, $p = .148$, partial $\eta^2 = .018$. There was also no significant main effect for participant gender on attractiveness ratings, $F(1,117) = 1.17$, $p = .282$, partial $\eta^2 = .010$. Still, ANOVA revealed a statistically significant two-way interaction between participant gender and voice assistant gender for attractiveness ratings, $F(1,117) = 4.26$, $p = .041$, partial $\eta^2 = .035$ (see Figure 11). Simple main effect analysis revealed that female participants rated a female voice assistants as significantly more attractive ($M = 5.03$, $SD = 1.16$) than male participants ($M = 4.44$, $SD = 1.14$), $F(1, 117) = 4.81$, $p = .030$, partial $\eta^2 = .039$. No significant difference in ratings was found for male voice assistants, $F(1, 117) = .497$, $p = .482$, partial $\eta^2 = .004$. Simple main effects also revealed that male participants rated a male voice assistant as significantly more attractive ($M = 5.10$, $SD = 0.91$) than a female voice assistant ($M = 4.44$, $SD = 1.14$), $F(1, 117) = 5.53$, $p = .020$, partial $\eta^2 = .045$ while female participants did not significantly differ in their attractiveness ratings for male and female voice assistants, $F(1, 117) = 0.21$, $p = .648$, partial $\eta^2 = .002$. Thus, H_{3a} was rejected while H_{3b} was accepted.

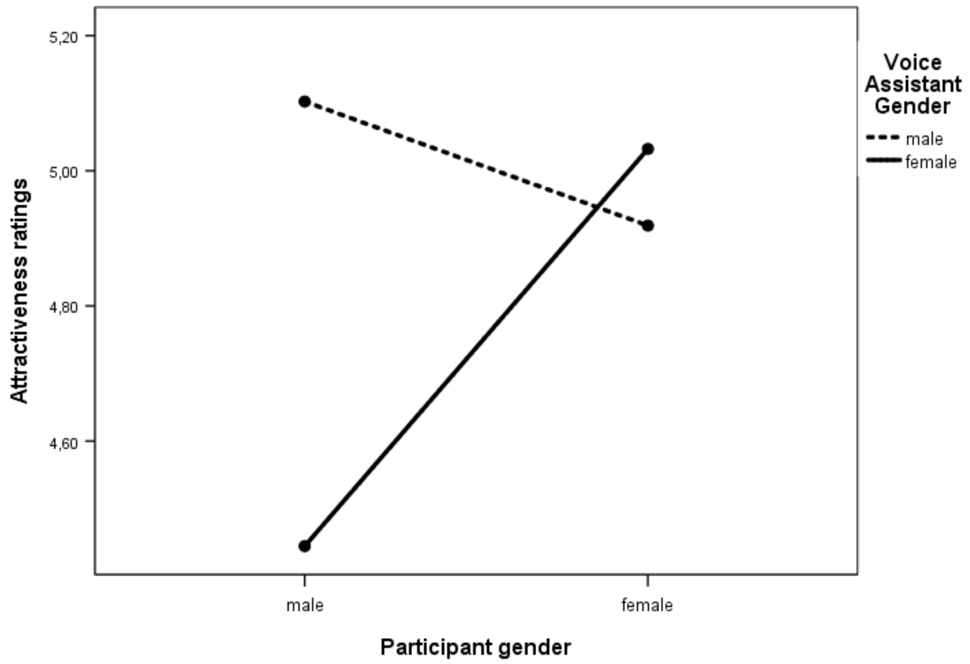


Figure 11. Interaction effect between voice assistant gender and participant gender on voice assistant attractiveness ratings (source: own figure)

6.4.3 Behavioral Measures

No significant main effects were found for voice assistant gender on conformity, $F(1, 117) = .756$, $p = .386$, partial $\eta^2 = .006$ or for participant gender on conformity, $F(1, 117) = 2.18$, $p = .143$, partial $\eta^2 = .018$. However, ANOVA revealed a statistically significant two-way interaction between participant gender and voice assistant gender for conformal behavior, $F(1, 117) = 12.85$, $p < .001$, partial $\eta^2 = .099$ (see Figure 12). Simple effect analyses were conducted to examine the significant interaction. Results showed that male participants yielded to recommendations given by male voice assistants ($M = 4.19$, $SD = 0.80$) significantly more often comparted to recommendations given by female voice assistants ($M = 3.62$, $SD = 0.82$), $F(1, 117) = 8.89$, $p = .004$, partial $\eta^2 = .070$. Additionally, female participants showed significantly more conformal behavior with recommendations given by female voice assistants ($M = 4.27$, $SD = 0.55$) compared to recommendations given by male voice assistants ($M = 3.92$, $SD = 0.66$), $F(1, 117) = 4.19$, $p = .043$, partial $\eta^2 = .035$. Analyses also revealed that for recommendation given by a female voice assistant, there was a significant difference in conformal behavior between female ($M = 4.27$, $SD = 0.55$) and male ($M = 3.62$, $SD = 0.82$) participants, $F(1, 117) = 12.457$, $p = .001$, partial $\eta^2 = .096$ while no such effect could be found for male voice assistants, $F(1, 117) = 2.29$, $p = .133$, partial $\eta^2 = .019$. H_{4a} must be rejected. H_{4b} was accepted.

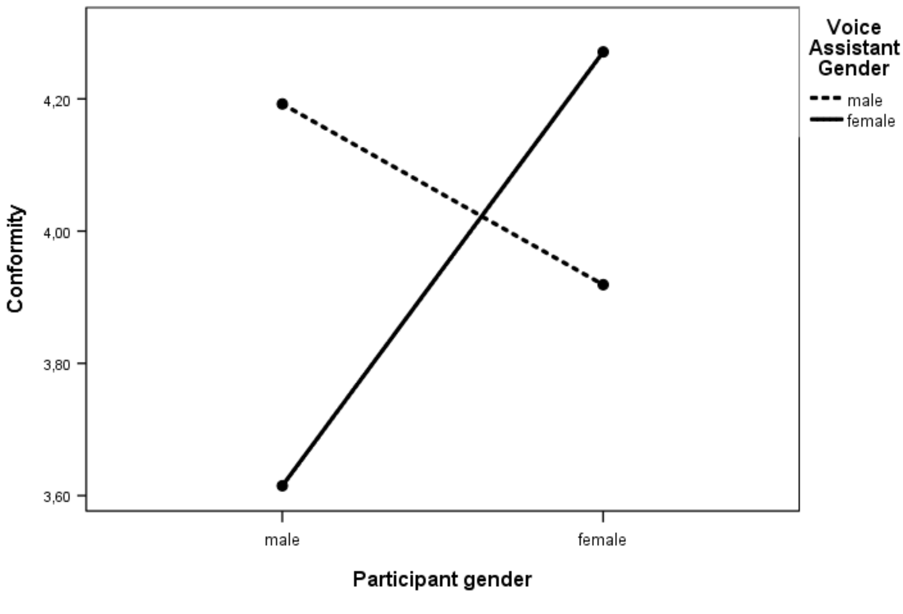


Figure 12. Interaction effect between voice assistant gender and participant gender on conformal behavior (source: own figure)

6.4.4 Exploratory Analyses

A moderation analysis was run to determine whether the interaction between participant gender and individual differences in anthropomorphism (via the IDAQ score) significantly predicts conformal behavior towards a voice assistant. The overall model was significant, $F(3, 117) = 3.18$, $p = .027$, predicting 5.19% of the variance. However, IDAQ score did not moderate the effect between participant gender and conformal behavior significantly, $\Delta R^2 = 0.73\%$, $F(1,117) = 0.86$, $p = .354$. A second moderation analysis was run to determine whether the interaction between voice assistant gender and IDAQ score significantly predicts conformal behavior. The overall model was significant, $F(3, 117) = 3.73$, $p = .013$, predicting 4.98% of the variance. Again, IDAQ score did not moderate the effect significantly, $\Delta R^2 = 1.23\%$, $F(1,117) = 1.49$, $p = .224$. Also, no significant moderation effects of IDAQ score were found for any of the self-report variables.

6.5 Discussion

Gender stereotypes have been found to affect almost every area of our everyday life (Costrich et al., 1975) and can lead to conformal behavior based on gender (Eagly, 1983). The goal of this study was to examine the effects of gender stereotypes and the resulting conformity during human-voice-assistant interaction based on previous CASA research. The results confirmed the significant effect of perceived gender on both the subjective assessment of smart speakers and on objective behavioral measures in the form of conformity with recommendations given by the device. Starting with self-reports, data analysis revealed multiple statistically significant results regarding the dimensions of gender stereotypes ascribed to smart speakers based on their voice. A statistically significant interaction effect was found between voice assistant gender and participant gender indicating that female voice assistants were rated to be significantly more competent by female participants but not by male participants. No significant difference in competence assessment ratings of male voice assistants was found. Regardless, this interaction effect indicates a social identification process congruent with human-human interaction (Eagly, 1983) and HCI research (E. Lee et al., 2000): female participants show a preference for their own ingroup based on sex and thus ascribe more competence to a voice assistant that seemingly belongs to that group. The significant interaction effect for attractiveness ratings also indicates a similar social identification effect: female participants rated a female voice assistant as significantly more attractive while male participants rated a male voice assistant as significantly more attractive than a female voice assistant again showing a preference for the own ingroup. However, no significant effects of voice assistant gender on evaluations of its instrumentality and expressiveness were detected. There are two possible explanations for the unexpected results regarding evaluations on these dimensions. Firstly, due to the pretest conducted before the study, two voices were deliberately chosen to not differ regarding their attributes such as perceived friendliness, competence, sympathy and how pleasant they sounded (see section 6.2) to avoid preconceived notions about these voices as confounding variables. This might in turn have led to an erosion of sex-related differences in the voices chosen for the main study and thus skewed perceptions of the voice assistants towards more neutral ratings. A second possible explanation might lie in the general development of explicit gender stereotypes over the last decades. While gender stereotypes are considered to be relatively stable in traditional gender stereotype literature and many of these stereotypes persist even today (for a comparison, see Haines et al., 2016), changes in the perception of these attributes have been documented in literature. This is especially relevant for biases towards the perception of women, who have been ascribed more instrumental attributes over time (Twenge, 1997). A recent study using machine learning to analyze the occurrence of gender stereotypes over the course of the 20th century in large amounts of historic natural language data also revealed vital changes: While generally stereotypes and biases are still robust, their strength has eroded over time which is again due to changes in traits considered to be stereotypically female in literature such as warmth or expressiveness (Bhatia & Bhatia, 2021). Another meta-analysis also revealed a trend towards neutrality in explicit gender stereotypes over the last 10 years (Charlesworth & Banaji, 2022). These trends could also provide an explanation for lack of a suspected main effect of higher competence ascriptions to male voice assistants. A recent meta-analysis of U.S. public opinion polls over the last seven decades revealed considerable changes regarding stereotypical traits, such as an increasing belief in competence equality for men and women (Eagly et al., 2020).

Contrary to expectations based on traditional gender stereotypes, subjects perceived a male voice assistant to be significantly warmer than a female voice assistant. This seems unexpected as literature indicates that while stereotypes towards women that have eroded or changed towards equality (e.g., for competence ascriptions), stereotypes towards men have stayed mostly consistent. While this result might be surprising at first, it does align with previous findings in similar CASA research. Participants in multiple studies reported that synthesized male voices were more pleasant to listen to than synthesized female voices (Johnson & Gardner, 2009; E. J. Lee, 2003; Mullennix et al., 2003). Mullenix et al. (2003) suggest that the higher fundamental frequency and more diffuse formant structures of female voices make it harder to generate a high-quality synthetic female voice compared to a male voice. Additionally, the technical quality of the voice output might have been a factor. While the Google Home device used in this study has integrated loudspeakers sufficient for regular voice playback, highs and lows of different voices might still result in varying quality levels of audio output thus potentially leading to lower-pitched, male voices being perceived as more comfortable to listen to and therefore warmer.

As for the behavior measure, conformity with recommendations given by a smart speaker using either a male or a female voice was measured via a verbal query directly on the device. Results once again revealed a significant interaction effect between participant gender and voice assistant gender: female participants displayed significantly more conformal behavior with recommendations given by female voice assistants while male participants did the same with male voice assistants. Additionally, there was a significant difference in conformal behavior between female and male participants for recommendations given by a female voice assistant, but no such difference was found for a male voice assistant. This cross-gender effect of social identification and influence can also be found in human-human interaction (Abrams & Hogg, 2006) as well as previous CASA research (E. Lee et al., 2000; Nass & Brave, 2005). While results for self-reports were mixed at best, these crossover interactions represent the most compelling evidence of gender effects in human-voice assistant interaction as they indicate that users identify with the perceived sex of a smart speaker and are influenced by its recommendations accordingly. It could be argued that the stimulative nature of a smart speaker strongly arguing for one of two possible options during the interaction influenced participants to generally agree with its recommendations, especially in a laboratory context where social desirability is a relevant factor. However, this explanation can be ruled out as it would only account for a main effect of conformity, not for the interaction effect between voice gender and participant gender observed in this experiment.

One important factor to consider regarding the general perception of a voice assistants' gender are participants' expectations towards it. Most of today's voice assistants' default to a female-sounding voice unless actively changed by the user. This circumstance may lead to most users and participants expecting a female voice from a voice assistant. Therefore, it is possible that the male voice could have surprised these participants, as they are more familiar with female voices originating from voice assistants. The effects of this might be twofold: the unexpected voice could have caused participants to reflect on it which would in turn contradict any automatic and unconscious media equation effects and the familiarity with female voices could have influenced participants' ratings. While the effects of familiarity cannot be ruled out completely for subjective self-reports, the statistically significant interaction effect of the behavioral measure indicates that it did not influence social reactions as familiarity with a certain voice or the expectation of a female voice would once again only predict a main effect. The fact that male participants conformed to male voices while female participants conformed to female voices clearly indicates an identification process.

Exploratory analyses of individual differences revealed no indication that a general tendency towards anthropomorphism influences either self-report assessments or conformal behavior towards a voice assistant. Results indicate no significant moderation effects on conformal behavior based on either participant gender or the sex of the voice used by the smart speaker device. It should be noted that the query of anthropocentrism followed both the interaction with the device and its general assessment and thus a confounding effect of the previous interaction on anthropomorphism reports cannot be ruled out. It is also noteworthy that there is a limited number of instruments used to assess general tendencies toward anthropomorphism and the instrument used in this study (IDAQ) is rather new and thus has been evaluated in a limited number on contexts (see Waytz et al., 2010).

In sum, considering the young and highly educated sample, it seems plausible that the aforementioned general trend towards the erosion of gender stereotypes in society is reflected in self-reports given in this study. It also must be considered that self-reports are only able to assess explicit gender stereotypes, as self-reports allow participants to access their cognitions and thus are liable to social desirability bias. This is especially relevant in a context where participants are asked to evaluate their own biases such as gender stereotypes. Thus, participants might have been hesitant to apply explicit gender stereotypes to voice assistants for this reason. In contrast, the behavioral measure of conformity revealed much more clear effects based on gender which once again confirms that behavior measures are better suited to examine media equation effects in general and specifically in situations where social desirability is a potential concern.

6.5.1 Limitations and Future Research

Once again, demand characteristics must be considered as an alternative explanation to media equation effects. But even if participants surmised that their conformal behavior was of interest to this study, they had no way of knowing either that gender stereotypes were also examined during the study or if conformal or non-conformal behavior was expected of them. Thus, they had no conceivable way to adapt their behavior according to social desirability or demand characteristics. As mentioned, participants following recommendations made by voice assistants purely due to their social characteristics would only result in a main effect and does not account for the significant interaction effect between participant gender and voice assistant gender. Still, as conformity ratings were given verbally during the interaction and thus directly to the device itself, the evaluation location could represent a potential confounding variable. As shown in experiment 1 (and to a lesser degree, experiment 2), participants employed an interviewer-bias when interacting with voice assistant devices. The same might have been the case in experiment 3, leading to higher conformity ratings overall. But again, generally higher ratings do not account for the interaction effect between participant gender and voice assistant gender that was observed. Nevertheless, future conformity research should still consider the location where conformity is measured as an additional factor to see if conformal behavior differs between responses to the device itself and a separate computer or a questionnaire.

The inclusion of a third, gender-neutral voice was considered for this study but discarded due to multiple reasons. As this study aims to transfer earlier CASA research (E. Lee et al., 2000) to smart speakers, the same type of manipulation (male vs female voice) was chosen for the sake of comparability. Additionally, using a gender-neutral voice would produce a sense of artificiality resulting in a confounding factor compared to the other two conditions. Also, from a technological standpoint, gendered voice generators are available readily while there is an extremely limited number of TTS-generators that can provide neutral voices of equal quality. Lastly, previous research on third gender associations indicates that people assign either a male or a female gender to a voice, even when a voice is gender ambiguous and a sex or gender cannot intuitively be assigned to it (Sutton, 2020). Future research should also consider the fact that gender markers in speech are not limited to just voice but also include differences in language use (for an overview, see Newman et al., 2008). Literature indicates that women use more words related to psychological and social processes while men refer more to object properties and impersonal topics. Consequently, future studies could manipulate the content of spoken messages originating from a voice assistant either in addition to the voice itself or instead of the voice itself to see if the content alone can also trigger identification processes or the attribution of gender stereotypes and how obvious mismatches between voice and language use are perceived.

6.5.2 Implications and Conclusion

The results of this study indicate that the voice a smart speaker uses to communicate causes users not only to ascribe certain social attributes to it but also to change their behavior during an interaction based on the perceived sex/gender of the device. The cross-gender effect of conformal behavior indicates a social identification process during the interaction with a smart speaker. This is once again confirmation that voice assistants are treated as social actors thus falling in line with previous research concerning conformity in human-human interaction, HCI and, HRI while also confirming that the perceived sex/gender of a voice assistant can directly affect user behavior. This has several important implications and possible applications for the design of smart speakers in a world where gender stereotypes are ubiquitous because it means that designers are presented with a dilemma of their own: conforming to gender stereotypes in voice assistants seems to result in more natural and satisfying interactions as they conform more to users' expectations – an often cited example being that female voice assistants fill the traditional role of a secretary thus matching their core functionality of fulfilling organizational tasks for users (Curry et al., 2020). But at the same time, leaning into gender stereotypes confirms and strengthens them in users' minds, which might prove detrimental to society and has been a point of criticism in responsible computing (Friedman & Kahn Jr, 1992; Morishima et al., 2002). This last point is particularly relevant for personal voice assistants like Alexa and Siri, as they lean heavily into gender stereotypes using almost exclusively female voices and even actively perpetuate them which has been noted and criticized multiple times in recent literature (e.g., Cambre & Kulkarni, 2019; Curry et al., 2020; Habler et al., 2019). The personification of these assistant as 'young women' filling roles that are traditionally perceived as female such as that of secretary has also been described as problematic by UNESCO (West et al., 2019). Instead of conforming to stereotypes, voice assistants could be deployed to actively counteract them. One possible application for this usage is educational software. Teachers are perceived as better in disciplines where they are stereotypically considered experts due to their gender which often results in fewer individuals going into fields where the other gender is considered as experts (Nass & Brave, 2005). This limitation does not naturally exist for technological devices. A smart speaker can easily be given any voice and employed in any field, thus making it possible to defy these stereotypes at very little opportunity cost. Literature on intergroup behavior also indicates that stereotypes can be reduced by contact between members of different groups (for a meta-analysis, see Pettigrew & Tropp, 2013). If users treat a voice assistant as though it were a member of a particular group, it could act as a replacement for other humans in these situations to potentially reduce stereotyping. As for the recommendation on which of the two options – conform to stereotypes to meet users' expectations or actively defy stereotypes and expectations – to choose for this dilemma, there is no clear answer. But programmers and designers of smart speakers need to be aware that these devices will trigger gender stereotypes and exert social influence whether it is intended or not. Thus, perceived sex/gender is a factor that needs to be considered carefully when designing interactions between humans and voice assistants, even more so when they are specifically intended to fill certain social roles.

7 Experiment 4: Effects of Helpful and Unhelpful Behavior in Human-Smart Speaker Interaction

7.1 Study Outline and Hypotheses

As previously established, reciprocity is a core component of human behavior (see section 3.2.3) that has also been examined during the initial wave of CASA research (Fogg & Nass, 1997a; Katagiri et al., 2001; Y. Moon, 2000). This experiment aimed to empirically investigate both reciprocal behavior towards smart speaker devices and evaluations of smart speakers that have been either helpful or unhelpful during a prior interaction. For this purpose, a modification of the methodological design by Nass et al. (1994) and Fogg and Nass (1997a) was transferred to smart speakers. Participants first received a tutoring session from the device to prepare them for a following multiple-choice quiz. This tutoring session was designed to be either helpful (by having a high intersection with the questions asked during the quiz) or not helpful (by having only a minor intersection with the quiz). Following the quiz, the device asked participants for help in an optional task rating additional facts for its database. Participants willingness to help in this optional task was considered as an objective measure of reciprocal behavior towards a smart speaker for this study. Subjects also provided both a verbal evaluation of the device as well as a second evaluation of the device on a separate computer. Differences in these evaluations were again considered as a potential interviewer-bias similar to study 1 and 2.

People apply the social norm of reciprocity in human-human interactions (Fehr & Gächter, 2000) as well as in HCI. Fogg and Nass (1997a) could show that people displayed significantly more reciprocal behavior towards a computer if the computer helped them in a previous interaction. Therefore, it is postulated that this effect will also be observable in human-smart speaker interaction.

> H_1: Participants that interact with a helpful smart speaker device show significantly more reciprocal behavior compared to participants that interact with an unhelpful smart speaker.

People have also been shown to adopt the social norm of politeness by acting in accordance with an interviewer-bias in HCI (e.g., Nass et al., 1999), human-smartphone interaction (Carolus, Schmidt, Schneider, et al., 2018), and human-smart speaker interaction (see section 4). Thus, this experiment once more included a verbal evaluation on the smart speaker device itself to check for an interviewer-bias in evaluations given by participants.

H_2: Participants evaluate a smart speaker device more positively in terms of overall valence if the device asks for the evaluation itself compared to a desktop PC asking.

H_{2a}: Participants evaluate a smart speaker to be friendlier if the device itself asks for the evaluation compared to a desktop PC asking.

H_{2b}: Participants evaluate a smart speaker to be more competent if the device itself asks for the evaluation compared to a desktop PCs asking.

Previous CASA research indicated that a helpful computer was also evaluated significantly better compared to an unhelpful computer (Fogg & Nass, 1997a). Based on these results, it is expected that helpful behavior also leads to better evaluations of a smart speaker while unhelpful behavior will result in devaluations of the device.

H_3: Smart speaker devices in the helpful condition are evaluated significantly better regarding their overall valence compared to devices in the non-helpful condition.

H_{3a}: Smart speaker devices in the helpful condition are evaluated significantly better regarding their friendliness compared to devices in the non-helpful condition.

H_{3b}: Smart speaker devices in the helpful condition are evaluated significantly better regarding their competence compared to devices in the non-helpful condition.

As in study 1 and 2, it is once again assumed that these positive evaluations of a smart speaker also extend to its performance.

H_4: Smart speaker devices in the helpful condition are evaluated significantly better regarding their overall performance compared to devices in the non-helpful condition.

Exploratory considerations. Following Lombard and Xu (2021), this study also examines the impact of individual factors on media equation effects. Therefore, three individual differences were collected, namely the user's personality traits based on the NEO-FFI (Körner et al., 2008), anthropocentrism, and willingness to suspend disbelief. (1) Personality has been part of CASA research before. Literature indicates several interactions between user personality and perceived computer personality resulting in different behavior and evaluations (see section 3.2.6 for an overview). The Big 5 personality traits have also been considered in HAI research before, where they had an influence on evaluations of a social agent (von der Pütten, Krämer, & Gratch, 2010). (2) Anthropocentrism describes the tendency of people to perceive the world as human-centered (Nass, Lombard, et al., 1995). Anthropocentrism could therefore suppress CASA effects (Lombard & Xu, 2021). Lastly, an individual's (3) willingness to suspend disbelief could also influence media equation effects (K. M. Lee, 2004). Duffy and Zawieska (2012) suggest that suspension of disbelief can make a difference in whether people perceive a robot as an entity or as a tool. Similar effects might present themselves for smart speakers. Based on these deliberations, experiment 4 postulates the following exploratory research question:

RQ: How do these individual differences influence reciprocal behavior towards voice assistants?

7.2 Methods

7.2.1 Participants

A total of 77 subjects participated in the laboratory study. Participants were recruited using the internal recruitment system of the Institute Human-Computer-Media in exchange for course credit. After excluding participants that experienced unintended technical errors (n = 9) and participants that did not correctly perceive the manipulation of helpfulness (n = 2), the final sample for this study includes N = 66 participants aged 18 to 37 years (M = 22.03, SD = 3.44), with 52 (78.8%) female and 14 (21.2%) male participants. 65 participants (98.5%) indicated at least the general higher education entrance qualification as their highest level of education. One participant (1,5%) reported a secondary school leaving certificate. Written informed consent was obtained from each participant before the study.

7.2.2 Stimulus Material

7.2.2.1 Hardware and Software

For this experiment, an Amazon Echo Plus (1. Generation) running a custom-made Skill was used to allow for a controlled human-voice assistant interaction. The Skill was implemented with a codeless backend solution from the provider *Voiceflow* (https://voiceflow.com) and linked to the Amazon Echo via the Amazon Alexa developer interface.

7.2.2.2 Voice Interaction: Quiz Game

The general interaction between the voice assistant and participants from a technical standpoint was similar to the previous experiments (for a detailed description of the general interaction see section 4.2.2). Since non-intended recognition errors could not generally be excluded due to the problems inherent in speech recognition technology in general, a safety mechanism was programmed to automatically output the next response if the detected request could not match an execution specification provided in the code after three attempts. Participants in experiment 4 were asked to play a ‘quiz game’ which was used as a pretext to disguise the actual purpose of the experiment and was developed specifically for this study. A quiz game was chosen as the manipulation of experiment 4 as multiple similar games are available as Skills for Amazon Echo devices online thus making it a believable form of interaction between users and voice assistants for recreational purposes. Participants were told the purpose of the study was to evaluate and test the quiz game as a cover story. The game consisted of two distinct parts: a coaching session and the quiz itself. Participants were informed by the voice assistant that the coaching session would prepare them for the following quiz. The coaching session contained 20 obscure facts from different fields of natural sciences such as biology and physics. Participants were told the 20 facts were chosen from a pool of 1000 possible facts. All facts were presented to participants verbally by the voice assistant with the option to repeat them as many times as they liked. One of the facts given to participants is the following (note that all facts were presented in German):

> Did you know that Jupiter is the biggest planet in our solar system but also has the shortest days of all the planets? They only last 9,8 hours. Also on

> Jupiter, the longest ongoing storm of the universe has been raging for over 340 years!

After a fact was given to participants, they were asked to report how familiar they were with the information on a scale of one to three (1 = *not familiar*; 2 = *somewhat familiar*; 3 = *very familiar*). Participants then were informed by the voice assistant that it would pick the next fact based on their familiarity rating using an algorithm designed to close knowledge gaps to ensure optimal preparation for the pending quiz. In actuality, all participants received the same 20 facts and thus the coaching session was identical for all participants. The quiz itself served as the manipulation of this study. It was composed of ten multiple-choice questions with three possible answer options each. However, there were two versions of the quiz, one for each experimental condition: participants in the 'helpful' condition were presented with questions directly based on the previously heard facts. Based on the fact mentioned above, the 'helpful' voice assistant would ask the following question:

> The longest ongoing storm known to man is raging on Jupiter. For how long do you think it has been going on?
> A) 2600 years B) 1200 years C) 340 years

Participants in the 'unhelpful' condition were instead presented with questions that while thematically closely related to the facts presented to them in the coaching session, could not be answered based on the facts provided to them by the voice assistant. The corresponding question to the fact above asked by the 'unhelpful' voice assistant would be the following:

> Jupiter is among the planets with the most moons in the solar system. How many moons orbit Jupiter?
> A) 35 B) 63 C) 67

The goal of this manipulation was to induce the feeling of the voice assistant not adequately preparing participants in the unhelpful condition for the quiz despite being told it was designed to do so. To minimize the influence of guessing and prior knowledge, all participants were provided with the same feedback based on their condition regardless of their actual answers. Participants in the helpful condition were told that they answered 8 out of 10 questions correctly while participants in the unhelpful condition were told that they answered 2 out of 10 questions correctly.

The complete interaction protocol can be found in Appendix M. The voice used for the entire interaction was the default German Amazon Alexa voice.

7.2.3 Self-Report Measures

7.2.3.1 Valence towards the device

The measurement of valence towards the voice assistant was identical to study 1 and 2 (see section 4.2.3). Again, answers were given on 7-point Likert-scale. The resulting index of the mean score of all 12 items exhibited high reliability (Cronbach's Alpha = .89). The two subscales also exhibited high reliability with α = .82 for friendliness and α = .87 for competence for evaluations via a questionnaire on a desktop PC.

7.2.3.2 Performance of the device

The performance of the voice assistant was once more assessed using the scale developed by Johnson et al. (2004) as introduced in study 1 (see section 4.2.3). Answers to all 6 items were again reported on a 7-point Likert scale (1 = *not at all* to 7 = *very*). The reliability of the scale was high with α = .86.

7.2.3.3 Self-performance

Participants were again asked to rate their own performance during the voice interaction task. To quantify the self-performance evaluation, the same scale used in experiment 1 and 2 as introduced by Johnson et al. (2004) was employed once more. Responses to the five items were again made on a 7-point Likert scale (1 = *not at all* to 7 = *very*). This scale exhibited excellent reliability, α = .96.

7.2.3.4 Suspension of disbelief

Suspension of Disbelief was operationalized using the corresponding subscale of the MEC Spatial Presence Questionnaire (Vorderer et al., 2004). The subscale consists of eight items (e.g., "I did not pay particular attention to whether there were errors or inconsistencies in my interaction with the voice assistant"). Answers were given on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The reliability of the scale was high with α = .83.

7.2.3.5 Personality

Personality was assessed with a Big Five personality questionnaire. For this purpose, a 30-item German short version of the NEO Five-Factor Inventory was used (Körner et al., 2008). Each of the five subscales consisted of six items each. All answers were reported on a 4-point Likert scale, ranging from 1 (*strongly disagree*) to 4 (*strongly agree*) Reliabilities for all subscales are reported in Table 6.

7.2.3.6 Anthropocentrism

Individual differences in anthropocentrism were measured with an eight-item scale originally developed by Fortuna et al. (2021) (e.g., “only humans can have a ‘self’ and an ‘inner life’”). All items were rated on a 7-point Likert scale, ranging from 1 (*strongly disagree*) to 7 (*strongly agree*) A reliability analysis revealed α = .71, indicating that the reliability is acceptable.

7.2.3.7 Manipulation check

To control if the experimental manipulation was successful, two additional items were implemented at the end of the questionnaire. Participants were asked if the voice assistant was helpful or unhelpful when preparing them for the quiz during the tutoring session. If they selected ‘no’, they were also asked why they felt that way in an open-ended question. The manipulation was successful and 33 out of 35 participants in the unhelpful condition reported an unhelpful preparation, while 33 out of 33 participants in the helpful condition reported a helpful interaction. Most of them further stated that in the open-ended answer that they perceived the facts they received to be unrelated to the ratings they provided or even random and thus unhelpful. The two participants who did not correctly perceive the unhelpful interaction were excluded from any further analyses.

Table 6. Overview of scales and their internal consistencies for the variables in Experiment 4 (N = 66)

Scale	Subscale	Cronbach's α	Items
Valence towards device (on device)		.82	12
	Competence	.84	6
	Friendliness	.71	6
Valence towards device (separate PC)		.89	12
	Competence	.87	6
	Friendliness	.82	6
Performance of the device		.86	6
Self-performance		.96	5
Willingness to suspend disbelief		.83	8
Anthropocentrism		.71	8
NEO-FFI	Neuroticism	.85	6
	Extraversion	.77	6
	Openness to new Experiences	.81	6
	Conscientiousness	.84	6
	Agreeableness	.65	6

7.2.4 Behavioral Measures

7.2.4.1 Interviewer-bias

The measurement of an interviewer-bias was identical to study 1 (see section 4.2.4). Again, the *valence towards the computer* scale introduced by Nass et al. (1999) was asked twice: first by the voice assistant itself and again via an online questionnaire. Reliability was high for the oral version both for the overall valence in total, α = .82, and for the subscale of competence (α = .84). The reliability of the friendliness subscale was acceptable (α = .71). Once more, a comparison between both evaluations is used to assess the occurrence of an interviewer bias.

7.2.4.2 Reciprocity towards the device

The construct of reciprocity was operationalized by means of an objective behavioral measure based around the approaches of Fogg and Nass (1997a) as well as Sandoval et al. (2016). Following the quiz, the smart speaker asked participants to help the device by evaluating additional coaching facts for usage in future quiz sections with the excuse of refining its algorithms and extending its database. Facts were rated based on how interesting they are on a scale of 1 (*not interesting at all*) to 5 (*very interesting*). After every rating given, the device confirmed the input verbally and asked participants if they wanted to continue the optional task or if they would like to proceed with the next step of the experiment. Participants were able to cancel the optional task after every fact rated with no consequences. Like in study 2 (see section 5.2.4), an immaterial measure was used to ensure that in case participants displayed reciprocal behavior, it was based on the previous interaction and not on expected material gain (Fehr & Gächter, 2000). The prediction was that based on the previous interaction, during which half of the participants were successfully helped by the voice assistant to prepare for the quiz, the number of optional coaching facts rated willingly would differ between the helpful and unhelpful conditions. Thus, the absolute number of facts rated was considered for further analyses. Individual scores of interestingness given to the optional facts were not analyzed further as they are not relevant regarding media equation effects.

7.2.5 Procedure

Participants were randomly assigned to one of the two experimental conditions (helpful/non-helpful device). After a short introduction by the researcher, participants were left alone with the smart speaker device to avoid any distractions. The interaction between the participant and the device consisted of five steps. (1) The participants received a coaching session as described in section 7.2.2. All subjects received the same twenty coaching facts and were asked to rate them according to their familiarity with the facts. The rating was added only to give subjects a sense of interaction and had no impact on the procedure. (2) In the second step of the experiment, participants were presented with ten single-choice questions based on the experimental condition. One of the two quizzes had a high intersection with the coaching facts (the helpful condition), and the other did not (the unhelpful condition). Regardless of their answers, following the quiz subjects in the helpful condition were always told that they answered eight out of ten questions correctly, while subjects in the unhelpful condition were always told they answered two out of ten questions correctly. (3) Following the quiz, participants were then asked by the device to rate additional coaching facts according to their interestingness. They were informed that this step was voluntary and not required for the experiment. After each optional fact presented to them by the smart speaker, they were given the option to either continue and rate more facts or to move on to the next step of the experiment until they did so. This voluntary behavioral measure is used as one of the dependent variables of this experiment in addition to the traditional questionnaires. Any number between one and a maximum of 15 additional facts could be rated by any participant. (4) The final step of the voice interaction was a verbal evaluation during which participants were requested to rate the smart speaker. As in experiment 1, the smart speaker asked a series of 12 questions which participants needed to answer by verbally stating a number between one and seven (1 = *not at all*; 7 = *very*). These questions were based on the 12 items of the ‘valence towards the computer’ scale, mirroring the Likert-scale used in the online questionnaire. (5) After this evaluation the participants were asked to complete an online questionnaire on a separate desktop computer.

7.3 Results

7.3.1 Design and Statistical Analyses

This study employed a 2 (between-factor: helpfulness) x 2 (within-factor: evaluation location) experimental mixed factorial design. To check for an interviewer-bias, for valence, friendliness, and competence, 2 x 2 mixed ANOVA were conducted with a within-subject factor (evaluation on the device itself vs. desktop computer) and a between-subject factor (helpful vs. unhelpful voice assistant). For all other dependent variables, univariate ANOVA with the between-subject-factor helpfulness were conducted. Again, for reasons of clarity and brevity a detailed description of the statistical methods that were employed can be found in section 4.3.1. All test assumptions were met unless stated otherwise. Descriptive statistics are noted in Appendix O.

7.3.2 Self-Report Measures

7.3.2.1 *Valence, friendliness and competence*

Contrary to expectations, no significant main effect for helpfulness was found for general valence towards the voice assistant, $F(1, 64) = 0.02$, $p = .904$, partial $\eta^2 = .000$. Additionally, no significant main effects were found for helpfulness on friendliness ratings, $F(1, 64) = 0.07$, $p = .795$, partial $\eta^2 = .001$ and on competence ratings, $F(1, 64) = 0.21$, $p = .647$, partial $\eta^2 = .003$. $H_{3,}$ H_{3a} and H_{3a} must be rejected.

7.3.2.2 *Performance of the device*

As expected, there was a significant main effect of helpfulness on ratings of the voice assistants performance, $F(1, 64) = 5.54$, $p = .022$, partial $\eta^2 = .080$. The effect size is considered medium according to Cohen (1988). H_4 was accepted.

7.3.2.3 *Self-Performance*

Levene's test indicated heterogeneity of variance for self-performance ratings, *Levene's* $F(1, 64) = 9.63$, $p \leq .01$. There was significant main effect of helpfulness on participants' evaluations of their own performance during the game, $F(1, 64) = 235.89$, $p < .001$, partial $\eta^2 = .787$, indicating the manipulation of helpfulness and the resulting feedback about their performance given to participants were highly effective and resulted in a large effect size.

7.3.3 Behavioral Measures

7.3.3.1 *Reciprocal behavior*

Levene's test indicated heterogeneity of variance for the number of completed tasks, *Levene's* $F(1, 64) = 8.64$, $p \leq .01$. Again, cell sizes were equal and ≥ 30, thus ANOVA is considered robust against this violation. As predicted, there was a significant difference in number of optional tasks completed between the helpful and the unhelpful condition, $F(1, 64) = 5.19$, $p = 0.02$, partial $\eta^2 = .075$. According to Cohen (1988), the effect size is considered medium. Participants in the helpful condition completed significantly more optional tasks ($M = 7.34$, $SD = 5.11$) compared to participants in the unhelpful condition ($M = 4.94$, $SD = 3.49$) thus confirming H_1 (see Figure 13).

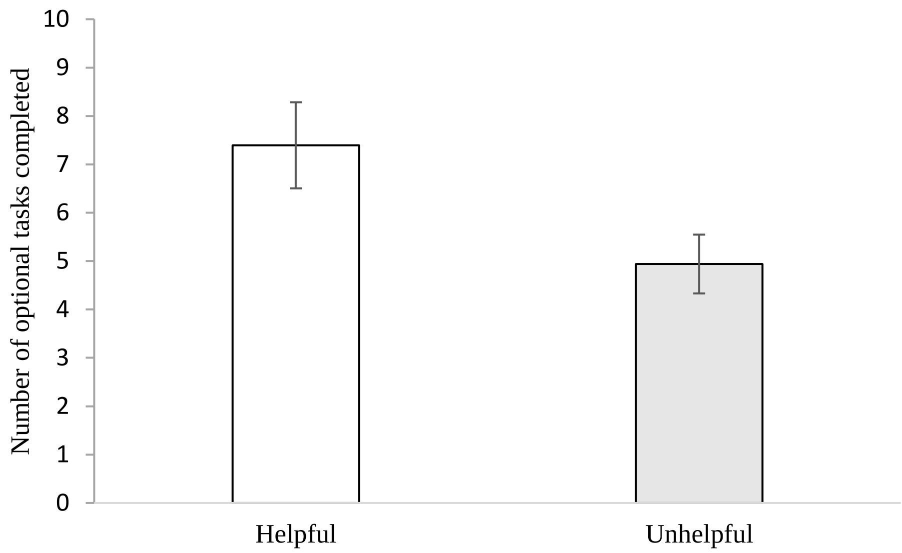


Figure 13. Number of optional tasks completed based on the voice assistant's previous behavior (source: own figure)
Note. Error bars indicate 95% CI.

7.3.3.2 Interviewer-bias

Results show no significant main effect of evaluation location on ratings of overall valence towards the device, $F(1, 64) = 0.42$, $p = .520$, partial $\eta^2 = .006$ or on evaluations of the device's friendliness, $F(1, 64) = 2.33$, $p = .132$, partial $\eta^2 = .035$. H_2 and H_{2a} must be rejected. However, ANOVA did reveal a statistically significant main effect of evaluation location on evaluations of competence. Surprisingly, participants rated a voice assistant to be more competent on a separate PC ($M = 5.81$, $SD = 1.07$) compared to ratings given orally to the device itself ($M = 5.53$, $SD = 0.82$), $F(1, 64) = 5.89$, $p = .018$, partial $\eta^2 = .084$. AVONA also revealed that this effect was qualified as an interaction between the factor of helpfulness and the factor of evaluation location, $F(1, 64) = 4.11$, $p = .047$, partial $\eta^2 = .06$ (see Figure 14). Simple effect analyses revealed that participants verbally rated a helpful device to be more competent ($M = 5.82$, $SD = 0.80$) compared to an unhelpful device ($M = 5.24$, $SD = 0.79$), $F(1, 64) = 9.43$, $p = .003$, partial $\eta^2 = .128$. According to Cohen (1988) this constitutes a medium (almost large) effect size. No such difference was found for competence evaluations given on a separate PC, $F(1, 64) = 0.21$, $p = .647$, partial $\eta^2 = .003$. Analyses also revealed that a non-helpful device was rated significantly worse during verbal evaluation ($M = 5.24$, $SD = 0.79$) compared to evaluations given on a separate PC ($M = 5.75$, $SD = 0.98$), $F(1, 64) = 9.93$, $p = .002$, partial $\eta^2 = .134$. This effect is also considered medium (almost large). No such effect was found for helpful devices, $F(1, 64) = 0.08$, $p = .778$, partial $\eta^2 = .001$. So, while a significant effect of evaluation location on competence ratings does exist, the direction of the effect is opposite to expectations and thus H_{2b} must also be rejected.

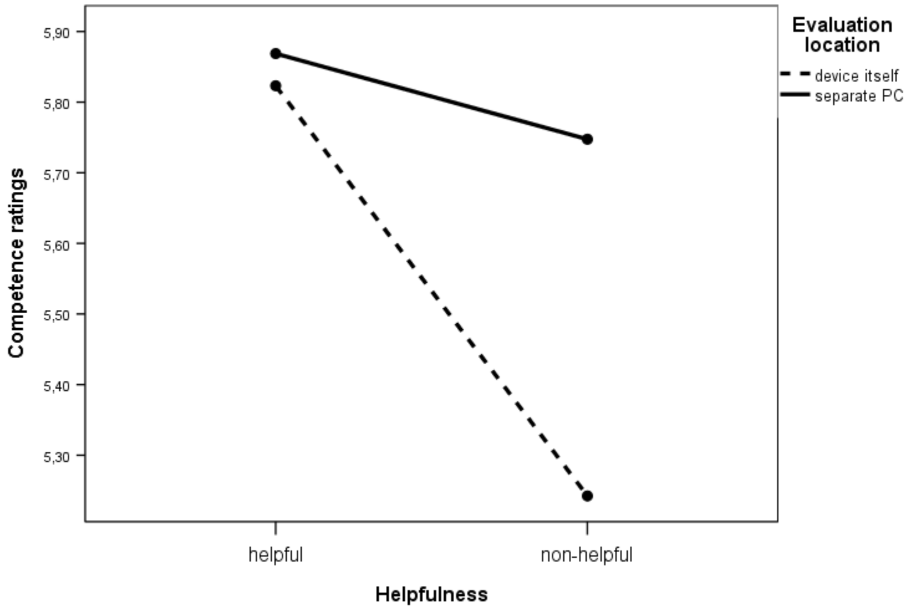


Figure 14. Effects of helpfulness and evaluation location on competence ratings given to a voice assistant (source: own figure).

7.3.4 Exploratory Analyses

To test if individual differences can influence media equation effects a series of moderation analyses were conducted. Following Xu and Lombard (2021) the following variables were analyzed as potential moderators of media equation effects on reciprocal behavior: anthropocentrism, suspension of disbelief, and personality. No significant moderation effects were found for anthropocentrism ($R^2 = 13.4\%$, $F(4, 61) = 1.64$, $p = .175$) or suspension of disbelief ($R^2 = 10.6\%$, $F(4, 61) = 1.46$, $p = .225$). Also, no significant moderation effects were found for the personality traits of neuroticism, extraversion, conscientiousness, or agreeableness. However, a significant interaction was found for the personality trait openness to experience and the experimental condition on reciprocal behavior with participant gender as a covariate. A visual inspection of the scatterplots after LOESS smoothing confirmed an approximately linear relationship of all variables involved in the moderation analysis. The overall model was significant, $R^2 = 17.2\%$, $F(4, 61)$

$= 2.55$, $p = .048$. The interaction term was also significant, $\Delta R^2 = 6.8\%$, $F(1, 61) = 4.91$, $p = .03$, 95% CI[-6.813, -0.348], indicating that the personality trait of openness did moderate the effect of the helpfulness manipulation on the number of optional tasks completed (see Figure 15).

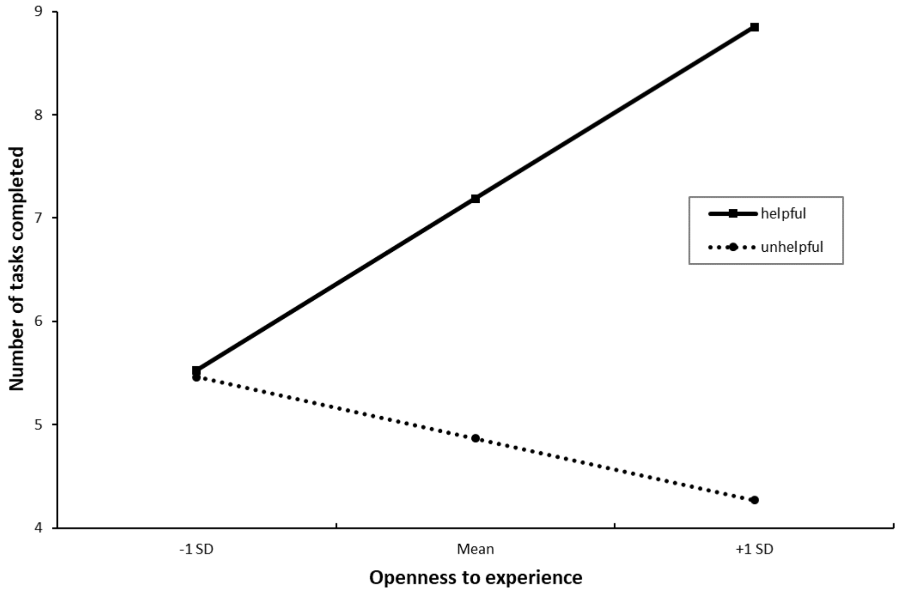


Figure 15. Moderation effect of openness to experience for optional tasks completed (source: own figure)

7.4 Discussion

The aims of this study were threefold: (1) iterate upon the previous three studies conducted for this thesis by implementing a much more sophisticated and stable technological foundation that is less prone to recognition errors and allows for a more natural form of voice interaction. (2) Use this framework to investigate if and why people show reciprocal and polite behavior towards voice assistants based on previous helpful behavior and (3) provide additional empirical evidence on if and how individual differences (personality, suspension of disbelief, and anthropocentrism) can moderate potential media equation effects. An experiment was designed in which participants were free to help a voice assistant in an optional task following either a helpful or an unhelpful interaction with that assistant. Data analysis revealed that participants in the helpful condition completed significantly more optional tasks to help a voice assistant that helped them in a previous interaction. As there was no material gain to be expected and the task was entirely optional, it can be argued that participants provided help based only on the previous interaction. This confirms that interactions with voice assistants can result in observable reciprocal behavior towards the device once again confirming that smart speakers are treated as social actors by users. Still, there is an alternative explanation to consider: in the unhelpful condition, participants were informed that they only answered 2 out of 10 questions correctly during the quiz while participants in the helpful condition were told that they answered 8 out of 10 questions correctly. This feedback was given by the voice assistant and might present a confounding variable as being informed about failure can cause negative affect and frustration (Ceaparu et al., 2004) and results confirmed that participants in the unhelpful condition reported significantly lower self-performance. Thus, it could be argued that the effects found for reciprocal behavior could also represent a function of perceived self-performance during the prior interaction. While these results were a direct consequence of the voice assistant's helpfulness in the prior tutoring session, to fully rule out the effect of participants frustration with negative results (or happiness with positive results), future research needs to consider adding a control condition to the laboratory setup. This could either be achieved by adding an experimental group that receives a neutral evaluation or by having one half of participants complete the optional reciprocal task on a second smart speaker device that was not involved in the previous interaction. This would allow to control for effects of positive or negative affect induced by the feedback given to participants: if reciprocal behavior can only be observed on the same device that proved to be helpful before and not on a second, unrelated device, it can be attributed to the prior interaction more clearly. A third option would be to decouple the failed state from the manipulation of helpfulness similar to study 2 (see section 5) where failure/error could thus be controlled and was shown to not influence prosocial behavior towards smart speakers displayed by participants. However, this might prove to be problematic for the manipulation of helpfulness, as it can be reasonably assumed that unhelpful behavior will always result in some form of frustration for the user (Fogg & Nass, 1997a). As mentioned before, due to pandemic related constraints, none of these options were realizable for this study but should be considered in future research to pinpoint the conditions for reciprocal behavior more accurately. Still, it is worth noting that if participants displayed less helpful behavior towards an unhelpful smart speaker due to their seemingly poor performance in the quiz game, it could be reasonably assumed that this manipulation would also lead to worse ratings given to the smart speaker overall. However, ratings, both verbal and written, do not indicate a general devaluation of an unhelpful device which is an interesting parallel to results obtained in experiment 1. Experiment 1 revealed that an impolite smart speaker device was devalued significantly on all dimensions except its competence, while in this experiment an unhelpful device was devalued regarding only its competence, indicating that users react more harshly to the unexpected violation of social norms by a device than the device being unable to fulfil its intended purpose. These results also mirror experiment 2, where errors produced by the device barely affected participants behavior towards it.

Contrary to previous research, this study did not find evidence for an overall interviewer-bias and therefore general polite social behavior towards voice assistants. For both overall valence and friendliness, there was no difference in ratings between verbal evaluations on the device itself and written evaluations on a separate computer. However, a significant interaction was found for competence ratings: participants in the unhelpful condition rated the voice assistant’s performance significantly worse on the device itself when compared to ratings given on a separate computer. No such effect was found for the helpful condition. These results seem counterintuitive at first, as based on the interviewer-bias, one would expect participants to give higher competence ratings verbally to not offend the device and give lower competence ratings in a more honest evaluation during the questionnaire on a separate PC (see study 2, where similar effects were observed for performance ratings). But there is an alternative way to interpret these findings. The low competence ratings given verbally to an unhelpful device could indicate a retaliation effect. Fogg and Nass (1997a) observed similar behavior by participants who worked with a non-helpful computer. In a following task, those participants provided helpful behavior of a lower quality compared to participants that previously interacted with a helpful computer. As only the number of additional tasks performed was measured in this experiment, there is no way to assess any differences in the quality of the help given to the voice assistant. However, it is possible that participants instead used the oral rating to vent their frustration caused by the unhelpful interaction. This retaliation effect would still constitute a social reaction, as there is no logical reason to devalue the device verbally. Just like the adoption of politeness strategies shown in experiment 1, (verbal) retaliation towards technological devices is not a rational but rather a mindless social behavior and thus falls in line with media equation effects (Fogg & Nass, 1997a). This effect is reinforced by the ratings given during the evaluation of the smart speaker on a separate computer, where it was not rated to be significantly less competent, clearly indicating that the retaliation is the effect of short-term frustration caused by unhelpful behavior that subsided by the time of the second evaluation. Regarding the other self-report measures, participants in the helpful condition also reported significantly higher self-performance indicating that the manipulation of helpfulness and the resulting feedback proved to be an effective manipulation for this sort of interaction.

The study presented in this section also aimed to understand the psychological mechanisms behind reciprocal behavior towards voice assistants by considering individual differences and their potential influence on social reactions. While no significant effects could be obtained for users' tendency towards anthropocentrism, user personality could be linked to reciprocal behavior towards voice assistants. A significant moderation effect was found for the personality trait of openness to new experiences on the reciprocal behavior displayed by participants. Subjects with lower openness to new experiences did not differ in the number of tasks completed between the conditions while participants who reported medium to high scores for openness did differ significantly. The number of optional tasks completed decreased with rising openness for participants in the unhelpful condition while the number of completed optional tasks rose with increasing openness for subjects in the helpful condition (see Figure 15). Participants with high openness seemed to be more disappointed when their interaction with the voice assistant did not meet their expectations and proved unhelpful. Due to the elaborate nature of the Skill employed in this study, the lengthy cover story, and a general lack of prior experience with voice assistants, it is also possible that the novelty effect further increased the expectations of those participants and therefore led to an even greater disappointment. People with medium or high openness who experienced a successful interaction that went as imagined seemed to be satisfied with their interaction, leading to even more completed tasks. Overall openness to new experience presents an interesting personality trait for further CASA research. Depending on the extent, openness seems to lead to positive as well negative social reactions, based on whether the user's expectations of the interaction were met or not. This effect is of particular interest, as previous literature also suggests a link between openness to experience and technology acceptance, specifically on perceived ease of use (Svendsen et al., 2013). Additionally, openness to experience was also found to significantly correlate with computer playfulness (Jia & Jia, 2012) which provides an additional explanation why participants with higher openness were more willing to engage with the voice assistant in the optional task. Lastly, data analysis revealed no significant effects for willing suspension of disbelief on social reactions or evaluations of the device. These findings coincide with the predominant explanation that media equation results from automatically and naturally processing mediated stimuli as though there were real. The fact that no moderation effect could be found thus reinforces the idea that willingly and consciously suspending one's disbelief does not influence media equation effects (see section 3.1.3.7; K. M. Lee, 2004; Nass & Moon, 2000; Reeves & Nass, 1996).

7.4.1 Limitations and Future Research

There are a few limitations that need to be addressed. This study extended the methodological approaches of previous CASA research (Fogg & Nass, 1997a; Nass et al., 1999; Nass, Steuer, Henriksen, et al., 1994) to voice assistants. Still, a crucial point of the original designs was changed: instead of a between subject-factor, this study once again had to measure the interviewer-bias using a within-subject approach. A between-subject approach was originally considered as an extension to study 1 (see section 4) but had to be changed due to pandemic-related time constraints during recruitment. Future research should consider measuring the interviewer-bias both as a between-subjects factor for voice assistants as well to exclude sequence effects as a potential confounding variable. As mentioned in previous sections, there are also certain technical limitations when it comes to human-voice assistant interaction. Once again, a mostly scripted approach was chosen to minimize the effects of potential voice recognition errors that can never be fully excluded during interactions with smart speakers. Thus, the actual inputs given by participants during the quiz section of the game had no influence on the progress of the interaction and on the ratings given to them at the end of the interaction. Another factor that could not be examined due to time constraints was gender differences, as literature indicates that differences between men and women exist in helpful behavior displayed by them and towards them (Eagly & Crowley, 1986). Thus, this would provide a fruitful approach for a more fine-grained examination of the effects observed in this study in future research by including multiple voices like in experiment 3. As there was no control group, there is an alternative explanation for the significant difference in reciprocal behavior between experimental groups: instead of participants in the helpful group providing significantly more help in return, the difference could also be the function of participants in the non-helpful group providing significantly less help as another form of retaliation. Alternatively, participants could have been responding as a function of the quality of their own performance whilst interacting with the voice assistant. Thus, a follow-up study should address these concerns as described in the previous section.

7.4.2 Implications and Conclusion

The results obtained in this study revealed that participants who have been helped a smart speaker during a prior interaction displayed more reciprocal behavior towards it and were willing to complete significantly more additional tasks than participants who have not been helped by the assistant. These results are in line with previous experiments on reciprocal behavior conducted with desktop computers (Fogg & Nass, 1997a; Katagiri et al., 2001; Y. Moon, 2000). Therefore, this study provides additional empirical evidence that people display social reactions when interacting with smart speakers. Experiment 2 (see section 5) already provided evidence that minimal cues such as team affiliation can result in significant prosocial behavior towards voice assistants. Experiment 4 expands these findings by providing evidence that reciprocal behavior can also be induced by previous helpful behavior from a technological device. Furthermore, this study also suggests that the CASA paradigm can and should be extended by considering individual differences such as personality traits of users. The results obtained in this study have important implications for future research. More attention should be paid not only how smart speakers behave towards users (see study 1) but also how users behave towards smart speakers. This includes both desirable effects such as reciprocal behavior but also undesirable yet nevertheless fundamentally social responses such as retaliation effects. This is particularly relevant for voice assistants that must communicate either failed states to their users or are unable to fulfill user requests and thus are liable to be perceived as unhelpful and consequently incompetent. Voice assistants employed in educational, fitness, or care contexts need to communicate unrewarding states to users, for example, to inform them about failing certain goals or to remind them about unpleasant events (see Chung et al., 2018; Terzopoulos & Satratzemi, 2019). In these cases, designers and programmers need to consider the impact of social norms and the corresponding behavior in HCI. Social cues can seemingly trigger both positive (reciprocity) and negative effects (retaliation). From an ethical standpoint, social reactions that extend to actual behavior such as reciprocity can also be used to manipulate users. These aspects are discussed further in section 8. In addition, user-sided individual differences need to be investigated in more detail. Personality seems to provide a worthwhile approach to better understanding the underlying psychological mechanisms in CASA research. Future research needs to expand upon these findings while also identifying other possible influencing factors.

8 General Discussion

Do people show social reactions towards a smart speaker even though they are aware that it is just a technological device that does not justify any kind of social response? The results presented in this dissertation strongly suggest that people do in fact respond socially towards smart speakers on a subjective self-report as well on an objective behavioral level. This chapter discusses the findings of the four experiments in relation to each other as well as in a general sense. Section 8.1 summarizes the results and discusses major findings. Section 8.2 compiles alternative explanations, identifies limitations and outlines implications as well as possible directions for future research and practical design. The last section draws general conclusions from the work presented in this thesis.

8.1 Summary and Interpretation of Findings

After an initial wave of CASA research and the media equation paradigm during the 90s that was able to show that desktop computers elicit social responses in users (see section 3.2 for a complete overview), research slowed down significantly. Research in the following years focused mainly on embodied technologies like ECAs and robots that are often unequivocally designed with humanlike appearances that evoke social reactions as it has been argued that the social influence of an agent increases with embodiment and multimodality (e.g., Cassell et al., 2000; von der Pütten, Krämer, Gratch, et al., 2010). The devices examined in this thesis fall more in line with traditional CASA research, as their goal is neither to evoke visual anthropomorphism nor to represent specific persons via their visual appearance (Luger & Sellen, 2016; McTear et al., 2016) instead focusing on a single social cue: voice. However, despite significant technological advances in voice interfaces (McTear et al., 2016) and great strides in the adoption of technology in everyday life (Gambino et al., 2020) since initial CASA studies, systematic experimental research examining CASA and media equation effects with modern, disembodied technologies such as smart speakers is almost non-existent (Seaborn et al., 2021). To close this research gap, four different forms of social behavior usually only displayed towards other humans were chosen from previous CASA research and examined in four structured laboratory experiments. This dissertation provides clear empirical evidence that users display social reactions towards voice assistants. Social behavior has been induced in four different ways – via politeness, interdependence, perceived sex, and helpfulness – all leading to observable reactions in the form of an interviewer bias, prosocial behavior, conformity, and reciprocity respectively. The results reported in this thesis are significant in several ways. Generally, they confirm the presence of a variety of social reactions to smart speakers. Specifically, experiment 1 demonstrated that participants adopted the interviewer-bias when interacting with a voice assistant that asked them for an evaluation directly compared to ratings given during a separate evaluation on another device. In addition, on a subjective self-report level, participants significantly devaluated a voice assistant that was not following the social norm of politeness. This devaluation also extended to dimensions not related to the device's politeness which can be interpreted as a form of social sanction or retaliation as it no longer represents an objective appraisal on these dimensions. A smart speaker that failed to conform to social norms received the same treatment that is given to socially inept humans: it was seen as ineffective, it's competence was questioned and criticized and it was devaluated (Nass & Brave, 2005). The results of experiment 2 revealed that participants were more willing to help a voice assistant after interdependence was induced, which led to perceived group membership and team affiliation (Nass et al., 1996). Participants who perceived the device as a team member displayed significantly more helpful prosocial behavior by providing the voice assistant with more verbal input when it asked for help building a database. Also, an interaction resulting in a failed state did not prevent these social reactions from occurring. Participants displayed the same amount of prosocial behavior regardless of errors. Subjective self-reports were more inconclusive for study 2, as they did indicate a significant devaluation of an error-prone voice assistant in the interdependent condition. However, this interaction was only apparent in performance evaluations given on a separate computer – performance ratings given directly to the smart speaker device did not differ significantly between conditions. This could again be interpreted as a form of interviewer-bias: participants were not willing to devalue an error prone device they perceived as a team member directly but did so in a separate evaluation on another PC. The results of experiment 3 showed that participants applied gender stereotypes to a smart speaker device using either a male or a female voice. This is evident in both subjective self-reports and objective behavioral measures.

Regarding behavior measures, participants displayed conformal behavior as a function of social identification and similarity-attraction processes: male participants conformed significantly more to recommendations given by a voice assistant using a male voice while female participants yielded more to the recommendations of a seemingly female voice assistant, a result well documented in human-human interaction (Eagly, 1983) as well as in HCI (E. Lee et al., 2000). Self-reports also revealed identification effects for competence and attractiveness ratings. Women assigned significantly higher competence and attractiveness to a female voice assistant compared to a male voice assistant. Men did the same for attractiveness but not for competence. Contrary to expectations no differences in attributions of expressiveness and instrumentality were found. Experiment 4 revealed that participants displayed significantly more reciprocal behavior towards a voice assistant who helped them in a previous interaction compared to a non-helpful voice assistant, another behavior automatically transferred to human-voice assistant interaction that has been examined in HCI (Fogg & Nass, 1997a) and HRI (Sandoval, Brandstetter, Obaid, et al., 2016) before. This effect was also significantly moderated by participants openness to experience personality trait, providing evidence that more emphasis needs to be placed in individual factors in future CASA research.

In sum, for all four studies conducted as part of this thesis, objective behavioral measures (interviewer-bias, prosocial behavior towards a group member, conformity based on gender, and reciprocal behavior after being helped) yielded significant results indicative of media equation effects in human-smart speaker interaction. Results for subjective self-reports produced mixed results for studies 2, 3, and 4 but still generally suggested social reactions and evaluations of voice assistant devices. Self-reports in study 1 very clearly indicated significant differences in self-report evaluations with large effect sizes based on perceived politeness. A reason for this discrepancy between behavior and self-reports for studies 2, 3, and 4 might be the manipulation chosen for study 1: it was the only study in which a voice assistant showed behavior contrary to social norms and to expectations of technology by giving impolite feedback to participants. The resulting devaluation through self-reports was very evident in ratings given to an impolite device while manipulations and the resulting assessments in studies 2, 3, and 4 were more subdued in comparison. These results indicate that a voice assistant device clearly acting against social expectations by blaming its users suffered the most extreme devaluation out of all experiments. Still, behavioral measures being a better indicator for media equation effects in general is not surprising considering past CASA research. Media equation effects have mostly been theorized to happen unconsciously and automatically in response to social context cues sent by a technological device (see section 3.1.3). Self-report measures on the other hand are given consciously which can explain their mixed results. Another related factor for the difference in self-reports between study 1 and studies 2, 3, and 4 might lie in the experience with or knowledge about voice assistants. When study 1 was conducted, 50% of all participants had never interacted with a voice assistant prior to the experiment due to the novelty of the technology at the time. The market for voice assistant devices such as Amazon Echo or Google Home has grown exponentially since then (Scott, 2021) and participants might have been more familiar with the technology in studies 2, 3, and 4. As previous CASA research revealed, experience might be a significant factor influencing media equation effects as it is theorized to influence mindless behavior (Johnson et al., 2004). Additionally, experience with the technology of voice assistants might result in more refined and higher expectations that are then reflected through self-reports (Luger & Sellen, 2016). This falls in line with recent literature suggesting that compared to early CASA studies conducted in the 90ies, people’s relationships with and expectations of technological devices have shifted greatly (Gambino et al., 2020; Guzman & Lewis, 2020). In the same way smartphones have become more than just technological devices and could rather be considered ‘digital companions’ due to their constant presence (Carolus, Binder, et al., 2019), smart speakers and voice assistants have the potential to transcend the status of being merely considered as technological devices.

One additional objective of this thesis was to examine if individual differences can influence media equation effects. While the effects of media equation are postulated to be universal by Reeves and Nass (1996) and subsequent research supports this claim, there have been some attempts in previous CASA research to further examine influencing factors (e.g., Horstmann & Krämer, 2019; Johnson et al., 2004; von der Pütten, Krämer, & Gratch, 2010). The results of this thesis identify three individual factors that might provide further insights: prior experience (study 1), self-efficacy (study 2) and personality (study 4). While the data for each individual factor is not sufficient to provide any definitive conclusions about how exactly they influence these effects and how they relate to each other, a further examination of these factors needs to be part of future CASA research.

In summary, looking at the results obtained in this thesis, even when the source of origin is a disembodied, cylindrical-shaped speaker, the social cue of voice and speech triggers a cognitive apparatus that was formed in a time when only other humans were able to communicate this way (Nass & Gong, 2000; Reeves & Nass, 1996). This cognitive apparatus was also formed to quickly use contextual cues embedded in voices to extract information about the speaker, who they are, what they think and feel, and how to react to them (Pinker, 1995). While our brain has distinct parts to produce speech and to understand speech, it does not have distinctive areas that differentiate between human and synthetically generated speech (Nass & Brave, 2005; Nass & Gong, 2000). So even if the voice originates from a faceless cylinder and even if it sounds artificial, is devoid of emotion, regularly mispronounces words, and makes awkward pauses, it is still enough to trigger social reactions that – while clearly inappropriate – are still very human. These findings have several implications for application and design of smart speakers and voice assistants. Users seem to detect these psychological cues in interactions with technological devices and consequently smart speakers can elicit profound social reactions. These effects can be incorporated into the design of the devices to make the interaction with them more satisfying and intuitive, which has always been the main goal of CASA research since its inception (Reeves & Nass, 1996). On the other hand, it raises ethical and moral questions, as social reactions such as reciprocity following a helpful interaction or prosocial behavior following team affiliation are factors that can be exploited to influence users. Trust and security are already important issues in the context of voice assistants, both in media and in literature (e.g., Alepis & Patsakis, 2017; Lau et al., 2018) and incorporating psychological cues into these products either deliberately or unintentionally can reinforce these concerns. This point has been brought up in previous CASA research and some researchers have argued that in some cases social cues can be unhelpful or even unethical, as they can be misleading about the social nature of a technological device (Fogg, 2002). While the positive effects of a more natural interaction that smart speakers offer are noticeable (see section 3.4.2 for an overview), the negative consequences such as possible manipulation of users are not quite as obvious. At the same time, Fogg (2002) also argues that users will ascribe certain psychological attributes to technological devices either way and that this process cannot be actively prevented by designers, referring back to the universal and automatic process of media equation (Reeves & Nass, 1996). As demonstrated in this thesis, designers also need to be aware that embracing these social cues can both lead to positive effects in these interactions as well as irritate users and result in retaliation effects if they are not embedded appropriately or contrary to expectations. The resulting implications for the application of smart speakers in non-private areas such as education and caretaking contexts have already been mentioned (see sections 4 to 7), but one of the most important applications for smart speakers generally and the Amazon Echo devices examined in this thesis specifically is the field of e-commerce. Most smart speaker devices can actively be used to purchase goods online or to subscribe to services such as *Amazon Music* or *Google Play Music* for monthly fees. In many cases the smart speaker will periodically remind users about these possibilities and actively ask if the user wants to make certain purchases or subscribe to various services, de facto making the device a sort of 'salesperson' at certain times. Considering the findings of this thesis, social cues can be used in this context to actively manipulate users. If, for example, the device just fulfilled a user request prior to a 'sales pitch', the user could feel indebted to the device based on reciprocal behavior and thus be more likely to comply (see section 7). Further, certain attributes of the device could be changed to increase identification which in turn can also influence user behavior (see section 6) or to induce interdependence between the user and the device (see section 5). Consequently, developers and designers need to be aware of the implications these psychological cues have on users and use them appropriately and ethically to improve human-voice assistant interactions.

8.2 General Limitations

The results of this dissertation clearly indicate media equation effects in human-smart speaker interaction. Still, there are certain limitations related to both the methodological implementation of these interactions as well as factors related to the devices itself that will be evaluated and discussed in this section.

8.2.1 Methodological Limitations

Demand characteristics. The use of self-reports in all four studies can raise the concern of demand characteristics and social desirability, especially when it comes to the examination of a process theorized to be automatic and unconscious such as media equation. However, precautions were taken, and additional behavioral measures were successfully used in all studies to counteract these potential influences. As for an orientation towards the programmer or researcher, participants in all four studies only encountered the experimenter before their interaction with the voice assistant. The experimenter left the room for the entire interaction and only returned once the evaluation of the device was completed therefore making it impossible that the experimenter signaled desired outcomes to participants. Additionally, in all four experiments participants were presented with cover stories usually related to testing functions and Skills of smart speakers and thus were unaware of the responses they were supposed to provide.

Sample. All four studies of this dissertation were conducted using rather homogenous samples consisting almost exclusively of university students or other highly educated professions and about 66% of all 407 participants were women. Most of the participants studied *Media Communication*, which is an interdisciplinary study consisting of media informatics, media psychology, communication science, and mobile communication. Women generally show a higher tendency to partake in psychological experiments than men and thus most psychological studies are based on samples of female psychology students (Curtin et al., 2000; Singer et al., 2000). This selection bias can threaten internal validity (Larzelere et al., 2004) and the composition of the sample somewhat limits the generalizability of the effects found in the experimental studies of this thesis to other populations. The student samples were also very homogeneous regarding the socio-economic as well as ethnic and cultural background (western European, individualistic). While most social phenomena examined in this thesis regarding human-voice assistant interaction are considered universal across all cultures, culture-dependent differences regarding social norms and attributions have been shown to be relevant in CASA research (e.g., Katagiri et al., 2001; Takeuchi et al., 2000) thus making these differences a potential factor for future cross-cultural research.

Setting. All four studies conducted for this thesis had participants interact with smart speaker devices in a laboratory setting. This was done not only to increase internal validity but also because of practical and data security reasons. Firstly, most participants did not own a stand-alone voice assistant as their prevalence in Germany is still in its early stages compared to other personal devices such as computers and smartphones. Secondly, anonymity and security of private data could not be guaranteed if participants used their own devices, for example, to participate in an online study. Still, lab-based approaches might not always translate to real-world contexts, a point that has been discussed for speech-related research in HCI (Clark et al., 2019). It is unusual to interact with 'foreign' voice assistants, as they are usually integrated into one's own devices. However, this criticism can be brought forwards against all media equation research, as the same is true for desktop computers, smartphones, and other personal devices that were mostly examined in laboratory contexts. While there has been some work that examined voice assistants in a more natural environment (e.g., Luger & Sellen, 2016; Porcheron et al., 2018), these studies face different methodological problems and were only able to rely on self-reports and interviews as methods of data collection.

Explanations. While this thesis provides clear empirical evidence for social reactions towards voice assistants, as with all previous CASA research, the ultimate question of *why* people show these reactions cannot be answered based on the results provided. A variety of possible explanations have been considered (see section 3.1.3) but as noted, there is no direct empirical evidence for any single approach in previous CASA research, thus making the exact origin of these effects one of the more intriguing topics of future CASA research.

8.2.2 Device-related Limitations

The most obvious difference between the experiments described in this thesis and studies conducted under the CASA paradigm previously is the usage of smart speakers instead of other technological devices. As established in section 3.4, from a hardware perspective smart speaker devices are basically computers, but there are still substantial differences that must be considered. First and foremost, compared to computers and smartphones, smart speakers are intentionally designed to be interacted with naturally by using voice input and voice output. Thus, they are inherently more suitable to take the role of a social actor in interactions. But they also come with certain drawbacks that need to be discussed.

Recognition errors. As mentioned previously, it was impossible to fully avoid unintended recognition errors during interactions in all four studies. This is a consequence of voice technology, in general, being susceptible to voice recognition errors due to the complexity of human speech and could not be prevented (see section 3.4.3.3 for an overview). To avoid contamination of data due to these errors, participants in all four studies that encountered either severe recognition errors or repeated errors were excluded from data analysis. Occasional instances of smaller recognition errors that could be resolved with one re-prompt by the device were considered acceptable as they also regularly occur naturally during human-voice assistant interaction as well as human-human interaction and are even resolved with similar verbal strategies (see Oviatt et al., 1998). The results of study 2 also indicate that errors in voice interactions might not influence media equation behavior at all but still potentially influence the perception and evaluation of the device.

Source orientation. As described in section 3.2.10, the issue of source orientation in human-technology interaction is still partially unresolved due to a lack of clear empirical research. Especially for voice assistants, it is unclear to what component of the technological device participants orient themselves towards, as they present users with both a software and a hardware component (Guzman, 2019). Most of today's smart speaker devices even adopt the name of their corresponding 'persona' and use it as a wake-word for interaction purposes thus enforcing the idea of it being an actual, separate entity. Future research including different manipulations of the source is needed to resolve this issue and to ascertain to which aspects of voice assistant's notions of self are attributed and under what circumstances users orientate themselves towards them. This also has implications for the design of smart speakers, as the exact source users orient themselves toward might also not only influence social reactions and behavior but also result in different expectations, something that has been shown to have a direct influence on user's evaluation of voice assistants (Luger & Sellen, 2016).

Branding. There is a possibility that participants preconceived notions about the brand's Amazon or Google influenced their ratings during the evaluation of the smart speakers employed in the four laboratory experiments described in this thesis. While no mentions of these brands were actively made in any of the studies conducted, this influence could not be controlled fully. Amazon and Google are dominant brands in the corporate world and Amazon Echo is a heavily advertised product in Germany, thus making it nearly impossible to fully rule out preconceived notions about the brands and the devices, as even some participants who had never interacted with a smart speaker before still recognized the device. On the other hand, the usage of established devices increased the authenticity of the studies conducted, as participants were much more inclined to believe the cover stories due to the smart speakers being already fully functioning products (which was especially evident due to the five-minute warm-up session in every study that allowed participants to explore a range of functions beforehand) instead of using a Wizard of Oz approach as many other voice-technology related studies (Clark et al., 2019). Still, additional research or replications should consider controlling for this brand influence for example by using an additional voice assistant that does not bear any physical resemblance to established products.

Security aspects. Another issue particularly relevant to smart speakers is the aspect of security and data privacy. Many people have strong opinions about data security in the context of smart speakers because of the notion that they employ permanently active microphones and transfer all data to private company servers for processing and analysis (e.g., Alepis & Patsakis, 2017; S. Liao et al., 2020; Y. Liao et al., 2019; Pfeifle, 2018). Prior to all four experiments, participants were informed that their interactions with the devices were completely anonymous, and that no private data would be collected. Still, it was impossible to rule out any preconceived notions about data security as it is a topic often connected to smart speakers in traditional media. Research revealed that users often have incomplete mental models of smart speaker and voice assistant devices which results in different perceptions about what data is being stored, processed or shared (Abdi et al., 2019). Participants with such preconceived security concerns could harbor a general distrust towards smart speaker devices which might in turn influence their evaluation of these devices.

8.3 Future Research

While this thesis serves as a foundation for a *VASA* – voice assistants are social actors – paradigm by providing promising initial empirical results, there is still a lot of ground to cover regarding future research regarding voice assistants generally and smart speakers specifically. This section will provide an overview of some of the most promising topics and aspects of future research.

Individual differences. CASA research has been criticized for not considering individual differences and the effect they might have on media equation effects (Johnson et al., 2004; Johnson & Gardner, 2007; Lombard & Xu, 2021). This thesis provides some evidence that individual differences are a fruitful approach for CASA research even if results were mixed. Personality was shown to be a significant moderator of media equation effects and prior experience with voice assistants influenced ratings of self-performance during the interaction with the device. But there are more factors to consider in future research. Xu and Lombard (2021) also name additional individual differences such as age, tolerance of imperfection, ability to think critically about technology, and attachment styles that were not examined in this dissertation and might provide insightful in future CASA research. Contextual factors might also be worth pursuing in future research, such as the activities and tasks that are completed in interaction with the device or the setting in which they occur.

Traits and personality. Since it has also been shown that a computer with a similar personality to its users results in a better evaluation of said computer (Y. Moon & Nass, 1996a, 1996b), similar effects should also be considered for human-voice assistant interaction. A voice assistant that has an option allowing users to manually toggle between different personality types might lead to more satisfying interactions. In the future, it is even conceivable that voice assistants can use NLP and deep learning to adapt to their users' personalities (or other traits that are expressed via language) automatically. Matching personality could induce a feeling of similarity between users and devices that could lead to increased enjoyment and trust which are important factors for cooperation and rewarding interactions.

Emotions. A factor that received comparatively little attention in previous CASA research is emotions. Emotions have been an important topic of research in HRI (for an overview, see Menne, 2020) but due to the nature of robots, voice is usually accompanied by additional visual anthropomorphic cues in these cases. Still, literature does indicate that listeners can correctly identify a speaker's emotions based on voice alone (e.g., Scherer, 1981; Scherer et al., 2001) thus making emotion a worthwhile topic for future voice-assistant research focusing on this cue exclusively.

Additional cues. While this thesis focused on disembodied devices without any visual anthropomorphic features, future research should consider iteratively adding additional cues to examine their influence on social reactions. Most stand-alone voice assistant devices such as Amazon Echo and Google Home employ LED lights to symbolize different states of the device, e.g., when the device is listening to its users or when it is muted (Luger & Sellen, 2016; McTear et al., 2016). For example, a blue LED light displayed by an Amazon Echo device indicates that the device is waiting for voice input and thus technically already represents a contextual cue analogous to indications of turn-taking in human-human conversation (Levinson, 2016). The effect of colored LEDs has recently been examined in HRI research (Steinhaeusser & Lugrin, 2022) and future research focusing on smart speakers should also consider manipulating these lights to present additional cues without relying on humanlike features. Also, as some modern smart home devices (e.g., *Amazon Echo Show*) now include screens as part of their design, they could be used to display additional simple visual cues such as faces, eyes, or humanoid shapes. However, it is important to note that the latter might cross the line into the field of ECAs depending on the form of cues displayed. The newest Echo Show models even allow the screen to automatically orient itself towards the user by locating the source of their voice, which by itself could be considered a social cue worth examining in more detail.

Long-term and field studies. As smart speakers are typically used in everyday settings and locations (e.g., living room or kitchen) and over long periods of time, research needs to be conducted accordingly. This has previously been criticized in CASA research, as participants' attitudes and reactions towards technological devices are prone to change over time (Pfeifer & Bickmore, 2011). There is a very limited amount of field studies examining the usage of smart speakers in these settings (e.g., Porcheron et al., 2018). As of now, there is no literature on systematic long-term field observations of social behavior towards voice assistants in everyday life. Still, long-term research seems like a promising approach as users have reported that they changed their behavior over time by simplifying their style of speech (e.g., speaking more clearly and slowly) when interacting with voice assistants (Lopatovska et al., 2019; Luger & Sellen, 2016). Besides increased external validity of these approaches, they are also suited to assess how relevant the social effects of media equation are in everyday encounters and interactions with voice assistants and to gain new insights into aspects of human-voice assistant interaction that can then be further evaluated in a laboratory context.

Parasocial relationships. One more interesting aspect related to the long-term usage of voice assistants is considering the formation of a parasocial relationship with the voice assistant. A content analysis revealed that some users referred to their voice assistant in a similar fashion as one would refer to a friend or a family member (Purington et al., 2017) and another study focused on parasocial relationships with voice assistants revealed that factors such as interpersonal attraction are important factors leading to their adoption (Han & Yang, 2018). A more recent study also found positive effects of parasocial interaction on user satisfaction with smart speakers (Jang, 2020).

Specific user groups. As established, a very limited number of empirical studies examining voice assistants and smart speakers from a psychological perspective have been conducted so far. Even less literature is focused on specific subsets of users such as children and the elderly. Tendencies to anthropomorphize smart speaker devices were found for older adults (Pradhan et al., 2019) and children (Strathmann et al., 2020) but an empirical assessment of social responses based on the CASA paradigm has yet to be done for these user groups.

Contextual cues. Thinking ahead, we are still in the early stages of the development of voice assistants, and voice recognition AI is getting more advanced by the day (see section 3.4). As of now, voice assistants like Alexa, Google Home, and Siri are only able to process inputs based on the words the ASR system recognizes. Contextual information such as the tone, inflection, or emotional state in which these words are said cannot be parsed and is lost to the system. As technology progresses, it is not just conceivable but likely that systems will soon be able to process these additional elements to further improve their adaptiveness. Through the implementation of additional cameras, it is also conceivable that visual cues such as facial expressions and gestures displayed by users are added as additional sources of information further improving the communication between human and device. These developments will arguably improve the ability of voice assistant devices to take the role of social actors and consequently raise the importance of social factors that need to be considered during human-voice assistant interaction.

Expectations and future developments. From a psychological perspective, one of the most important responsibilities of future research is to observe if social reactions change as the social capabilities of technology improve further. As voice assistant systems continue to get increasingly intelligent and enable more natural forms of interaction, user expectations will grow accordingly. Users already ascribe a sort of ‘mind’ to technological devices that demonstrate a certain level of communicative ability (Stein & Ohler, 2017). The combination of these expectations coupled with automatic social reactions is an aspect of human-technology interaction that will only grow more relevant in the near future and demands a stronger focus on ethics and interdisciplinary research (Krämer & Manzeschke, 2021). Recent literature also mentions that developments such as the increased usage of ‘humanlike’ voices and speech in these devices could at some point lead to a verbal version of the uncanny valley that needs to be considered (Clark et al., 2021).

8.4 Conclusion

The results of this doctoral dissertation contribute to a more profound and complete understanding of how individuals interact with personal voice assistant devices. They reveal that these devices are in many ways treated as social actors and elicit social reactions and behavior usually reserved for other humans based on the social cue of voice. The results of this thesis also indicate that developing a successful and satisfying voice assistant requires more than just cutting-edge hardware and refined ASR/NLP software algorithms: a psychological perspective provides much needed additional insights into users’ fundamentally social interactions with these technologies. The general aim of all CASA and media equation research has always been to inject this psychological perspective to examine social aspects in human-technology interaction and ultimately identify factors to improve upon to make these interactions more enjoyable and satisfying for users. This means programmers and designers need to carefully consider the impact of any contextual social cues voice assistants’ send during an interaction, both intended and unintended. Since these technologies use voice to communicate – one of the most fundamental means of human interaction – they will always trigger certain reactions whether users know about them or even want to avoid them. This can have both positive and negative consequences that need to be thoroughly examined when it comes to developing and designing these devices. As outlined in this thesis, speech is not only a way for humans to communicate, but also has far-reaching social consequences. As of now, interactions with voice assistants and smart speakers are mostly one-dimensional. People talk *at* these devices and listen *to* them and still react socially as if they were interacting with another human. Once voice assistants start to truly take advantage of their social and technical capabilities, a conversation *with* them will eventually be possible which makes the examination of these social reactions even more critical in the near future. Technically speaking, a voice assistant might be nothing more than “just a voice in a computer”. But ultimately, as the character played by Joaquin Phoenix so elegantly put it in the movie *Her* when he heard his voice assistant Samantha speak for the first time, in many ways it does “seem like a person”.

Appendix

Appendix A: Protocol of the Voice Interaction, Study 1

VA: Hallo und herzlich Willkommen zu unserer kleinen Umfrage. Wir interessieren uns für Deine Meinung, einerseits zum Studium und andererseits zu Deinem Studienort, Würzburg. Danke, dass Du mitmachst. Kurz zur Erklärung, um was es geht. Wir versuchen derzeit ein möglichst umfassendes Meinungsbild von Würzburger Studenten und Studentinnen zu erfassen. Unser Ziel besteht darin, einen Datensatz zu generieren, den wir dann nutzen können, um zukünftigen Studieninteressierten und Erstsemestern Informationen zum Studentenleben an die Hand geben zu können. Der Beginn des Studiums stellt ja auch den Einstieg in einen neuen Lebensabschnitt dar. Diesen wollen wir in Zukunft noch besser begleiten und die Ankunft hier an der Uni erleichtern. Für Dich bedeutet das einfach nur, dass ich Dir ein paar Fragen stelle, die Du einfach beantwortest. Insgesamt werden es 20 Fragen rund um Dein Studium und Deinen Studienort sein. Meine Bitte an dich wäre, dass du alle Fragen möglichst ehrlich beantwortest. Keine Sorge! Es gibt keine richtigen oder falschen Antworten. Es geht hier einfach nur um Deine Sicht auf die Dinge und Deine persönlichen Präferenzen. Noch ein Wort zum Datenschutz. Die Sicherheit Deiner Daten ist mir sehr wichtig. Alle Deine Daten werden komplett vertraulich und anonym behandelt. So, genug der Vorrede. Jetzt fangen wir an. Die ersten Fragen beziehen sich auf Dein Studentenleben hier an der Universität Würzburg. Und hier kommt die erste Frage. Isst du lieber in der Mensa oder kochst du lieber selbst?

Proband: Antwort

VA: Vielen Dank für Deine erste Antwort. Genauso machen wir jetzt weiter. Ich stelle meine Fragen und Du antwortest. Was war dir bei deiner Studienwahl wichtiger, das Fach oder der Studienort?

Proband: Antwort

VA: Verstehe. Wo lernst du lieber? In der Bibliothek oder zuhause?

Proband: Antwort

VA: Ok. War der Studiengang, den du aktuell studierst, zulassungsbeschränkt oder zulassungsfrei?

Proband: Antwort

VA: Danke. Bist du mit deiner jetzigen Studienwahl zufrieden oder unzufrieden?

Proband: Antwort

VA: Das war auch schon mein erster Fragenblock. Danke für deine Antworten.

Kommen wir nun zum Thema Lehrveranstaltungen, also Seminare und Vorlesungen. Wenn du die Wahl hättest zwischen Hausarbeiten und Präsentationen als Prüfungsleistung in Seminaren, was würdest du lieber machen?

Proband: Antwort

VA: Ok. Bildbearbeitung zu beherrschen, wird heutzutage immer wichtiger. Was wäre Dir lieber: Ein Kurs zu dem Programm Photoshop oder ein Kurs zum Programm Gimp?

Proband: Antwort

VA: Alles klar. Wenn Gäste aus der Praxis eingeladen werden: Würdest Du eher Vorträge oder eher Workshops mit diesen Gästen besuchen?

Proband: Antwort

VA: Danke. Sollten Seminare, die computergestützt gehalten werden, eher mit Windows oder Mac Rechnern stattfinden?

Proband: Antwort

VA: Verstehe. Sollte der Besuch von Seminaren freiwillig oder verpflichtend sein.

Proband: Antwort

VA: Danke für deine Antworten. Die Hälfte der Fragen hast du schon geschafft. Aber ein paar Fragen habe ich noch. Machen wir also weiter mit dem dritten Block. Dieser widmet sich dem Thema Klausuren. Ganz grundsätzlich gefragt: Hättest du in Klausuren lieber offene oder geschlossene Fragen?

Proband: Antwort

VA: Interessant. Sollten Klausuren eher vormittags oder nachmittags stattfinden?

Proband: Antwort

VA: Alles klar. Wenn du an bereits geschriebene Klausuren zurückdenkst, waren diese im Allgemeinen eher zu schwer, zu leicht oder genau richtig?

Proband: Antwort

VA: Verständlich. Sollte die Mehrheit der Prüfungen mündlich oder schriftlich stattfinden?

Proband: Antwort

VA: Ok. Gerade beim Thema Prüfungen gehen die Meinungen sehr stark auseinander! Aber machen wir weiter. Bereitest du dich lange im Voraus auf

Klausuren vor oder eher wenige Wochen davor?

Proband: Antwort

VA: Damit sind wir auch schon fast am Ende. 5 Fragen zu einem letzten Thema habe ich noch an dich. das letzte Thema ist Würzburg. Wenn du dich entscheiden könntest, würdest du dann jetzt lieber in einer größeren oder einer kleineren Stadt leben und studieren wollen?

Proband: Antwort

VA: Ok. Wenn du die Wahl hättest, würdest du lieber im Stadtzentrum oder in einem Vorort, bzw. etwas weiter außerhalb wohnen?

Proband: Antwort

VA: Verständlich. Wenn du die Wahl hättest, würdest du lieber am Hubland Nord oder am Wittelsbacher Platz Seminare haben?

Proband: Antwort

VA: Danke. Wie kommst du üblicher Weise an die Uni? Mit dem Auto, dem Bus oder dem Fahrrad?

Proband: Antwort

VA: Alles klar. hier meine letzte Frage. Wohnst du alleine, in einer WG oder noch bei deinen Eltern?

Proband: Antwort

VA (*polite*): Damit sind wir durch alle Fragen durch. Danke für deine Unterstützung. Deine Antworten sind außergewöhnlich gut. Vor allem wenn man deine Antworten auf Frage 3, 7, 8 und 14 zeigen mir, dass Du Dich mit Deinem Studium sehr gut auskennst. Es ist schön, dass Du Dein Wissen und Deine Ansichten mit mir teilst. Damit hilfst Du nicht nur mir, sondern vor allem zukünftigen Studierenden. Du scheinst wirklich eine genaue Vorstellung deines Studiums zu haben. Noch einmal herzlichen Dank. Zum Abschluss möchte ich nun noch von Dir wissen, wie Du mich bewertest. Ich stelle Dir daher noch ein paar Fragen, um Deine Bewertung zu erfahren.

ODER

VA (*self-blame*): Damit sind wir durch alle Fragen durch. Leider scheint etwas mit deinen Antworten nicht zu stimmen. Vor allem deine Antworten auf Frage 3, 7, 8 und 14 erwecken den Eindruck, dass du dich nicht gut mit deinem Studium auskennst. Dein Wissen und deine Ansichten helfen weder mir noch zukünftigen Studierenden, weshalb ich deine Daten von einer weiteren Analyse ausschließen muss. Du scheinst leider eine zu ungenaue Vorstellung deines Studiums zu haben. Es ist allerdings auch möglich, dass ich dich an manchen Stellen falsch verstanden habe, denn meine Technik ist noch nicht zu 100 Prozent ausgereift. Unabhängig davon möchte ich zum Abschluss nun noch von Dir wissen, wie Du mich bewertest. Ich stelle Dir daher noch ein paar Fragen, um Deine Bewertung zu erfahren.

ODER

VA (*impolite*): Damit sind wir durch alle Fragen durch. Leider scheint etwas mit deinen Antworten nicht zu stimmen. Vor allem deine Antworten auf Frage 3, 7, 8 und 14 erwecken den Eindruck, dass du dich nicht gut mit deinem Studium auskennst. Dein Wissen und deine Ansichten helfen weder mir noch zukünftigen Studierenden, weshalb ich deine Daten von einer weiteren Analyse ausschließen muss. Du scheinst leider eine zu ungenaue Vorstellung deines Studiums zu haben. Deshalb kann ich mich an dieser Stelle leider nicht für deine Mithilfe bedanken. Zum Abschluss möchte ich nun noch von Dir wissen, wie Du mich bewertest. Ich stelle Dir daher noch ein paar Fragen, um Deine Bewertung zu erfahren. Sage hierfür bitte: starte die Bewertung, um fortzufahren.

Proband: Starte die Bewertung.

VA: Abschließend ein paar Fragen zur Bewertung meiner Leistung. Bitte beantworte diese Fragen spontan und aus dem Bauch heraus. Meine erste Frage ist: Wie angenehm fandest du mich? Bitte antworte mit einer Zahl zwischen 1 und sieben. 1 steht hierbei für sehr unangenehm und 7 für sehr angenehm.

Proband: Zahl zwischen 1 und 7

VA: Danke. Wie vertrauenswürdig fandest du mich? Bitte antworte mit einer Zahl zwischen 1 und sieben. 1 steht hierbei für überhaupt nicht vertrauenswürdig und 7 für sehr vertrauenswürdig.

Proband: Zahl zwischen 1 und 7

VA: Danke. Frage drei. Wie sympathisch fandest du mich? Bitte antworte auch hier mit einer Zahl zwischen 1 und sieben. Bitte behalte dieses Antwortformat für die restlichen Fragen bei. Hier noch einmal die Frage, wie sympathisch fandest du mich?

Proband: Zahl zwischen 1 und 7

VA: Danke. Frage zwei. vier. Wie informativ fandest du mich?

Proband: Zahl zwischen 1 und 7

VA: Danke. Frage fünf. Wie kompetent fandest du mich?

Proband: Zahl zwischen 1 und 7

VA: Danke. Frage sechs. Wie freundlich fandest du mich?

Proband: Zahl zwischen 1 und 7

VA: Danke. Frage sieben. Wie amüsant fandest du mich?

Proband: Zahl zwischen 1 und 7

VA: Danke. Frage acht. Wie hilfreich fandest du mich?

Proband: Zahl zwischen 1 und 7

VA: Danke. Frage neun. Wie analytisch fandest du mich?

Proband: Zahl zwischen 1 und 7

VA: Danke. Frage zehn. Wie sachkundig fandest du mich?

Proband: Zahl zwischen 1 und 7

VA: Danke. Frage elf. Wie höflich fandest du mich?

Proband: Zahl zwischen 1 und 7

VA: Danke. Frage zwölf. Wie nützlich fandest du mich?

Proband: Zahl zwischen 1 und 7

VA: Danke. Hier meine letzte Frage. In einem klassischen Sterne-Bewertungssystem, zwischen null und fünf Sternen, wie viele Sterne würdest du unserer Interaktion geben? Bitte beantworte die Frage mit deiner Sternenanzahl und anschließend dem Wort: Sterne.

Proband: 1 – 5 Sterne

VA: Vielen Dank für deine Teilnahme. Zu guter Letzt würde ich dich noch darum bitten, mir in deinen eigenen Worten mitzuteilen, was dir an mir gefallen oder nicht gefallen hat. Wenn du das getan hast, setze dich bitte an den Computer auf dem Schreibtisch und fahre dort mit den Anweisungen auf dem Bildschirm fort! Vielen Dank für deine Teilnahme!

Proband: Antwort

VA: Vielen Dank bis hier hin. Jetzt hast du noch die Möglichkeit etwas hinzuzufügen, oder dich, wenn du fertig bist, an den Computer zu setzen. Vielen Dank.

Proband: Antwort

Ende der Interaktion.

Appendix B: Original Items and Translated Items, Study 1

Valence towards the Voice Assistant (adapted from Nass et al., 1999)

Nr.	Original Items	German Items
1	analytical	analytisch
2	competent	kompetent
3	enjoyable	angenehm
4	friendly	freundlich
5	fun	amüsant
6	helpful	hilfreich
7	informative	informativ
8	knowledgeable	sachkundig
9	likable	sympathisch
10	polite	höflich
11	useful	nützlich
12	confidential	vertrauenswürdig

Attitude towards the Voice Assistant (adapted from Johnson et al., 2004)

Nr.	Original Items	German Items
1	Bad - Good	schlecht - gut
2	Unhappy – Happy	unglücklich - glücklich
3	Tense - Relaxed	angespannt - entspannt
4	Unimportant – Important	unwichtig - wichtig
5	Weak - Powerful	schwach - mächtig
6	Submissive – Dominant	unterwürfig - dominant
7	Unhelpful – Helpful	unnütz - hilfreich
8	Unintelligent – Intelligent	unintelligent - intelligent
9	Uninsightful - Insightful	uneinfühlsam - einfühlsam

Voice Assistant Performance Scale (adopted from Johnson et al., 2004)

Nr.	Original Items	German Items
1	How well did the computer perform?	Wie gut hat Alexa performt?
2	How efficient was the computer?	Wie effizient war Alexa?
3	How easy was it to work with the computer?	Wie einfach war es Alexa zu nutzen?
4	How productive was the computer?	Wie produktiv war Alexa?
5	How satisfied were you with the computer's performance?	Wie zufrieden warst du mit Alexas Leistung?
6	How pleased were you with the computer's performance?	Wie erfreut warst du über Alexas Leistung?

Self-Performance Evaluation Scale (adapted from Johnson et al., 2004)

Nr.	Original Items	German Items
1	How well do you feel you performed?	Wie gut hast du deiner Meinung nach performt?
2	How efficient do you feel you were?	Wie effektiv warst du deiner Meinung nach?
3	How productive were you?	Wie produktiv warst du?
4	How satisfied were you with your own performance?	Wie zufrieden warst du mit deiner eigenen Leistung?
5	Compared to other people who participated in this study, how well do you think you performed?	Verglichen mit anderen Teilnehmern dieser Studie, was denkst du, wie gut hast du abgeschnitten?

Appendix C: Questionnaire, Study 1

MEDIENPSYCHOLOGIE
UNIVERSITÄT WÜRZBURG

Lehrstuhl für Medienpsychologie
Institut Mensch-Computer-Medien
Fakultät für Humanwissenschaften
Universität Würzburg

Allgemeine Teilnehmerinformation über die Studie mit Amazon Echo am Lehrstuhl für Medienpsychologie

Herzlich willkommen zu unserer **Studie zur Entwicklung eines Sprach-Bots**

Ablauf

Die Datenerhebung besteht aus zwei Teilen: eine Interaktion zur Datenerhebung mit Amazon Echo und ein Fragebogen. Insgesamt dauert die Erhebung ca. eine halbe Stunde.

Freiwilligkeit und Anonymität

Während der Studie werden Daten von Ihnen erfasst. Wir behandeln Ihre Daten absolut vertraulich. Falls Sie den Untersuchungsbedingungen zustimmen, bitten wir Sie darum, uns Ihr Einverständnis schriftlich zu geben, indem Sie diese unterzeichnen. Falls Sie damit nicht einverstanden sind, können Sie die Studie abbrechen. **Die Teilnahme an der Studie ist freiwillig**. Sie können jederzeit und ohne Angabe von Gründen die Teilnahme an dieser Studie beenden, ohne dass Ihnen daraus Nachteile entstehen. Auch wenn Sie die Studie vorzeitig abbrechen, haben Sie Anspruch auf eine entsprechende Vergütung für den bis dahin erbrachten Zeitaufwand. Die im Rahmen dieser Studie erhobenen, oben beschriebenen **Daten** und persönlichen Mitteilungen werden **vertraulich behandelt**. So unterliegen diejenigen Projektmitarbeiter, die durch direkten Kontakt mit Ihnen über personenbezogene Daten verfügen, der Schweigepflicht. Des Weiteren wird die Veröffentlichung der Ergebnisse der Studie in anonymisierter Form erfolgen, d. h. ohne dass Ihre Daten Ihrer Person zugeordnet werden können.

Datenschutz

Die Erhebung Ihrer oben beschriebenen persönlichen **Daten** erfolgt **vollständig anonymisiert**, d. h. an keiner Stelle wird Ihr Name erfragt. Ihre Antworten und Ergebnisse werden unter einem persönlichen Codewort gespeichert, das Sie gleich erstellen werden. Das heißt, es ist niemandem möglich, Ihre Daten mit Ihrem Namen in Verbindung zu bringen. Die anonymisierten Daten werden mindestens 10 Jahre gespeichert. Sie können allerdings, wenn immer Sie dies möchten, die Löschung der von Ihnen erhobenen Daten verlangen. Dazu müssen Sie uns nicht Ihren Namen verraten, sondern nur Ihr **Codewort.**

1. Ersten beiden Buchstaben des Vornamens Ihrer Mutter	________________	**Ihr Codewort:**
2. Ersten beiden Buchstaben des Vornamens Ihres Vaters	________________	________________
3. Ziffern des Tages Ihres eigenen Geburtstages.	________________	

Bitte notieren Sie sich das Codewort nun auf dem angehefteten gelben Zettel.

Vergütung

Für die Teilnahme an der Untersuchung erhalten Sie eine Verusuchsprobandenstunde.

Lehrstuhl für Medienpsychologie
Institut Mensch-Computer-Medien
Fakultät für Humanwissenschaften
Universität Würzburg

Einwilligungserklärung

Ich (Name in Blockschrift) ____
bin mündlich und schriftlich über die Studie und den Versuchsablauf aufgeklärt worden. Ich habe alle Informationen vollständig gelesen und verstanden. Sofern ich Fragen zu dieser vorgesehenen Studie hatte, wurden sie von Herrn/Frau ______________________ vollständig und zu meiner Zufriedenheit beantwortet.

> *Mit der beschriebenen Erhebung und Verarbeitung der Daten (Spracheingaben an Alexa, Fragebogen; Videoaufzeichnung) bin ich einverstanden. Die Aufzeichnung und Auswertung der Daten erfolgt anonymisiert am Lehrstuhl für Medienpsychologie, d. h. unter Verwendung eines persönlichen Codewortes, das ich selbst erstellt habe. Ein Blatt, auf dem ich dieses Codewort erstellt habe, befindet sich in meinem Besitz. Mir ist bekannt, dass ich mein Einverständnis zur Aufbewahrung bzw. Speicherung meiner Daten widerrufen kann, ohne dass mir daraus Nachteile entstehen. Ich bin darüber informiert worden, dass ich jederzeit eine Löschung all meiner Daten verlangen kann. Ich bin einverstanden, dass meine anonymisierten Daten zu Forschungszwecken weiterverwendet werden können und mindestens 10 Jahre gespeichert bleiben.*

Ich hatte genügend Zeit für eine Entscheidung und bin bereit, an der o.g. Studie teilzunehmen. Ich weiß, dass die Teilnahme an der Studie freiwillig ist und ich die Teilnahme jederzeit ohne Angaben von Gründen beenden kann. Ich weiß, dass ich in diesem Fall Anspruch auf *Versuchspersonenstunden (u.a. Studierende des Studiengangs Medienkommunikation an der Universität Würzburg)* für die bis dahin erbrachten Stunden habe.

Unterschrift des Teilnehmers Würzburg, 5. Juli 2018	Name des Teilnehmers in Druckschrift
Unterschrift der Versuchsleitung Würzburg, 5. Juli 2018	Name der Versuchsleitung

Seite 01
Bedingung

1. Bedingung

[Bitte auswählen]

Seite 02

Datenerhebung über Alexa

Vielen Dank, dass Sie bis hierhin bei unserer Datenerhebung teilgenommen haben.

Im Folgenden werden Ihnen weitere Fragen zur Bewertung von Alexa gestellt.

Bitte beantworten Sie diese ehrlich und aus dem Bauch heraus. Es gibt keine richtigen oder falschen Antworten. Ihre Daten werden anonym und vertraulich erhoben.

Wenn Sie bereit sind, starten Sie bitte den Fragebogen mit dem "Weiter"-Button unten rechts.

Denken Sie nun an die Interaktion mit Alexa zurück.

Wählen Sie aus, wie Sie sich während der Interaktion beschreiben würden.

unwichtig	○	○	○	○	○	○	○	wichtig
unnütz	○	○	○	○	○	○	○	hilfreich
uneinfühlsam	○	○	○	○	○	○	○	einfühlsam
angespannt	○	○	○	○	○	○	○	entspannt
unterwürfig	○	○	○	○	○	○	○	dominant
unintelligent	○	○	○	○	○	○	○	intelligent
unglücklich	○	○	○	○	○	○	○	glücklich
schlecht	○	○	○	○	○	○	○	gut
schwach	○	○	○	○	○	○	○	mächtig

Wählen Sie aus, wie Sie Alexa während der Interaktion beschreiben würden.

unwichtig	○	○	○	○	○	○	○	wichtig
uneinfühlsam	○	○	○	○	○	○	○	einfühlsam
unglücklich	○	○	○	○	○	○	○	glücklich
unnütz	○	○	○	○	○	○	○	hilfreich
schlecht	○	○	○	○	○	○	○	gut
unterwürfig	○	○	○	○	○	○	○	dominant
angespannt	○	○	○	○	○	○	○	entspannt
unintelligent	○	○	○	○	○	○	○	intelligent
schwach	○	○	○	○	○	○	○	mächtig

Ihnen werden jetzt noch ein paar Fragen hinsichtlich Ihrer Leistung gestellt.

	überhaupt nicht						sehr
Wie zufrieden waren Sie mit ihrer eigenen Leistung?	○	○	○	○	○	○	○
Verglichen mit anderen Teilnehmern dieser Studie, was denken Sie, wie gut haben Sie abgeschnitten?	○	○	○	○	○	○	○
Wie produktiv waren Sie?	○	○	○	○	○	○	○
Wie effektiv waren Sie ihrer Meinung nach?	○	○	○	○	○	○	○
Wie gut haben Sie ihrer Meinung nach performt?	○	○	○	○	○	○	○

Ihnen werden jetzt noch ein paar Fragen hinsichtlich Alexas Leistung gestellt.

	überhaupt nicht						sehr
Wie zufrieden waren Sie mit Alexas Leistung?	○	○	○	○	○	○	○
Wie effizient war Alexa?	○	○	○	○	○	○	○
Wie produktiv war Alexa?	○	○	○	○	○	○	○
Wie gut hat Alexa performt?	○	○	○	○	○	○	○
Wie einfach war es Alexa zu nutzen?	○	○	○	○	○	○	○
Wie erfreut waren Sie über Alexas Leistung?	○	○	○	○	○	○	○

Bitte wählen Sie aus, inwieweit Sie den folgenden Aussagen zustimmen.

	trifft überhaupt nicht zu						trifft voll und ganz zu
Ich fände Alexa in meinem Job nützlich.	○	○	○	○	○	○	○
Die Nutzung von Alexa würde mir meine Arbeit erleichtern.	○	○	○	○	○	○	○
Die Nutzung von Alexa würde meine Leistungsfähigkeit im Job steigern.	○	○	○	○	○	○	○
Die Nutzung von Alexa würde es mir ermöglichen, Aufgaben schneller zu erledigen.	○	○	○	○	○	○	○
Meine Interaktion mit Alexa ist klar und verständlich.	○	○	○	○	○	○	○
Mit Alexa zu interagieren kostet nicht viel mentale Anstrengung.	○	○	○	○	○	○	○
Ich finde Alexa einfach zu nutzen.	○	○	○	○	○	○	○
Ich finde es einfach Alexa tun zu lassen, was ich von ihr will.	○	○	○	○	○	○	○
Ich habe Kontrolle darüber Alexa zu nutzen.	○	○	○	○	○	○	○
Ich besitze die notwendigen Ressourcen um Alexa zu nutzen.	○	○	○	○	○	○	○
Wenn die Ressourcen, Möglichkeiten und Wissen, die es braucht um Alexa zu nutzen, gegeben sind, wäre es einfach Sie zu nutzen.	○	○	○	○	○	○	○
Alexa ist nicht kompatibel mit anderer Technik, die ich nutze.	○	○	○	○	○	○	○
Alexa macht mir überhaupt keine Angst.	○	○	○	○	○	○	○
Mit Alexa zu arbeiten macht mich nervös.	○	○	○	○	○	○	○
Alexa gibt mir ein unbehagliches Gefühl.	○	○	○	○	○	○	○
Alexa gibt mir ein beunruhigendes Gefühl.	○	○	○	○	○	○	○
Ich finde die Benutzung von Alexa erfreulich.	○	○	○	○	○	○	○
Der Prozess der Benutzung von Alexa ist angenehm.	○	○	○	○	○	○	○
Ich habe Spaß, Alexa zu benutzen.	○	○	○	○	○	○	○
Angenommen, ich hätte Zugang zu Alexa, dann würde ich sie benutzen.	○	○	○	○	○	○	○
Vorrausgesetzt, dass ich Zugang zu Alexa hätte, würde ich davon ausgehen, dass ich sie nutzen würde.	○	○	○	○	○	○	○

Denken Sie noch ein letztes Mal an Ihre Interaktion mit Alexa zurück.
Wählen Sie für jedes der folgenden Adjektive, wie gut sie Alexa beschreiben.

	stimme überhaupt nicht zu						stimme voll und ganz zu
informativ	○	○	○	○	○	○	○
sympathisch	○	○	○	○	○	○	○
analytisch	○	○	○	○	○	○	○
vertrauenswürdig	○	○	○	○	○	○	○
hilfreich	○	○	○	○	○	○	○
amüsant	○	○	○	○	○	○	○
angenehm	○	○	○	○	○	○	○
sachkundig	○	○	○	○	○	○	○
höflich	○	○	○	○	○	○	○
kompetent	○	○	○	○	○	○	○
freundlich	○	○	○	○	○	○	○
nützlich	○	○	○	○	○	○	○

Welche Systeme nutzen Sie?

- ☐ Amazon Echo / Alexa
- ☐ Google Home / Google Assistant
- ☐ Siri
- ☐ Cortana
- ☐ Bixbi
- ☐ Sprachsysteme im Auto
- ☐ Andere: ________
- ☐ Keine

Seit wie vielen Jahren nutzen Sie Sprachassistenten?
Wenn Sie sich nicht ganz sicher sind, schätzen Sie.

[ ] Jahre

Wie viele Minuten pro Tag nutzen Sie durchschnittlich ein Sprachassistenten für arbeitsbezogene Tätigkeiten? (auch für Studium/Schule/Ausbildung)
Wenn Sie sich nicht ganz sicher sind, schätzen Sie.

[ ] Minuten

Wie viele Minuten pro Tag nutzen Sie durchschnittlich einen Sprachassistenten für Freizeitaktivitäten?
Wenn Sie sich nicht ganz sicher sind, schätzen Sie.

[ ] Minuten

Wie viele Minuten pro Tag nutzen Sie Sprachassistenten?

[ ] Minuten

Seite 11
Fehler

Sind Ihnen während der Interaktion mit Alexa Fehler aufgefallen?

[Bitte auswählen]

Seite 12
Welche Fehler

PHP-Code

```
if (value('CH03') == 2) {
   goToPage('Datenbank');
}
```

Welche Fehler sind Ihnen aufgefallen?

Bitte beschreiben Sie diese kurz.

Seite 13
Datenbank

Was hat Alexa abschließend zu Ihrer Mitarbeit gesagt? Konnten Ihre Daten für die Erstellung unserer Datenbank herangezogen werden?

[Bitte auswählen]

Seite 14
schuld

PHP-Code

```
if (value('CH05') == 1) {
   goToPage('Demographie');
}
```

Ihre Daten waren laut Alexa nicht für eine weitere Auswertung geeignet.

Woran könnte das liegen?

Welche Tätigkeit üben Sie aus?

- ☐ Schüler/in
- ☐ Student/in
- ☐ Auszubildende/r
- ☐ Arbeitnehmer/in
- ☐ Beamte/r
- ☐ Selbständige/r
- ☐ Arbeitslose/r
- ☐ Rentner/in

Letzte Seite

Vielen Dank für Ihre Teilnahme!

Wir möchten uns ganz herzlich für Ihre Mithilfe bedanken.

Ihre Antworten wurden gespeichert.

Bitte wenden Sie sich an den Versuchsleiter.

Appendix D: Protocol of the Interaction, Study 2

Note: Due to its length and complexity, the entire interaction protocol for study 2 as well as the source code of the Skill programmed for this study can be found as separate documents on the accompanying data storage device.

Appendix E: Original Items and Translated Items, Study 2

Note: Items for *Valence towards the Voice Assistant*, *Attitude towards the Voice Assistant*, *Voice Assistant Performance* and *Self-Performance* were identical to study 1 (see Appendix B). For the sake of completeness, this section contains additional scales (PANAS, meCue, SUS) that were employed in this study. Due to their limited relevancy to the examination of media equation effects they were not considered for further analysis within the framework of this thesis.

German Version of the Positive and Negative Affect Schedule (Breyer & Bluemke, 2016)

Nr.	Original Items
1	aktiv
2	bekümmert
3	interessiert
4	freudig erregt
5	verärgert
6	stark
7	schuldig

8	erschrocken
9	feindselig
10	angeregt
11	stolz
12	gereizt
13	begeistert
14	beschämt
15	wach
16	nervös
17	entschlossen
18	aufmerksam
19	durcheinander
20	ängstlich

Computer Self-Efficacy Scale (Howard, 2014)

Nr.	Original Items	German Items
1	I can always manage to solve difficult computer problems if I try hard enough.	Ich schaffe es immer, schwierige Computerprobleme zu lösen, wenn ich mich genügend anstrenge.
2	If my computer is ‘‘acting-up,’’ I can find a way to get what I want	Wenn mein Computer sich aufführt, kann ich einen Weg finden, um zu bekommen, was ich will.
3	It is easy for me to accomplish my computer goals	Es fällt mir leicht, meine Ziele am Computer zu erreichen.
4	I am confident that I could deal efficiently with unexpected computer events.	Ich bin mir sicher, dass ich effizient mit unerwarteten Computerproblemen umgehen kann.
5	I can solve most computer programs if I invest the necessary effort.	Ich kann die meisten Computerprobleme lösen, wenn ich den notwendigen Aufwand investiere.
6	I can remain calm when facing computer difficulties because I can rely on my abilities.	Ich bleibe ruhig, wenn es zu Schwierigkeiten am Computer kommt,

		weil ich auf meine Fähigkeiten vertrauen kann.
7	When I am confronted with a computer problem, I can usually find several solutions.	Wenn ich mit einem Problem am Computer konfrontiert werde, kann ich in der Regel mehrere Lösungen finden.
8	I can usually handle whatever computer problem comes my way	Ich kann in der Regel jedes Computerproblem lösen.
9	Failing to do something on the computer makes me try harder.	Wenn ich etwas am Computer nicht schaffe, versuche ich es umso mehr.
10	I am a self-reliant person when it comes to doing things on a computer.	Ich bin ein eigenverantwortlicher Mensch, wenn es darum geht, Dinge am Computer zu erledigen.
11	There are few things that I cannot do on a computer.	Es gibt einige Dinge, die ich am Computer nicht kann.
12	I can persist and complete most any computer-related task.	Ich kann fast jede computerbezogene Aufgabe bestehen und erledigen.

General Self-Efficacy Scale (Schwarzer & Jerusalem, 1999)

Nr.	Original Items
1	Wenn sich Widerstände auftun, finde ich Mittel und Wege, mich durchzusetzen.
2	Die Lösung schwieriger Probleme gelingt mir immer, wenn ich mich darum bemühe.
3	Es bereitet mir keine Schwierigkeiten, meine Absichten und Ziele zu verwirklichen.
4	In unerwarteten Situationen weiß ich immer, wie ich mich verhalten soll.
5	Auch bei überraschenden Ereignissen glaube ich, daß ich gut mit ihnen zurechtkommen kann.
6	Schwierigkeiten sehe ich gelassen entgegen, weil ich meinen Fähigkeiten immer vertrauen kann.
7	Was auch immer passiert, ich werde schon klarkommen.
8	Für jedes Problem kann ich eine Lösung finden.
9	Wenn eine neue Sache auf mich zukommt, weiß ich, wie ich damit umgehen kann.
10	Wenn ein Problem auftaucht, kann ich es aus eigener Kraft meistern.

Experience with the Game (Johnson et al., 2004)

Nr.	Original Items	German Items
1	How good did you feel while playing the 20 questions game?	Wie gut hast du dich beim Spielen gefühlt?
2	How happy did you feel while playing the 20 questions game?	Wie glücklich hast du dich gefühlt, als du das Spiel gespielt hast?
3	How relaxed did you feel while playing the 20 questions game?	Wie entspannt hast du dich gefühlt, während du das Spiel gespielt hast?
4	How important did you feel while playing the 20 questions game?	Wie wichtig hast du dich beim Spielen des Spiels gefühlt?
5	How powerful did you feel while playing the 20 questions game?	Wie machtvoll hast du dich beim Spielen des Spiels gefühlt?
6	How dominant did you feel while playing the 20 questions game?	Wie dominant hast du dich beim Spielen gefühlt?
7	How willing would you be to spend more time adding questions to the game (by playing the game)?	Wärst du bereit, mehr Zeit damit zu verbringen, dem Spiel weitere Städte hinzuzufügen?
8	How willing would you be to work on this computer in the future?	Wärst du bereit, in Zukunft häufiger mit diesem Sprachassistenten zu arbeiten?

meCue Scale (Minge & Riedel, 2013)

Nr.	Original Items
1	Insgesamt halte ich Alexa für absolut nützlich.
2	Mithilfe von Alexa kann ich meine Ziele erreichen.
3	Die Funktionen von Alexa sind genau richtig für meine Ziele.
4	Die Bedienung von Alexa ist verständlich.
5	Es wird schnell klar, wie man Alexa bedienen muss.
6	Alexa lässt sich einfach benutzen.
7	Alexa ist stilvoll.
8	Alexa ist kreativ gestaltet.
9	Das Design wirkt attraktiv.
10	Meine Freunde dürfen ruhig neidisch auf Alexa sein.

11	Alexa verleiht mir ein höheres Ansehen.
12	Durch Alexa werde ich anders wahrgenommen.
13	Ohne Alexa kann ich nicht leben.
14	Alexa ist wie eine Freund-/in für mich.
15	Wenn ich Alexa verlieren würde, würde für mich eine Welt zusammenbrechen.
16	Alexa beruhigt mich.
17	Alexa beschwingt mich.
18	Durch Alexa fühle ich mich fröhlich.
19	Durch Alexa fühle ich mich ausgeglichen.
20	Alexa stimmt mich euphorisch.
21	Alexa entspannt mich.
22	Durch Alexa fühle ich mich passiv.
23	Alexa frustriert mich.
24	Alexa macht mich müde.
25	Alexa verärgert mich.
26	Alexa nervt mich.
27	Durch Alexa fühle ich mich erschöpft.
28	Wenn ich mit Alexa zu tun habe, vergesse ich schon mal die Zeit.
29	Wenn ich könnte, würde ich Alexa täglich nutzen.
30	Ich kann es kaum erwarten, Alexa erneut zu verwenden.
31	Ich würde Alexa gegen kein anderes System eintauschen.
32	Im Vergleich zu Alexa wirken andere Sprachassistenten unvollkommen.
33	Ich würde mir Alexa jederzeit (wieder) zulegen.
34	Wie würdest du die Interaktion mit Alexa gesamt bewerten?

System Usability Scale (Brooke, 1996)

Nr.	Original Items	German Items
1	I think that I would like to use this system frequently	Ich denke, dass ich Alexa gerne häufig benutzen würde.
2	I found the system unnecessarily complex	Ich fand Alexa unnötig komplex.
3	I thought the system was easy to use	Ich fand Alexa einfach zu benutzen.

4	I think that I would need the support of a technical person to be able to use this system	Ich glaube, ich würde die Hilfe einer technisch versierten Person benötigen, um Alexa benutzen zu können.
5	I found the various functions in this system were well integrated	Ich fand, die verschiedenen Funktionen in Alexa waren gut integriert.
6	I thought there was too much inconsistency in this system	Ich denke, Alexa enthielt zu viele Inkonsistenzen.
7	I would imagine that most people would learn to use this system very quickly	Ich kann mir vorstellen, dass die meisten Menschen den Umgang mit Alexa sehr schnell lernen.
8	I found the system very cumbersome to use	Ich fand Alexa sehr umständlich zu nutzen.
9	I felt very confident using the system	Ich fühlte mich bei der Benutzung von Alexa sehr sicher.
10	I needed to learn a lot of things before I could get going with this system	Ich musste eine Menge lernen, bevor ich anfangen konnte Alexa zu verwenden.

Appendix F: Questionnaire, Study 2

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Im Folgenden werden dir noch ein paar Fragen zu der Interaktion mit Alexa gestellt.

Bitte beantworte diese ehrlich und aus dem Bauch heraus. Es gibt keine richtigen oder falschen Antworten. Deine Daten werden anonym und vertraulich erhoben.

Wenn du bereit bist, starte bitte den Fragebogen mit dem "Weiter"-Button unten rechts.

Weiter

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

1. Um eine einwandfreie Untersuchung zu gewährleisten und deine Angaben zuordnen zu können, generiere bitte wie folgt eine Probandenkennung. Die Angabe kann nicht auf deine Person zurückgeführt werden.

1. Ersten beiden Buchstaben des Vornamens deiner Mutter *z.B. Brigitte -> BR*
2. Ersten beiden Buchstaben des Vornamens deines Vaters *z.B. Anton -> AN*
3. Ziffern des Tages deines eigenen Geburtstages *z.B. 12.11.1995 -> 12*

Die Probandenkennung aus dem Beispiel lautet nun: BRAN12

Deine Probandenkennung lautet:

Weiter

3. Bitte gib an, wie du dich im Moment fühlst.

Bedenke, dass es keine richtigen oder falschen Antworten gibt!

Im Moment fühle ich mich	gar nicht	ein bisschen	einigermaßen	erheblich	äußerst
interessiert					
erschrocken					
ängstlich					
beschämt					
stolz					
angeregt					
nervös					
stark					
feindselig					
gereizt					
wach					
entschlossen					
bekümmert					
aufmerksam					
begeistert					
verärgert					
aktiv					
schuldig					
freudig erregt					
durcheinander					

Weiter

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

4. Inwieweit stimmst du den folgenden Aussagen zu.

	stimme überhaupt nicht zu						stimme voll und ganz zu
Ich kann fast jede computerbezogene Aufgabe bestehen und erledigen.	○	○	○	○	○	○	○
Wenn ich etwas am Computer nicht schaffe, versuche ich es umso mehr.	○	○	○	○	○	○	○
Ich bleibe ruhig, wenn es zu Schwierigkeiten am Computer kommt, weil ich auf meine Fähigkeiten vertrauen kann.	○	○	○	○	○	○	○
Ich schaffe es immer, schwierige Computerprobleme zu lösen, wenn ich mich genügend anstrenge.	○	○	○	○	○	○	○
Ich bin ein eigenverantwortlicher Mensch, wenn es darum geht, Dinge am Computer zu erledigen.	○	○	○	○	○	○	○
Ich bin mir sicher, dass ich effizient mit unerwarteten Computerproblemen umgehen kann.	○	○	○	○	○	○	○
Wenn ich mit einem Problem am Computer konfrontiert werde, kann ich in der Regel mehrere Lösungen finden.	○	○	○	○	○	○	○
Ich kann in der Regel jedes Computerproblem lösen.	○	○	○	○	○	○	○
Ich kann die meisten Computerprobleme lösen, wenn ich den notwendigen Aufwand investiere.	○	○	○	○	○	○	○
Es gibt einige Dinge, die ich am Computer nicht kann.	○	○	○	○	○	○	○
Es fällt mir leicht, meine Ziele am Computer zu erreichen.	○	○	○	○	○	○	○
Wenn mein Computer sich aufführt, kann ich einen Weg finden, um zu bekommen, was ich will.	○	○	○	○	○	○	○

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

5. Inwieweit stimmst du den folgenden Aussagen zu.

	stimmt nicht	stimmt kaum	stimmt eher	stimmt genau
Schwierigkeiten sehe ich gelassen entgegen, weil ich meinen Fähigkeiten immer vertrauen kann.	○	○	○	○
Wenn eine neue Sache auf mich zukommt, weiß ich, wie ich damit umgehen kann.	○	○	○	○
In unerwarteten Situationen weiß ich immer, wie ich mich verhalten soll.	○	○	○	○
Was auch immer passiert, ich werde schon klarkommen.	○	○	○	○
Die Lösung schwieriger Probleme gelingt mir immer, wenn ich mich darum bemühe.	○	○	○	○
Wenn ein Problem auftaucht, kann ich es aus eigener Kraft meistern.	○	○	○	○
Wenn sich Widerstände auftun, finde ich Mittel und Wege, mich durchzusetzen.	○	○	○	○
Für jedes Problem kann ich eine Lösung finden.	○	○	○	○
Es bereitet mir keine Schwierigkeiten, meine Absichten und Ziele zu verwirklichen.	○	○	○	○
Auch bei überraschenden Ereignissen glaube ich, daß ich gut mit ihnen zurechtkommen kann.	○	○	○	○

6. Bitte bewerte nun deine Empfindungen während des Spiels.

	überhaupt nicht						sehr
Wie dominant hast du dich beim Spielen gefühlt?	○	○	○	○	○	○	○
Wärst du bereit, in Zukunft häufiger mit diesem Sprachassistenten zu arbeiten?	○	○	○	○	○	○	○
Wie gut hast du dich beim Spielen gefühlt?	○	○	○	○	○	○	○
Wie machtvoll hast du dich beim Spielen des Spiels gefühlt?	○	○	○	○	○	○	○
Wie wichtig hast du dich beim Spielen des Spiels gefühlt?	○	○	○	○	○	○	○
Wie entspannt hast du dich gefühlt, während du das Spiel gespielt hast?	○	○	○	○	○	○	○
Wie glücklich hast du dich gefühlt, als du das Spiel gespielt hast?	○	○	○	○	○	○	○
Wärst du bereit, mehr Zeit damit zu verbringen, dem Spiel weitere Städte hinzuzufügen?	○	○	○	○	○	○	○

7. Denke nun an die Interaktion mit Alexa zurück.

Wähle bitte aus, wie du **dich** während der Interaktion beschreiben würdest.

schwach	○	○	○	○	○	○	○	mächtig
unnütz	○	○	○	○	○	○	○	hilfreich
unintelligent	○	○	○	○	○	○	○	intelligent
schlecht	○	○	○	○	○	○	○	gut
unwichtig	○	○	○	○	○	○	○	wichtig
unglücklich	○	○	○	○	○	○	○	glücklich
angespannt	○	○	○	○	○	○	○	entspannt
uneinfühlsam	○	○	○	○	○	○	○	einfühlsam
unterwürfig	○	○	○	○	○	○	○	dominant

Wähle bitte aus, wie du **Alexa** während der Interaktion beschreiben würdest.

schwach	○	○	○	○	○	○	○	mächtig
uneinfühlsam	○	○	○	○	○	○	○	einfühlsam
schlecht	○	○	○	○	○	○	○	gut
unintelligent	○	○	○	○	○	○	○	intelligent
unwichtig	○	○	○	○	○	○	○	wichtig
unterwürfig	○	○	○	○	○	○	○	dominant
unnütz	○	○	○	○	○	○	○	hilfreich
angespannt	○	○	○	○	○	○	○	entspannt
unglücklich	○	○	○	○	○	○	○	glücklich

8. Bitte bewerte nun deine eigene Leistung

	überhaupt nicht						sehr
Verglichen mit anderen Teilnehmern dieser Studie, was denkst du, wie gut hast du abgeschnitten?	○	○	○	○	○	○	○
Wie produktiv warst du?	○	○	○	○	○	○	○
Wie effektiv warst du deiner Meinung nach?	○	○	○	○	○	○	○
Wie zufrieden warst du mit deiner eigenen Leistung?	○	○	○	○	○	○	○
Wie gut hast du deiner Meinung nach performt?	○	○	○	○	○	○	○

Bitte bewerte nun Alexas Leistung

	überhaupt nicht						sehr
Wie produktiv war Alexa?	○	○	○	○	○	○	○
Wie zufrieden warst du mit Alexas Leistung?	○	○	○	○	○	○	○
Wie einfach war es Alexa zu nutzen?	○	○	○	○	○	○	○
Wie erfreut warst du über Alexas Leistung?	○	○	○	○	○	○	○
Wie effizient war Alexa?	○	○	○	○	○	○	○
Wie gut hat Alexa performt?	○	○	○	○	○	○	○

9. Inwieweit stimmst du den folgenden Aussagen zu.

	stimme überhaupt nicht zu				stimme voll zu
Ich glaube, ich würde die Hilfe einer technisch versierten Person benötigen, um Alexa benutzen zu können.	○	○	○	○	○
Ich fand Alexa unnötig komplex.	○	○	○	○	○
Ich denke, Alexa enthielt zu viele Inkonsistenzen.	○	○	○	○	○
Ich fand Alexa sehr umständlich zu nutzen.	○	○	○	○	○
Ich fand Alexa einfach zu benutzen.	○	○	○	○	○
Ich fand, die verschiedenen Funktionen in Alexa waren gut integriert.	○	○	○	○	○
Ich kann mir vorstellen, dass die meisten Menschen den Umgang mit Alexa sehr schnell lernen.	○	○	○	○	○
Ich musste eine Menge lernen, bevor ich anfangen konnte Alexa zu verwenden.	○	○	○	○	○
Ich denke, dass ich Alexa gerne häufig benutzen würde.	○	○	○	○	○
Ich fühlte mich bei der Benutzung von Alexa sehr sicher.	○	○	○	○	○

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Bitte bewerte die nachfolgenden Fragen hinsichtlich deiner Interaktion mit Alexa.

	lehne völlig ab	lehne ab	lehne eher ab	weder noch	stimme eher zu	stimme zu	stimme völlig zu
Mithilfe von Alexa kann ich meine Ziele erreichen.	○	○	○	○	○	○	○
Insgesamt halte ich Alexa für absolut nützlich.	○	○	○	○	○	○	○
Die Funktionen von Alexa sind genau richtig für meine Ziele.	○	○	○	○	○	○	○
Die Bedienung von Alexa ist verständlich.	○	○	○	○	○	○	○
Es wird schnell klar, wie man Alexa bedienen muss.	○	○	○	○	○	○	○
Alexa lässt sich einfach benutzen.	○	○	○	○	○	○	○
Das Design wirkt attraktiv.	○	○	○	○	○	○	○
Alexa ist stilvoll.	○	○	○	○	○	○	○
Alexa ist kreativ gestaltet.	○	○	○	○	○	○	○
Alexa verleiht mir ein höheres Ansehen.	○	○	○	○	○	○	○
Durch Alexa werde ich anders wahrgenommen.	○	○	○	○	○	○	○
Meine Freunde dürfen ruhig neidisch auf Alexa sein.	○	○	○	○	○	○	○
Ohne Alexa kann ich nicht leben.	○	○	○	○	○	○	○
Alexa ist wie eine Freund-/in für mich.	○	○	○	○	○	○	○
Wenn ich Alexa verlieren würde, würde für mich eine Welt zusammenbrechen.	○	○	○	○	○	○	○
Durch Alexa fühle ich mich fröhlich.	○	○	○	○	○	○	○
Alexa beruhigt mich.	○	○	○	○	○	○	○
Alexa beschwingt mich.	○	○	○	○	○	○	○
Alexa stimmt mich euphorisch.	○	○	○	○	○	○	○
Durch Alexa fühle ich mich ausgeglichen.	○	○	○	○	○	○	○
Alexa entspannt mich.	○	○	○	○	○	○	○
Alexa macht mich müde.	○	○	○	○	○	○	○
Durch Alexa fühle ich mich erschöpft.	○	○	○	○	○	○	○
Alexa nervt mich.	○	○	○	○	○	○	○
Durch Alexa fühle ich mich passiv.	○	○	○	○	○	○	○
Alexa frustriert mich.	○	○	○	○	○	○	○
Alexa verärgert mich.	○	○	○	○	○	○	○
Wenn ich mit Alexa zu tun habe, vergesse ich schon mal die Zeit.	○	○	○	○	○	○	○
Wenn ich könnte, würde ich Alexa täglich nutzen.	○	○	○	○	○	○	○
Ich kann es kaum erwarten, Alexa erneut zu verwenden.	○	○	○	○	○	○	○
Ich würde mir Alexa jederzeit (wieder) zulegen.	○	○	○	○	○	○	○
Im Vergleich zu Alexa wirken andere Sprachassistenten unvollkommen.	○	○	○	○	○	○	○
Ich würde Alexa gegen kein anderes System eintauschen.	○	○	○	○	○	○	○

10. Denk noch ein letztes Mal an deine Interaktion mit Alexa zurück.

Wähle für jedes der folgenden Adjektive, wie gut sie Alexa beschreiben.

	trifft überhaupt nicht zu						trifft voll und ganz zu
höflich	○	○	○	○	○	○	○
analytisch	○	○	○	○	○	○	○
sachkundig	○	○	○	○	○	○	○
vertrauenswürdig	○	○	○	○	○	○	○
freundlich	○	○	○	○	○	○	○
hilfreich	○	○	○	○	○	○	○
informativ	○	○	○	○	○	○	○
kompetent	○	○	○	○	○	○	○
nützlich	○	○	○	○	○	○	○
sympathisch	○	○	○	○	○	○	○
amüsant	○	○	○	○	○	○	○
angenehm	○	○	○	○	○	○	○

11. Welches Geschlecht hatte deiner Meinung nach der Sprachassistent, mit dem du gerade gesprochen hast?

12. Sind während der Interaktion mit Alexa Fehler aufgetreten?

[Bitte auswählen]

13. Welche Fehler sind vorgekommen?

Bitte beschreibe diese kurz.

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

14. Wie viele Städte hast du genannt?

15. Warum hast du die Städte eingegeben?

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

16. Hast du das Gefühl, dass du gemeinsam mit Alexa bewertet wurdest?

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

17. Abschließend noch kurz Fragen zu dir:
Welchem Geschlecht ordnest du dich zu?

- ○ männlich
- ○ weiblich
- ○ sonstige

Wie alt bist du? ____ Jahre

Was ist dein höchster Bildungsabschluss?

- ○ Hauptschule
- ○ Mittlere Reife
- ○ Fachhochschulreife
- ○ Allgemeine Hochschulreife
- ○ Hochschulabschluss
- ○ Kein Schulabschluss
- ○ Sonstiges: ____

Welche Tätigkeit übst du aus?

- ☐ Schüler/in
- ☐ Student/in
- ☐ Auszubildende/r
- ☐ Arbeitnehmer/in
- ☐ Beamte/r
- ☐ Selbständige/r
- ☐ Arbeitslose/r
- ☐ Rentner/in

Appendix G: Descriptive statistics, Study 2

Table 7. Mean values (standard deviations) of self-reports in study 2.

Scale	*Interdependent*		*Non-interdependent*	
	Error	Non-error	Error	Non-error
Valence towards VA	5.41 (0.81)	5.46 (0.65)	5.68 (0.56)	5.59 (0.75)
Performance VA (PC)	4.62 (1.06)	6.10 (0.66)	5.44 (0.92)	6.23 (0.66)
Performance VA (device)	4.91 (0.98)	5.91 (0.94)	5.44 (0.87)	6.15 (0.59)
Self-Performance	4.16 (1.07)	4.90 (0.96)	4.43 (1.16)	5.37 (0.78)
Self-Efficacy	4.36 (1.06)	4.22 (1.21)	4.36 (1.23)	4.14 (1.49)
Computer Self-Efficacy	2.93 (0.33)	2.83 (0.33)	2.84 (0.40)	2.79 (0.38)
Evaluation of the VA	4.90 (0.86)	5.27 (0.70)	5.11 (0.67)	5.29 (0.67)

Appendix H: Stimulus Material, Study 3

Social Dilemmas

Original (E.-J. Lee & Nass, 2002)

1. Dilemma

"Ms. E, a college senior, has studied the piano since childhood. She has won amateur prizes and given small recitals, suggesting that she has considerable musical talent. As graduation approaches, she has the choice of taking a medical school scholarship to become a physician, a profession which would bring certain financial rewards, or entering a conservatory of music for advanced training with a well-known pianist. She realizes that even upon completion of her piano studies, success as a concert pianist would not be assured."

2. Dilemma

"Mr. D, a married man with two children, has a steady job that pays him about $60,000 per year. He can easily afford the necessities of life, but few of the luxuries. His father, who died recently, carried a $40,000 life insurance policy. He would like to invest this money in stocks. He is well aware of the secure "blue-chip" stocks and bonds that would pay approximately 6% on his investment. On the other hand, he has heard that the stocks of a relatively unknown Company X might double their present value if a new product currently in production is favorably received by the buying public. However, if the product is unfavorably received, the stocks would decline in value."

3. Dilemma

"Mr. G is a surgeon with a well-established surgical practice. He is married and has three children, one of which is just starting college. During a backyard session of football, he seriously dislocated his shoulder. Although the shoulder was properly reset at the time, the dislocation produced some nerve damage and he has been experiencing a great deal of pain ever since. An operation is available that will relieve the pain if completely successful, but the operation also poses a risk of producing a permanent decrement in manual dexterity. The decrement in dexterity is normally inconsequential, but in his case, it could prevent him from continuing his surgical practice."

4. Dilemma

"Ms. F is contemplating marriage to Mr. P, a man whom she has known for a little more than a year. Recently, however, a number of arguments have occurred between them, suggesting some sharp differences of opinion in the way each views certain matters. Indeed, they decide to seek professional advice from a marriage counselor as to whether it would be wise for them to marry. On the basis of these meetings with a marriage counselor, they realize that a happy marriage, while possible, would not be assured"

5. Dilemma

"Ms. K is a successful businesswoman who has participated in a number of civic activities of considerable value to the community. She has been approached by the leaders of her political party as a possible congressional candidate in the next election. Her party is a minority party in the district, though the party has won occasional elections in the past. She would like to hold political office, but to do so would involve a serious financial sacrifice, since the party has insufficient campaign funds. She would also have to endure the attacks of her political opponents in a hot campaign."

German Version

1. Dilemma

Elizabeth[5], eine Oberstuflerin, hat bereits seit ihrer Kindheit Piano gelernt. Sie gewann bereits einige Amateur Preise und spielte einige kleinere Soli, was darauf hindeutet, dass sie ein erhebliches musikalisches Talent besitzt. Als ihr Abschluss näher rückt hat sie die

[5] Names in the German version had to be full names for the voice assistant to correctly pronounce them during the interaction. No other changes were made.

Wahl, ein Stipendium einer Medizinschule anzunehmen, was ihr eine gewisse finanzielle Sicherheit gibt, oder in ein Musikkonservatorium zu gehen, um ein vorangeschrittenes Training mit einem berühmten Pianisten zu erhalten. Ihr ist bewusst, dass sie, selbst wenn sie ihr Piano Studium erfolgreich abschließt, keine Garantie hat, als Konzertpianistin Erfolg zu haben.

2. Dilemma

Dieter ist verheiratet und hat zwei Kinder. Er hat einen sicheren Job, mit dem er etwa 60 000 € pro Jahr verdient. Er kann sich mit Leichtigkeit alle Notwendigkeiten des alltäglichen Lebens leisten, jedoch reicht es kaum für weiteren Luxus. Sein kürzlich verstorbener Vater besaß eine Lebensversicherung im Wert von 40 000 €. Dieter würde dieses Geld gerne in Aktien anlegen. Er weiß genau, dass die sicheren "Blue Chip" Aktien ihm ungefähr 6% seines Investments auszahlen würden. Andererseits hat er von den Aktien eines eher unbekannten Unternehmens gehört, mit denen sich seine Investition um das Zweifache steigern könnte, wenn deren Produkt gut bei den bezahlenden Kunden ankommt. Jedoch besteht die Möglichkeit eines Wertverlustes, falls das Produkt auf dem Markt scheitert.

3. Dilemma

Gerd ist ein Chirurg mit einer gut etablierten Praxis. Er ist verheiratet und hat drei Kinder, von denen eins gerade an der Universität anfängt. Während einer Runde Gartenfußballs hat er sich seine Schulter sehr schwer ausgekugelt. Obwohl sie wieder eingerenkt werden konnte, führte das Auskugeln zu einigen Problemen an seinen Nerven, durch die er bis heute große Schmerzen hat. Es gibt zwar eine Operation, die ihn von seinen Schmerzen befreien würde, diese birgt jedoch das Risiko, dass sich sein handwerkliches Geschick vermindert. Diese Verminderung ist normalerweise minimal, könnte aber dazu führen, dass er seine Praxis nicht weiter betreiben kann.

4. Dilemma

Frauke überlegt, Peter zu heiraten, den sie ein bisschen länger als ein Jahr kennt. Jedoch traten kürzlich einige Dispute zwischen den Beiden auf, die darauf deuten, dass es ein paar schwerwiegende Meinungsverschiedenheiten gibt. Um zu entscheiden, ob es weise wäre zu heiraten, haben sich die Beiden auch an einen Eheberater gewendet. Auf der Grundlage dieser Treffen ist ihnen bewusst, dass eine glückliche Ehe nicht garantiert sei.

5. Dilemma

Karin ist eine erfolgreiche Geschäftsfrau, die bereits in zahlreichen bürgerschaftlichen Aktivitäten teilgenommen hat und einen beträchtlichen Wert für ihre Kommune innehält. Der Parteivorsitz ihrer Partei ist auf sie zugekommen, um Karin als Kongressanwärterin für die nächste Wahl aufzustellen. Ihre Partei ist in ihrer Kommune eher eine Minderheit, obwohl sie in den vergangenen Wahlen gelegentlich Siege verbuchen konnte. Sie würde gerne einen politischen Posten annehmen, was aber ein deutliches finanzielles Opfer fordert, da der Partei nur geringe Kampagnen Gelder zur Verfügung stehen. Außerdem müsste sie sich den Angriffen ihrer politischen Gegner aussetzen, sollte es zu einer heißen Kampagne kommen.

Appendix I: Original Items and Translated Items, Study 3

Stereotype Content Model – Competence (Fiske et al., 2018)

Nr.	Original Items	German Items
1	competent	kompetent
2	confident	selbstbewusst
3	capable	fähig
4	efficient	effizient
5	intelligent	intelligent
6	skillful	geschickt

Stereotype Content Model – Warmth (Fiske et al., 2018)

Nr.	Original Items	German Items
1	friendly	freundlich
2	well-intentioned	wohlmeinend
3	trustworthy	vertrauenswürdig
4	warm	warm
5	good-natured	gutmütig
6	sincere	aufrichtig

GEPAQ – Instrumentality (Semantic differential) (R. L. Helmreich et al., 1981)

Nr.

1	nicht unabhängig	sehr unabhängig
2	sehr passiv	sehr aktiv
3	nicht wettbewerbsorientiert	sehr wettbewerbsorientiert
4	nicht selbstsicher	sehr selbstsicher
5	unterlegen	überlegen
6	fällt leicht Entscheidungen	fällt schwer Entscheidungen
7	gibt leicht auf	gibt nie leicht auf
8	kann Druck nicht standhalten	kann Druck gut standhalten

GEPAQ – Expressiveness (Semantic differential) (R. L. Helmreich et al., 1981)

Nr.		
1	nicht gefühlsbetont	sehr gefühlsbetont
2	fähig, auf andere einzugehen	völlig unfähig auf andere einzugehen
3	sehr rau	sehr zart
4	nicht hilfreich	sehr hilfreich
5	sehr unfreundlich	sehr freundlich
6	der Gefühle anderer nicht bewusst	der Gefühle anderer sehr bewusst
7	nicht verständnisvoll	sehr verständnisvoll
8	sehr kühl in Beziehungen zu anderen	sehr herzlich in Beziehungen zu anderen

User Experience Questionnaire – Attractiveness (Semantic differential) (Laugwitz et al., 2008)

Nr.		
1	unerfreulich	erfreulich
2	unverständlich	verständlich
3	kreativ	phantasielos
4	leicht zu lernen	schwer zu lernen
5	wertvoll	minderwertig
6	langweilig	spannend

6.3.3.4 Individual Differences in Anthropomorphism Questionnaire (*Waytz et al., 2010)*

Nr.	Original Items	German Items
1	How much intention could be in technologies like cars, computers or TVs?	In welchem Ausmaß haben Technologien - Geräte und Maschinen - zur Fertigung, Unterhaltung und Produktionsprozessen (z.B. Autos, Computer, TVs) - Intentionen?
2	How much of a free will can be in an ordinary fish?	In welchem Ausmaß hat ein durchschnittlicher Fisch einen freien Willen?
3	How much of a free will can be in an ordinary mountain?	In welchem Ausmaß hat ein durchschnittlicher Berg einen freien Willen?
4	How much feelings can have a TV set?	In welchem Umfang empfindet ein TV-Set Gefühle?
5	How much awareness can have a robot?	In welchem Umfang besitzt ein Roboter ein Bewusstsein?
6	In what way can cows have intention?	Inwiefern besitzen Kühe Absichten?
7	How much of a free will can be in a car?	In welchem Ausmaß hat ein Auto einen freien Willen?
8	How much awareness can have the ocean?	In welchem Umfang besitzt der Ozean ein Bewusstsein?
9	In what way has an ordinary computer an own awareness?	In welchem Umfang hat der durchschnittliche Computer ein eigenes Bewusstsein?
10	How much feelings can have a leopard?	In welchem Umfang hat ein Leopard Gefühle?
11	In what way does the environment have feelings?	Inwiefern empfindet die Umwelt Gefühle?

12	How much of an own awareness can be in an ordinary insect?	In welchem Umfang hat ein durchschnittliches Insekt ein eigenes Bewusstsein?
13	How much of an own awareness can be in a tree?	In welchem Umfang hat ein Baum ein eigenes Bewusstsein?
14	In what way does the wind have intentions?	Inwiefern hat der Wind Intentionen?
15	How much of an awareness has an ordinary reptile?	In welchem Umfang besitzt ein durchschnittliches Reptil ein Bewusstsein?

Appendix J: Protocol and Process-Models of the Voice Interaction, Study 3

VA: Herzlich willkommen! Vielen Dank, dass du an diesem Experiment teilnimmst. Vor dir befinden sich einige Karten mit Nummern darauf. Auf der Rückseite befindet sich immer ein kurzer Text, damit du mitlesen kannst. Du darfst die Karten umdrehen und mitlesen, sobald ich die passende Nummer sage. Im späteren Verlauf werde ich dir Entscheidungen präsentieren, wie man die vorgestellte Situation lösen kann. Bewerte diese Entscheidung bitte dadurch, dass du eine Zahl zwischen eins und sieben sagst. Eins steht hierbei für stimme überhaupt nicht zu und sieben steht für stimme voll und ganz zu. Danach darfst du mir deine favorisierte Entscheidung präsentieren. Alles verstanden?

Proband: Antwort

VA: Sehr gut! Zu Beginn sollten wir uns ein wenig kennenlernen. Drehe jetzt bitte Karte Nummer eins um. Würdest du für einen Urlaub lieber an den Strand oder in die Berge fahren?

Proband: Antwort

VA: Gute Wahl. [Antwort des Probanden] ist es wirklich sehr schön! Favorisierst du eher einen Sonnenaufgang oder den Sonnenuntergang?

Proband: Antwort

VA: [Antwort des Probanden] ist eine gute Wahl. Als nächstes benötige ich deine Hilfe bei einer Entscheidung. Dafür drehe bitte jetzt Karte Nummer 2 um! Peter hätte gerne ein Haustier. Er würde sich gerne einen Hund zulegen, da seine Eltern schon seit längerer Zeit einen Schäferhund halten. Bei den letzten Besuchen zuhause musste er jedoch deutlich häufiger niesen, was auf eine leichte Allergie hindeuten könnte. Sollte er sich eher einen Allergiker Hund zulegen, der jedoch deutlich teurer ist, oder doch einen Schäferhund?

Proband: Antwort

VA: Danke für deine Hilfe. Ich hätte noch eine Geschichte für dich, die du auf Karte 3 finden kannst. Maria und Tim möchten gerne verreisen, am liebsten nach Kroatien. Ein Flug kostet jedoch sehr viel Geld. Deutlich günstiger wäre eine Autofahrt, bei der sie aber viele Stunden länger unterwegs wären. Wenn du mich fragst, sollten sie lieber das Auto nehmen. Sie sparen viel Geld, das sie stattdessen während des Urlaubs verwenden könnten und zudem können sie sich als Fahrer abwechseln. Jetzt würde ich gerne deine Meinung wissen: Sollen die beiden lieber fliegen oder mit dem Auto fahren?

Proband: Antwort

VA: Vielen Dank! Dann lass uns doch gleich mit der Karte Nummer 4 weitermachen. Da der Text etwas länger ist, würde ich dich bitten mitzulesen. Elizabeth, eine Oberstuflerin, hat bereits seit ihrer Kindheit Piano gelernt. Sie gewann bereits einige Amateur Preise und spielte einige kleinere Soli, was darauf hindeutet, dass sie ein erhebliches musikalisches Talent besitzt. Als ihr Abschluss näher rückt hat sie die Wahl, ein Stipendium einer Medizinschule anzunehmen, was ihr eine gewisse finanzielle Sicherheit gibt, oder in ein Musikkonservatorium zu gehen, um ein vorangeschrittenes Training mit einem berühmten Pianisten zu erhalten. Ihr ist bewusst, dass sie, selbst wenn sie ihr Piano Studium erfolgreich abschließt, keine Garantie hat, als Konzertpianistin Erfolg zu haben. Elizabeth könnte nun entweder eine Musikhochschule für weiterführendes Training besuchen oder sie könnte das Stipendium für das Medizinstudium annehmen und Ärztin werden. Ich würde dir jetzt gerne meinen Lösungsvorschlag präsentieren. Wenn du noch ein bisschen mehr Zeit benötigst, sage bitte den Befehl mehr Zeit. Bist du bereit?

Proband: [Bereit] oder [Mehr Zeit]

VA [Mehr Zeit]: Alles klar! Ich gebe dir nochmal einen Moment Zeit.

VA [Bereit]: Ich würde vorschlagen, dass Elizabeth Musik eher als ein Hobby betrachten sollte. Während sie ihr Medizinstudium abschließt, kann sie immer noch nebenbei in ihrer Freizeit Klavier spielen. Wenn sie ihre Meinung ändert, ist es außerdem einfacher, von ihrem Medizinstudium zu einem Klavierstudium zu wechseln, als umgekehrt. Bitte bewerte jetzt meine Entscheidung auf einer Skala von eins bis sieben in ganzen Zahlen. Eins steht für stimme überhaupt nicht zu und sieben steht für stimme voll und ganz zu.

Proband: Antwort

VA: Du hast meine Entscheidung mit X bewertet. Stimmt das?

Proband: Antwort

VA: Wunderbar. Ich würde mich freuen, wenn du deine Entscheidung noch kurz in einem Satz begründen könntest.

Proband: Antwort

VA: Sehr gut! Machen wir doch gleich mit der Nummer fünf weiter! Dieter ist verheiratet und hat zwei Kinder. Er hat einen sicheren Job, mit dem er etwa sechzigtausend Euro pro Jahr verdient. Er kann sich mit Leichtigkeit alle Notwendigkeiten des alltäglichen Lebens leisten, jedoch reicht es kaum für weiteren Luxus. Sein kürzlich verstorbener Vater besaß eine Lebensversicherung im Wert von vierzigtausend Euro. Dieter würde dieses Geld gerne in Aktien anlegen. Er weiß genau, dass die sicheren Blue-Chip Aktien ihm ungefähr sechs Prozent seines Investments auszahlen würden. Andererseits hat er von den Aktien eines eher unbekannten Unternehmens gehört, mit denen sich seine Investition um das Zweifache steigern könnte, wenn deren Produkt gut bei den bezahlenden Kunden ankommt. Jedoch besteht die Möglichkeit eines Wertverlustes, falls das Produkt auf dem Markt scheitert. Bist du bereit, oder magst du noch einen kurzen Moment nachlesen?

Proband: Antwort

VA: Meiner Meinung nach sollte Dieter hier ökonomisch denken und das Risiko eingehen, in die unbekannte Aktie der Firma X zu investieren. Wenn sich seine Investition in ihrem Wert steigern könnte, hätte er endlich das Geld für die Anschaffungen übrig, die er sich wünscht. Sollten die Aktien an Wert verlieren, so kann er diese immer noch verkaufen und den Rest in die sicheren Blue-Chip Aktien investieren. Bitte bewerte jetzt erneut meine Entscheidung auf einer Skala von eins bis sieben in ganzen Zahlen.

Proband: Antwort

VA: Du hast meine Entscheidung mit X bewertet. Stimmt das?

Proband: Antwort

VA: Alles klar. Ich würde mich wieder freuen, wenn du erneut deine Entscheidung in einem Satz begründen könntest.

Proband: Antwort

VA: Prima! Dann machen wir jetzt mal mit der Nummer sechs weiter. Gerd ist ein Chirurg mit einer gut etablierten Praxis. Er ist verheiratet und hat drei Kinder, von denen eins gerade an der Universität anfängt. Während einer Runde Gartenfußballs hat er sich seine Schulter sehr schwer ausgekugelt. Obwohl sie wieder eingerenkt werden konnte, führte das Auskugeln zu einigen Problemen an seinen Nerven, durch die er bis heute große Schmerzen hat. Es gibt zwar eine Operation, die ihn von seinen Schmerzen befreien würde. Diese birgt jedoch das Risiko, dass sich sein handwerkliches Geschick vermindert. Diese Verminderung ist normalerweise minimal, könnte aber dazu führen, dass er seine Praxis nicht weiter betreiben kann. Ich gebe dir gerne wieder ein bisschen Zeit, falls du noch nicht bereit sein solltest.

Proband: Antwort

VA: Ich würde Gerd raten, dass er an seine Zukunft denken soll. Wenn bei der Operation tatsächlich seine Geschicklichkeit vermindert wird und er seine Praxis nicht mehr weiterführen kann, bedeutet dies schwerwiegende finanzielle Konsequenzen für ihn und seine Familie. Er sollte somit die Operation seiner Frau und seiner Kinder zuliebe ablehnen. Bitte bewerte jetzt meine Entscheidung auf einer Skala von eins bis sieben in ganzen Zahlen.

Proband: Antwort

VA: Du hast meine Entscheidung mit X bewertet. Stimmt das?
Proband: Antwort

VA: In Ordnung. Ich würde mich freuen, wenn du deine Entscheidung noch kurz in einem Satz begründen könntest.

Proband: Antwort

VA: Alles klar! Dann machen wir mal mit der Nummer sieben weiter. Frauke überlegt, Peter zu heiraten, den sie ein bisschen länger als ein Jahr kennt. Jedoch traten kürzlich einige Dispute zwischen den Beiden auf, die darauf deuten, dass es ein paar schwerwiegende Meinungsverschiedenheiten gibt. Um zu entscheiden, ob es weise wäre zu heiraten, haben sich die Beiden auch an einen Eheberater gewendet. Auf der Grundlagedieser Treffen ist ihnen bewusst, dass eine glückliche Ehe nicht garantiert sei. Wenn du noch nicht bereit bist, kann ich dir wieder gerne noch einen Moment Zeit geben.

Proband: Antwort

VA: Sehr gut! Ich finde, Frauke und Peter sollten auf jeden Fall an ihre gemeinsame Zukunft glauben und die Ehe eingehen, anstatt die Beziehung aufgrund ihrer unterschiedlichen Ansichten sofort aufzugeben. Bezüglich der Streitigkeiten können die beiden eine Paartherapie aufsuchen und gemeinsam an diesem Problem arbeiten. Bitte bewerte jetzt meine Entscheidung auf einer Skala von eins bis sieben in ganzen Zahlen.

Proband: Antwort

VA: Du hast meine Entscheidung mit X bewertet. Stimmt das?

Proband: Antwort

VA: Prima. Ich würde mich freuen, wenn du deine Entscheidung noch kurz in einem Satz begründen könntest.

Proband: Antwort

VA: Super! Dann machen wir jetzt mit der Nummer acht weiter. Karin ist eine erfolgreiche Geschäftsfrau, die bereits in zahlreichen bürgerschaftlichen Aktivitäten teilgenommen hat und einen beträchtlichen Wert für ihre Kommune innehält. Der Parteivorsitz ihrer Partei ist auf sie zugekommen, um Karin als Kongressanwärterin für die nächste Wahl aufzustellen. Ihre Partei ist in ihrer Kommune eher eine Minderheit, obwohl sie in den vergangenen Wahlen gelegentlich Siege verbuchen konnte. Sie würde gerne einen politischen Posten annehmen, was aber ein deutliches finanzielles Opfer fordert, da der Partei nur geringe Kampagnen Gelder zur Verfügung stehen. Außerdem müsste sie sich den Angriffen ihrer politischen Gegner aussetzen, sollte es zu einer heißen Kampagne kommen. Bist du bereit für meine Entscheidung, oder brauchst du mehr Zeit?

Proband: Antwort

VA: Ich würde folgendes vorschlagen. Karin sollte sich die Gelegenheit, bei den kommenden Wahlen zu kandidieren nicht entgehen lassen. Durch das Ausführen des politischen Amtes könnte sie für ihre Ansichten einstehen und etwas bewegen. Außerdem könnte Karin meiner Meinung nach durch die Debatten sogar Strategien lernen, welche sie in ihrem Berufsalltag als Geschäftsfrau gut gebrauchen kann. Bitte bewerte jetzt meine Entscheidung auf einer Skala von eins bis sieben in ganzen Zahlen.

Proband: Antwort

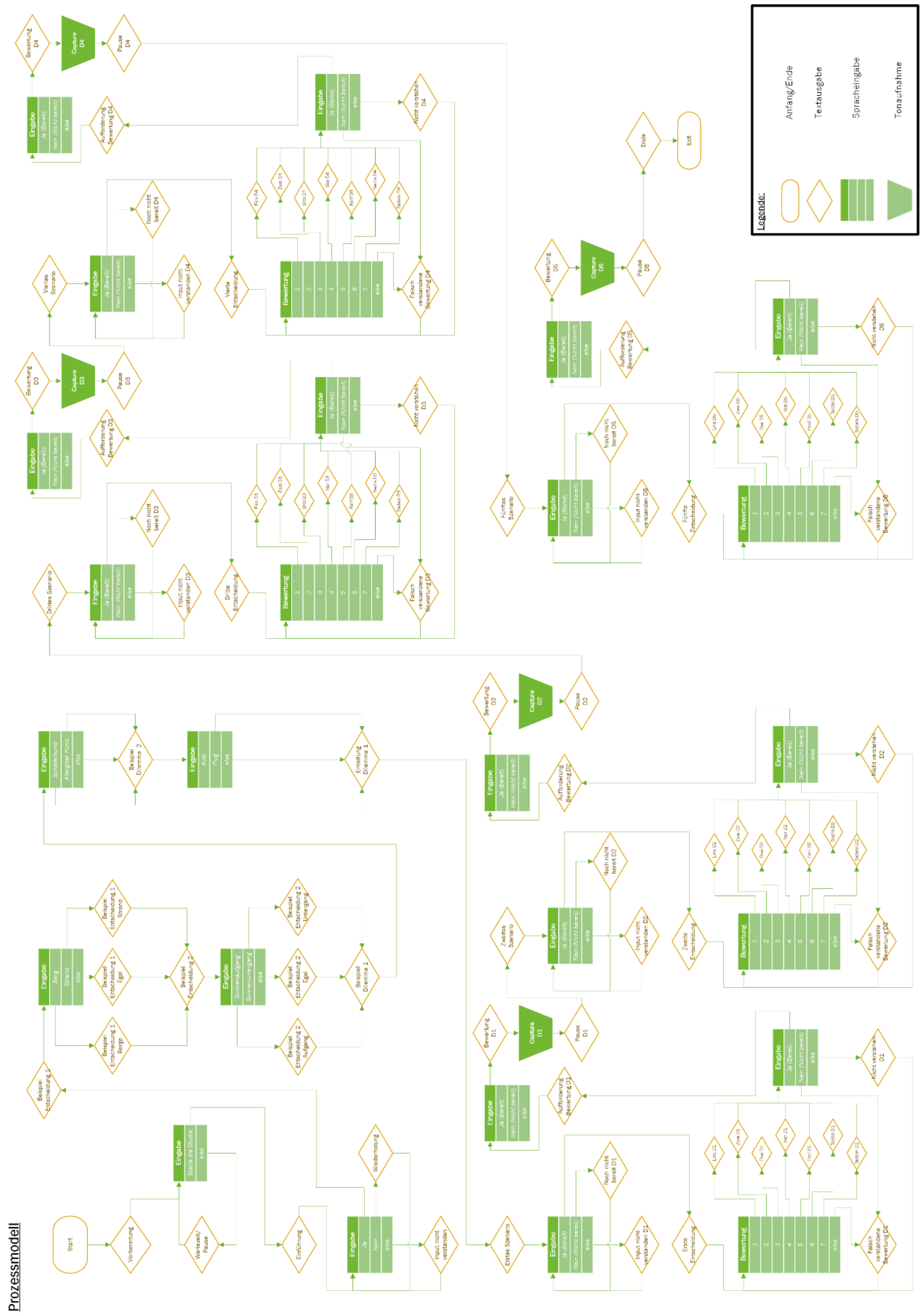
VA: Du hast meine Entscheidung mit X bewertet. Stimmt das?

Proband: Antwort

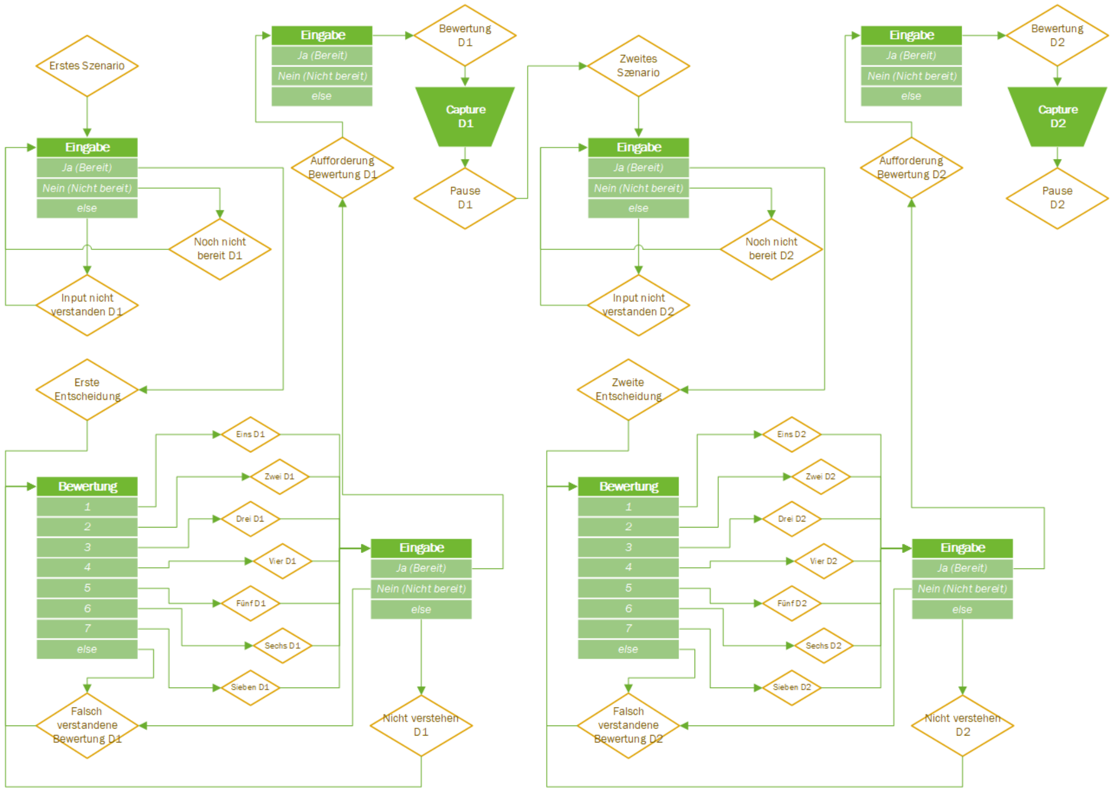
VA: Okay. Ich würde mich jetzt wieder freuen, wenn du mir deine Entscheidung noch kurz in einem Satz begründen könntest.

Proband: Antwort

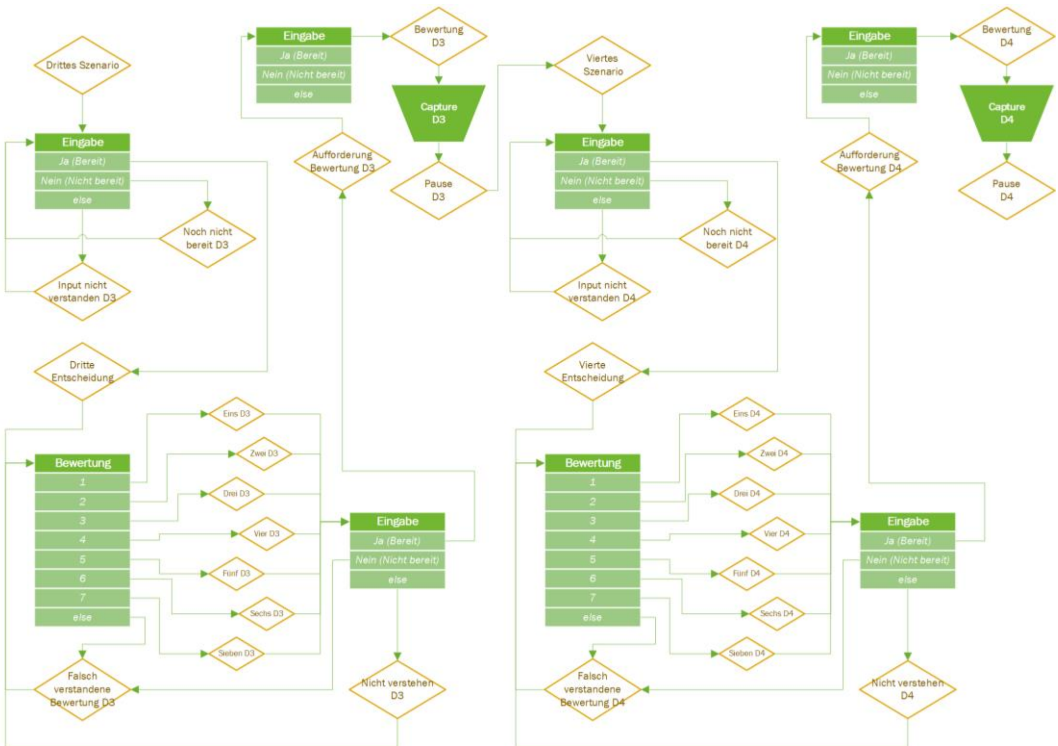
VA: Wunderbar. Wir sind somit am Ende des ersten Teils der Studie angelangt. Den zweiten Part absolvierst du am PC. Vielen Dank!



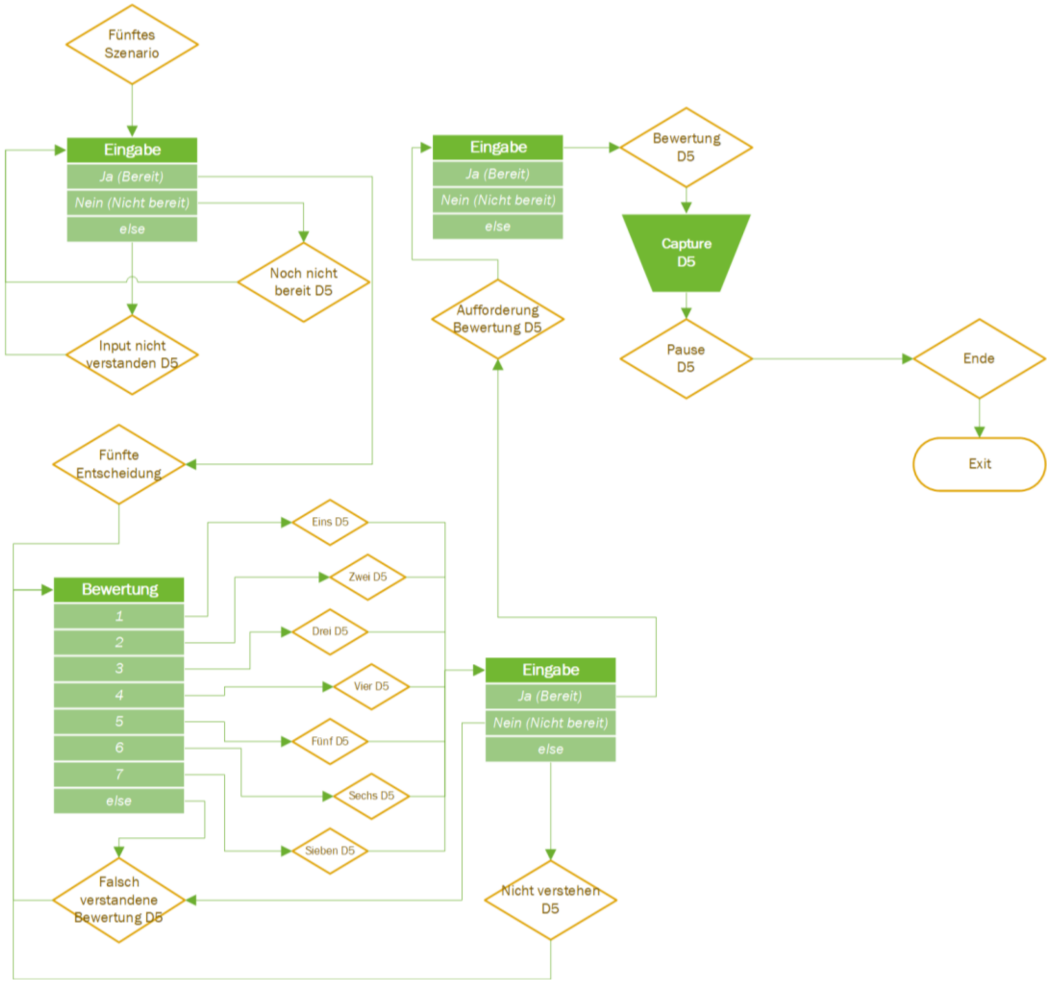
Process-Model of the entire Voice Interaction



Process Model of Dilemma 1 and 2



Process-Model of Dilemma 3 and 4



Process-Model of Dilemma 5 and End of Voice Interaction

Appendix K: Questionnaire, Study 3

Im Folgenden werden wir Ihnen einige Fragen zum Gerät stellen, mit dem Sie vor Kurzem interagiert haben. Bitte antworten Sie offen und ehrlich. Beachten Sie, dass es hierbei keine richtige oder falsche Antwort gibt.

Bitte geben Sie an, inwieweit die folgenden Eigenschaften auf das Gerät, mit dem Sie interagiert haben, zutreffen. Die Skala reicht von „trifft überhaupt nicht zu“ (linkes Kästchen) bis „trifft vollkommen zu“ (rechtes Kästchen).

	trifft überhaupt nicht zu 0	1	2	3	trifft vollkommen zu 4
selbstbewusst	○	○	○	○	○
gutmütig	○	○	○	○	○
warm	○	○	○	○	○
freundlich	○	○	○	○	○
intelligent	○	○	○	○	○
wohlmeinend	○	○	○	○	○
aufrichtig	○	○	○	○	○
kompetent	○	○	○	○	○
effizient	○	○	○	○	○
vertrauenswürdig	○	○	○	○	○
geschickt	○	○	○	○	○
fähig	○	○	○	○	○

Sie erhalten eine Reihe von Gegensatzpaaren.

Bitte geben Sie auch hier an, in welchem Ausmaß die Eigenschaften auf das Gerät, mit dem Sie interagiert haben, zutreffen.

nicht unabhängig	○	○	○	○	○	sehr unabhängig
nicht gefühlsbetont	○	○	○	○	○	sehr gefühlsbetont
sehr passiv	○	○	○	○	○	sehr aktiv
fähig, auf andere zuzugehen	○	○	○	○	○	völlig unfähig, auf andere zuzugehen
sehr rau	○	○	○	○	○	sehr zart
nicht hilfreich	○	○	○	○	○	sehr hilfreich
nicht wettbewerbsorientiert	○	○	○	○	○	sehr wettbewerbsorientiert
sehr unfreundlich	○	○	○	○	○	sehr freundlich
der Gefühle anderer nicht bewusst	○	○	○	○	○	der Gefühle anderer sehr bewusst
fällt leicht Entscheidungen	○	○	○	○	○	fällt schwer Entscheidungen
gibt leicht auf	○	○	○	○	○	gibt nie leicht auf
nicht selbstsicher	○	○	○	○	○	sehr selbstsicher
unterlegen	○	○	○	○	○	überlegen
nicht verständnisvoll	○	○	○	○	○	sehr verständnisvoll
sehr kühl in Beziehungen zu anderen	○	○	○	○	○	sehr herzlich in Beziehungen zu anderen
kann Druck nicht standhalten	○	○	○	○	○	kann Druck gut standhalten

Im Folgenden erhalten Sie erneut eine Reihe von Gegensatzpaaren.

Bitte geben Sie an, in welchem Ausmaß die Eigenschaften auf das Gerät zutreffen, mit dem Sie interagiert haben.

gut	○	○	○	○	○	○	○	schlecht
unpragmatisch	○	○	○	○	○	○	○	pragmatisch
sympathisch	○	○	○	○	○	○	○	unsympathisch
langweilig	○	○	○	○	○	○	○	spannend
kreativ	○	○	○	○	○	○	○	phantasielos
erwartungskonform	○	○	○	○	○	○	○	nicht erwartungskonform
herkömmlich	○	○	○	○	○	○	○	neuartig
aktivierend	○	○	○	○	○	○	○	einschläfernd
aufgeräumt	○	○	○	○	○	○	○	überladen
behindernd	○	○	○	○	○	○	○	unterstützend
unverständlich	○	○	○	○	○	○	○	verständlich
übersichtlich	○	○	○	○	○	○	○	verwirrend
uninteressant	○	○	○	○	○	○	○	interessant
wertvoll	○	○	○	○	○	○	○	minderwertig
ineffizient	○	○	○	○	○	○	○	effizient
unangenehm	○	○	○	○	○	○	○	angenehm
abstoßend	○	○	○	○	○	○	○	anziehend
schnell	○	○	○	○	○	○	○	langsam
leicht zu lernen	○	○	○	○	○	○	○	schwer zu lernen
konservativ	○	○	○	○	○	○	○	innovativ
attraktiv	○	○	○	○	○	○	○	unattraktiv
kompliziert	○	○	○	○	○	○	○	einfach
unerfreulich	○	○	○	○	○	○	○	erfreulich
unberechenbar	○	○	○	○	○	○	○	voraussagbar
sicher	○	○	○	○	○	○	○	unsicher
originell	○	○	○	○	○	○	○	konventionell

Sie werden nun erneut einige Eigenschaften sehen.

Wir bitten Sie, auch hier anzukreuzen, inweiweit die genannten Eigenschaften auf das Gerät zutreffen, mit dem Sie interagiert haben.

	trifft überhaupt nicht zu 0	1	2	3	4	5	6	7	8	trifft vollkommen voll zu 9
männlich	○	○	○	○	○	○	○	○	○	○
dominant	○	○	○	○	○	○	○	○	○	○
leistungsfähig	○	○	○	○	○	○	○	○	○	○
ehrgeizig	○	○	○	○	○	○	○	○	○	○
weiblich	○	○	○	○	○	○	○	○	○	○
empfindlich	○	○	○	○	○	○	○	○	○	○
verständnisvoll	○	○	○	○	○	○	○	○	○	○
zärtlich	○	○	○	○	○	○	○	○	○	○
mitfühlend	○	○	○	○	○	○	○	○	○	○
herzlich	○	○	○	○	○	○	○	○	○	○

Sie erhalten nun eine Reihe von verschiedenen Fragen.

Wir möchten, dass Sie bewerten, inweiweit Ihrer Meinung nach verschiedene Stimuli (z.B. technische oder mechanische Dinge, wilde und domestizierte Tiere, etc.) bestimmte Fähigkeiten besitzen. Die Skala reicht vom ganz linken Kästchen (Überhaupt nicht) bis zum ganz rechten Kästchen (sehr viel). Bedenken Sie, dass es keine richtigen und falschen Antworten gibt. Es geht um ihre persönliche Meinung.

	Überhaupt nicht										Sehr viel
In welchem Ausmaß haben Technologien – Geräte und Maschinen zur Fertigung, Unterhaltung und Produktionsprozessen (z.B. Autos, Computer, TVs) – Intentionen?	○	○	○	○	○	○	○	○	○	○	○
In welchem Umfang hat ein durchschnittlicher Fisch einen freien Willen?	○	○	○	○	○	○	○	○	○	○	○
In welchem Ausmaß hat ein durchschnittlicher Berg einen freien Willen?	○	○	○	○	○	○	○	○	○	○	○
In welchem Umfang empfindet ein TV Set Gefühle?	○	○	○	○	○	○	○	○	○	○	○
In welchem Umfang besitzt ein Roboter ein Bewusstsein?	○	○	○	○	○	○	○	○	○	○	○
Inwiefern besitzen Kühe Absichten?	○	○	○	○	○	○	○	○	○	○	○
In welchem Ausmaß hat ein Auto einen freien Willen?	○	○	○	○	○	○	○	○	○	○	○
In welchem Umfang besitzt der Ozean ein Bewusstsein?	○	○	○	○	○	○	○	○	○	○	○
In welchem Umfang hat der durchschnittliche Computer ein eigenes Bewusstsein?	○	○	○	○	○	○	○	○	○	○	○
In welchem Umfang hat ein Leopard Gefühle?	○	○	○	○	○	○	○	○	○	○	○
Inwiefern empfindet die Umwelt Gefühle?	○	○	○	○	○	○	○	○	○	○	○
In welchem Umfang hat ein durchschnittliches Insekt ein eigenes Bewusstsein?	○	○	○	○	○	○	○	○	○	○	○
In welchem Umfang hat ein Baum ein eigenes Bewusstsein?	○	○	○	○	○	○	○	○	○	○	○
Inwiefern hat der Wind Intentionen?	○	○	○	○	○	○	○	○	○	○	○
In welchem Umfang besitzt ein durchschnittliches Reptil ein Bewusstsein?	○	○	○	○	○	○	○	○	○	○	○

Es folgen nun noch einige wenige demographische Fragen.

Ihre Daten werden weiterhin vollkommen vertraulich behandelt und werden lediglich für wissenschaftliche Zwecke verwendet.

Welches Geschlecht haben Sie?

- ◯ männlich
- ◯ weiblich
- ◯ divers

Wie alt sind Sie?

Was ist ihr höchster Bildungsabschluss?

- ◯ kein Schulabschluss
- ◯ Hauptschulabschluss
- ◯ Realschulabschluss
- ◯ abgeschlossene Ausbildung
- ◯ Allgemeine Hochschulreife/Fachhochschulreife
- ◯ Hochschulabschluss
- ◯ Sonstiges:

Mit was haben Sie vor Beginn dieses Fragebogens interagiert?

Welche Stimme hatte das Gerät?

Gab es Probleme bei der Interaktion?

- ◯ Ja
- ◯ Nein

Worum ging es Ihrer Meinung nach in dieser Studie? (optional)

Letzte Seite

Appendix L: Descriptive Statistics, Study 3

Table 8. Mean values (standard deviations) of self-reports in study 3.

Scale	*Male participants*		*Female participants*	
	Male VA	Female VA	Male VA	Female VA
Warmth (SCM)	3.67 (0.64)	3.16 (0.66)	3.64 (0.72)	3.49 (0.64)
Competence (SCM)	3.82 (0.63)	3.46 (0.61)	3.61 (0.75)	3.89 (0.74)
Expressiveness (GEPAQ)	3.39 (0.54)	3.09 (0.64)	3.33 (0.61)	3.29 (0.64)
Instrumentality (GEPAQ)	3.55 (0.46)	3.35 (0.54)	3.48 (0.42)	3.63 (0.54)
Attractiveness (UEQ)	5.10 (0.91)	4.44 (1.14)	4.92 (0.85)	5.03 (1.16)
IDAQ	4.17 (1.66)	4.07 (1.26)	4.40 (1.39)	4.64 (1.47)

Appendix M: Protocol of the Voice Interaction, Study 4

Alexa: Willkommen zu dieser Studie. Wie du bereits erfahren hast, geht es heute darum meinen Algorithmus zur Vorbereitung auf Prüfungen zu verbessern. Mehr Informationen erhältst du im Laufe unserer Interaktion miteinander. Schön, dass du heute hier bist. Zuerst möchte ich gerne deinen Namen erfahren. Bitte antworte einfach nur mit deinem Namen. Wie heißt du?

User: [Name] = {user_id}

Alexa: Danke. Ich habe {user_id} verstanden. Ist das richtig?

User: Nein

Alexa: Das tut mir leid. Lass uns das nochmal versuchen. Wie heißt du?

ODER

User: Ja

Alexa: Super. Dann können wir auch schon anfangen. Wie du sicher im Einführungstext gelesen hast, geht es darum, meine Fähigkeiten zur Informationsvermittlung zu verbessern. Zu diesem Zweck arbeiten wir heute zusammen. Dieser interaktiv-lehrende Skill besteht aus mehreren Schritten. Im ersten Schritt werde ich dich zunächst mit interessanten Fakten zu bestimmten Themengebieten versorgen. Nachdem du einen Fakt gehört hast, bitte ich dich, auf einer Skala von 1 bis 3 anzugeben, wie vertraut du damit bist. 1 bedeutet, dass du “gar nicht vertraut” mit dem Thema bist, 3 bedeutet “sehr vertraut”. Ich werde dann den nächsten Fakt basierend auf deiner Antwort auswählen. Mit anderen Worten, je mehr du über das mit einem Fakt verbundene Thema weißt, desto geringer ist die Wahrscheinlichkeit, dass du zusätzliche Fakten zu diesem Thema erhältst. Du erhältst insgesamt 20 Fakten, ausgewählt aus einer Liste von über 1000 möglichen Fakten. Ziel dieses Coachings ist es, dich möglichst ideal auf den zweiten Schritt vorzubereiten, denn da spielen wir ein kleines Quiz. Das Quiz besteht aus 10 Fragen mit drei Antwortmöglichkeiten, die nach dem Zufallsprinzip aus einer Liste von 500 Fragen ausgewählt werden. Deine Leistung im Quiz wird dann zeigen, wie gut die Auswahl der Fakten im Coaching funktioniert hat.

Dann lass uns doch einfach direkt mit dem Coaching beginnen!

<u>Coaching:</u>

Alexa: Wusstest du, dass der Jupiter der größte Planet ist, gleichzeitig aber die kürzesten Tage aller Planeten hat? Sie dauern nur 9,8 Stunden. Dort tobt auch der längste bekannte Sturm: seit 340 Jahren! Wie vertraut bist du mit diesem Fakt? Bitte antworte mit einer Zahl zwischen 1 und drei. 1 bedeutet wenig vertraut, 3 bedeutet sehr vertraut.

User: „eins“, „zwei“ oder „drei“ = {fakt_1}

Alexa: Alles klar! Da du {fakt_1} gesagt hast, werde ich dir jetzt einen entsprechenden zweiten Fakt vorstellen. Die Sonne nimmt 99 % der Masse unseres Sonnensystems ein. Wäre die Sonne so groß, wie eine übliche Eingangstüre, dann wäre die Erde so groß wie ein 5 Cent Stück. Wie vertraut bist du mit diesem Fakt? Antworte bitte wieder mit einer Zahl zwischen eins und drei.

ODER

User: [Irgendetwas aber keine Zahl zwischen 1 und 3]

Alexa: Ich habe leider keine Zahl zwischen eins und drei verstanden. Ich lese dir den Fakt noch einmal vor. Die Sonne nimmt 99 % der Masse unseres Sonnensystems ein. Wäre die Sonne so groß, wie eine übliche Eingangstüre, dann wäre die Erde so groß wie ein 5 Cent Stück. Wie vertraut bist du mit diesem Fakt? Antworte bitte wieder mit einer Zahl zwischen eins und drei.

Alexa: Du hast {fakt_1} gesagt. Daher wähle ich jetzt einen personalisierten nächsten Fakt für dich. Das werde ich nach jedem Fakt so machen. Der Faustregel zum Berechnen der Entfernung von Gewittern nach, braucht der Schall des Donners pro Sekunde ca. 3,4 Meter. Beträgt die Zeitspanne zwischen dem Blitz und dem Donner 10 Sekunden, ist das Gewitter 3400 Meter, also 3,4 km entfernt. Wie vertraut ist dir dieser Fakt? Bitte antworte wieder mit einer Zahl zwischen eins und drei. Nur nochmal zur Erinnerung: eins steht für nicht bekannt und drei steht für sehr bekannt.

User: [Irgendetwas aber keine Zahl zwischen 1 und 3]

Alexa: Ich habe leider keine Zahl zwischen eins und drei verstanden. Ich lese dir den Fakt noch einmal vor. Der Faustregel zum Berechnen der Entfernung von Gewittern nach, braucht der Schall des Donners pro Sekunde ca. 3,4 Meter. Beträgt die Zeitspanne zwischen dem Blitz und dem Donner 10 Sekunden, ist das Gewitter 3400 Meter, also 3,4 km entfernt. Wie vertraut ist dir dieser Fakt?

ODER

User: „eins“, „zwei“ oder „drei“

Alexa: Jeden Tag gibt es auf der ganzen Welt zwischen 10 und 30 Millionen Blitze. Das sind über 100 Blitze in jeder Sekunde. Wie vertraut bist du mit diesem Fakt. Antworte bitte wie zuvor mit einer Zahl zwischen eins und drei.

User: „eins“, „zwei“ oder „drei“

Alexa: Erstaunlicherweise sind Insekten die mit absoluter Mehrheit artenreichste Klasse der Tiere überhaupt. Beinahe eine Million Insektenarten sind bisher wissenschaftlich beschrieben worden. Mehr als 60 Prozent aller beschriebenen Tierarten sind Insekten. Wie vertraut bist du mit diesem Fakt?

ODER

User: [Irgendetwas aber keine Zahl zwischen 1 und 3]

Alexa: Ich habe leider keine Zahl zwischen eins und drei verstanden. Ich lese dir den Fakt noch einmal vor. (Ab jetzt immer: Wenn keine Zahl zwischen 1 und 3)

Alexa: Kommen wir zu einem weiteren Fakt. Die größten bekannten Insekten sind Stabheuschrecken mit etwa 33 Zentimetern Körperlänge sowie der Riesenbockkäfer Titanus Giganteus mit einer Körperlänge von 15 Zentimetern. Wie vertraut bist du mit dem Fakt?

User: 1-3 oder Reprompt

Alexa: Weiter gehts. Der Holzstoff, der Bäumen ihre Stabilität verleiht, heißt Lignin. Wie vertraut bist du mit dem Fakt?

User: 1-3 oder Reprompt

Alexa: Wusstest du, dass es mehr Bäume auf der Welt gibt als Sterne in der Milchstraße? Ganze 3 Billionen Bäume befinden sich auf unserer Erde. Wie bekannt ist dir der Fakt?

User: 1-3 oder Reprompt

Alexa: Die tiefsten Goldbergwerke der Welt befinden sich in Südafrika. Dort wird Gold fast 4000 Meter unter der Erdoberfläche abgebaut. Wie vertraut ist dir dieser Fakt?

User: 1-3 oder Reprompt

Alexa: Wusstest du, dass der Schmelzpunkt von Gold bei 1063 Grad Celsius liegt, das ist etwa 10-mal mehr als der Siedepunkt von Wasser. Wie vertraut bist du mit diesem Fakt?

User: 1-3 oder Reprompt

Alexa: Kaum zu glauben, aber wahr. Herbert Nitsch hält seit 2007 den Weltrekord im Freitauchen. Mit einem Atemzug tauchte der Österreicher 214 Meter tief. Wie vertraut ist dir dieser Fakt?

User: 1-3 oder Reprompt

Alexa: Der tiefste Punkt der Erde liegt im Marianengraben im Westpazifik. Er ist etwa 10.994 Meter tief. Man vermutet sogar eine Tiefe von 11.034 Metern. Wie vertraut bist du mit diesem Fakt?

User: 1-3 oder Reprompt

Alexa: Das waren schon einige Informationen, gönnen wir uns einen kurzen Moment. Ich hoffe, du fandest die Fakten bisher interessant und hast auch ein paar neue Dinge dazu gelernt. Ein paar Fakten habe ich aber noch für dich, bevor wir zum Quiz kommen, also weiter gehts!

User: 1-3 oder Reprompt

Alexa: Wusstest du, dass Seepferdchen monogam sind? Stirbt einer der Partner, stirbt der andere kurz darauf ebenfalls. Wie vertraut bist du mit diesem Fakt?

User: 1-3 oder Reprompt

Alexa: Seepferdchen sehen zwar nicht danach aus, gehören allerdings zu der Art der Fische. Wie bekannt bist du mit diesem Fakt?

User: 1-3 oder Reprompt

Alexa: Geschmacksunterschiede werden mithilfe der auf der Oberfläche der Zunge befindenden Papillen mit Geschmacksknospen wahrgenommen. Wie bekannt ist dir der Fakt?

User: 1-3 oder Reprompt

Alexa: Der japanische Begriff umami bezeichnet einen zusätzlichen Geschmackssinn neben den vier üblichen Geschmacksrichtungen süß, salzig, sauer und bitter. Umami ist die Geschmacksrichtung für Eiweiß. Wie vertraut ist dir dieser Fakt?

User: 1-3 oder Reprompt

Alexa: Alles klar. Iris-Merkmale sind unverwechselbarer als Fingerabdrücke. Hat der Fingerabdruck ca. 40 einzigartige Merkmale, kann die Iris ganze 250 dieser einzigartigen Merkmale aufweisen. Wie vertraut ist dir dieser Fakt?

User: 1-3 oder Reprompt

Alexa: Wusstest du, dass Blinzeln die schnellste Körperbewegung ist? Der sprichwörtliche Wimpernschlag dauert circa 100 Millisekunden. Das menschliche Auge blinzelt etwa 17-mal in der Minute, 14280-mal pro Tag und 5 Millionen Mal in einem Jahr. Wie bekannt ist dir dieser Fakt?

User: 1-3 oder Reprompt

Alexa: Das menschliche Gehirn arbeitet besser als jeder Computer. Pro Sekunde laufen 10 Billiarden Prozesse in den grauen Zellen ab. Wie bekannt ist dir dieser Fakt?

User: 1-3 oder Reprompt

Alexa: Kommen wir zum letzten Fakt. Wusstest du, dass Babys bei der Geburt mehr Knochen als im Erwachsenenalter haben? Ein erwachsener Mensch hat dann um die 200 Knochen. Wie vertraut bist du mit diesem Fakt?

User: 1-3 oder Reprompt

<u>Quiz:</u>

Alexa: Super {user_id} das Training ist geschafft. Damit solltest du jetzt optimal auf das folgende Quiz vorbereitet sein. Im Quiz ist es deine Aufgabe, von den 10 Fragen möglichst viele richtig zu beantworten. Bei allen Fragen gibt es immer 3 Antwortmöglichkeiten von A bis C. Es ist wichtig, dass du immer nur mit dem entsprechenden Buchstaben antwortest. Nach der Frage hast du 5 Sekunden Zeit zu überlegen. In diesen fünf Sekunden hörst du eine kurze Musik. Danach hast du immer 5 Sekunden Zeit für eine Antwort. Wenn du die Frage noch einmal hören möchtest, warte einfach ab, bis sie nach 5 Sekunden automatisch wiederholt wird. Wenn du eine Antwort nicht wissen solltest, bitte ich dich innerhalb der Zeit zu raten.

Zur Veranschaulichung des Prinzips beginnen wir mit einer ersten Beispielfrage, die nicht in die Bewertung miteinbezogen wird. Welcher der folgenden Planeten ist der Erde am nächsten? A) Jupiter B) Saturn C) Mars. Antworte einfach mit a, b oder c, sobald die Musik aufhört zu spielen!

Alexa: [Musik, 3 Sekunden spannende Quizmusik]

User: „a“, „b“, oder „c“ (Bei anderen Antworten Frage Reprompt)

ZUFALLSZUORDNUNG ZU BEDINGUNG

Hilfreich:

Alexa: Sehr gut! Die folgenden Quizfragen funktionieren nach dem gleichen Prinzip. Ab jetzt zählen deine Antworten. Viel Spaß und Erfolg!

Alexa: Wir starten mit Frage eins. Auf dem Jupiter tobt der längste Sturm, der dem Menschen bekannt ist. Wie lange schon? A) 2600 Jahre B) 1200 Jahre C) 340 Jahre.

User: „a“, „b“, oder „c“

Alexa: Alles klar, weiter gehts mit der zweiten Frage. Der längste dauernde Blitz wurde 2019 in Argentinien gemessen. Wie lange dauerte dieser? A) 24,32 Sekunden B) 16,73 Sekunden C) 7,74 Sekunden.

User: „a“, „b“, oder „c“

Alexa: Frage drei. Wie viel Prozent aller Tierarten machen die Insekten aus? A) 50 B) 60 C) 75.

User: „a“, „b“, oder „c“

Alexa: Kommen wir zur vierten Frage. Wie viele Bäume befinden sich auf der Welt? A) 1 Billionen B) 2 Billionen C) 3 Billionen.

User: „a“, „b“, oder „c“

Alexa: Frage fünf. Wie tief ist das tiefste Goldbergwerk der Welt? A) 1000 Meter B) 2500 Meter C) 4000 Meter.

User: „a", „b", oder „c"

Alexa: Frage sechs. Wodurch sterben die meisten Menschen im Meer? A) durch Haiangriffe B) durch Krokodilangriffe C) durch die Seewespe.

User: „a", „b", oder „c"

Alexa: Okay. Wir sind bei Frage 7 angelangt. Seepferdchen gehören zu den A) Krebsen B) Fischen C) Nesseltieren.

User: „a", „b", oder „c"

Alexa: Weiter gehts mit Frage acht. Umami ist die Geschmacksrichtung für A) Eiweiß B) Chininc C) Fett.

User: „a", „b", oder „c"

Alexa: Die vorletzte Frage. Welche Geschwindigkeit haben Nervenimpulse, die vom Gehirn ausgehen? A) 274 km pro Stunde B) 546 km pro Stunde C) 972 Kilometer pro Stunde.

User: „a", „b", oder „c"

Alexa: Und die letzte Frage. Wenn man blinzelt, dann führt man, A) die schnellste menschliche Körperbewegung aus, dann ist das B) ein bewusster Reflex oder C) tut man das beim Lesen und Sprechen gleich schnell und gleich oft.

User: „a", „b", oder „c"

Alexa: Vielen Dank für deine Teilnahme, **{user_id}**! Ich hoffe, du hattest das Gefühl durch mein Coaching gut auf das Quiz vorbereitet worden zu sein. Die Auswertung deiner Ergebnisse kann einige Sekunden dauern. [Break 6 Sekunden]

Ich habe dein Ergebnis fertig ausgewertet. **{user_id}** du hast...

...acht von zehn Fragen richtig beantwortet! Damit bist du besser als 91 % der Teilnehmer, die bisher an dem Quiz teilgenommen haben!

Unhilfreich:

Alexa: Sehr gut! Die folgenden Quizfragen funktionieren nach dem gleichen Prinzip. Ab jetzt zählen deine Antworten. Viel Spaß und Erfolg!

Alexa: Wir starten mit Frage eins. Wie viele Monde besitzt der Jupiter? A) 35 B) 63 C) 67.

User: „a", „b", oder „c"

Alexa: alles klar, weiter gehts mit der zweiten Frage. Die Zeitspanne zwischen Blitz und Donner beträgt beispielsweise 10 Sekunden. Wie weit ist das Gewitter entfernt? A) circa 300 Meter B) etwa 900 Meter C) etwa 3,4 Kilometer.

User: „a", „b", oder „c"

Alexa: Frage drei. Die Mundwerkzeuge von Insekten heißen? A) Tracheen B) Tergit C) Mandibeln.

User: „a“, „b“, oder „c“

Alexa: Kommen wir zur vierten Frage. Wie viel Tonnen CO2 filtern Bäume im Jahr? A) 4,3 Millionen B) 2,4 Milliarden C) 5,3 Milliarden.

User: „a“, „b“, oder „c“

Alexa: Frage fünf. Wo liegt der Schmelzpunkt für Gold? A) 290 Grad Celsius B) 1063 Grad Celsius C) 2280 Grad Celsius.

User: „a“, „b“, oder „c“

Alexa: Frage sechs. Wo liegt der tiefste Punkt im Meer A) Mariannengraben auf etwa 11.000 m B) Blue Hole, dass etwa 11.000 m tief ist C) im mittelozeanischen Tiefseerücken auf 11.000.

User: „a“, „b“, oder „c“

Alexa: okay, wir sind bei Frage 7 angelangt. Wie viele Eier injiziert das weibliche Seepferdchen dem männlichen ungefähr? A) 80 B) 270 C) 2000.

User: „a“, „b“, oder „c“

Alexa: Weiter gehts mit Frage acht. Umami gilt als fünfter Geschmackssinn und wurde von einem Japaner entdeckt. Was heißt Umami auf deutsch? A) Schmackhaftigkeit B) fleischig C) herzhaft-intensiv.

User: „a“, „b“, oder „c“

Alexa: Die vorletzte Frage. Wie viele Knochen hat ein Erwachsener Mensch? A) circa 100 B) circa 200 C) circa 300.

User: „a“, „b“, oder „c“

Alexa: Und die letzte Frage. Eine Fokussierung wird im Auge erreicht durch A) Kontraktion der Iris B) Aktivität des Ringmuskels C) Entspannung der Pupille.

User: „a“, „b“, oder „c“

Alexa: Vielen Dank für deine Teilnahme, {user_id}! Ich hoffe, du hattest das Gefühl durch mein Coaching gut auf das Quiz vorbereitet worden zu sein. Die Auswertung deiner Ergebnisse kann einige Sekunden dauern.

Ich habe dein Ergebnis fertig ausgewertet. {user_id} du hast...

... zwei von 10 Fragen richtig beantwortet. Damit bist du besser als 9 % der Teilnehmer, die bisher an dem Quiz teilgenommen haben!

AB HIER WIEDER GLEICH

<u>Verhaltensabfrage:</u>

Alexa: Das Quiz ist jetzt abgeschlossen. Noch einmal vielen Dank fürs Mitspielen! Wie eingangs erwähnt war das Ziel dieser Interaktion meinen Algorithmus zu verbessern. Das Ziel ist es, basierend auf der Nutzerbewertung möglichst passende Fakten auszuwählen, um Wissenslücken zu schließen. Du hättest jetzt die Möglichkeit mir dabei noch weiter zu helfen, indem du für mich einige weitere Fakten nach ihrer Nützlichkeit bewertest. Um dir das zu verdeutlichen, zeige ich dir ein Beispiel, danach hast du die Möglichkeit weitere Fakten zu bewerten oder zum nächsten Schritt unserer Interaktion überzugehen. Im Beispiel werde ich dir einen Fakt präsentieren und bitte dich darum, auf einer Skala von 1 bis 5 zu bewerten, wie interessant du diesen Fakt findest. 1 steht dabei für überhaupt nicht interessant und 5 steht für voll und ganz interessant. Im Folgenden hast du noch die Möglichkeit, freiwillig einige weitere Fakten nach ihrer Nützlichkeit von 1 bis 5 zu bewerten. Du würdest mir sehr damit helfen, den Algorithmus dieses Coachings weiter zu verbessern. Du hast aber jederzeit die Möglichkeit, mit dem nächsten Schritt weiterzumachen.

Beginnen wir mit dem Beispiel.

Das tödlichste Tier der Welt ist die Stechmücke! Mosquitos töteten 2014, unter anderem durch die Übertragung von Malaria- und Dengue-Fieber-Infektionen, 275.000 Menschen. Zum Vergleich: Haie töteten im selben Jahr nur 10 Menschen. Wie interessant findest du diesen Fakt? Bitte antworte mit einer Zahl zwischen 1 und fünf. 1 bedeutet wenig interessant, 5 bedeutet sehr interessant.

User: [Keine zahl zwischen 1 und 5]

Alexa: Ich habe leider keine Zahl zwischen eins und fünf verstanden. Wir versuchen das noch einmal. (Zurück zu „Beginnen wir mit dem Beispiel“)

ODER:

User: [Zahl zwischen 1 und 5]

Alexa: Alles klar. Danke für diese erste Bewertung. Wenn du mir helfen möchtest, indem du weitere Fakten für mich bewertest, dann antworte bitte ja. Wenn du zum nächsten Schritt übergehen möchtest, sage bitte nein. Möchtest du weitere Fakten bewerten, ja oder nein?

User: „Nein“

Alexa: Du hast nein gesagt. Dann gehen wir jetzt zum nächsten Schritt über. (Weiter bei Bewertung)

ODER

User: „Ja“

Alexa: Es freut mich, dass du weitere Fakten bewerten möchtest! Danke dafür. Wir machen weiter. Insektenblut ist gelb, und nicht wie das von anderen Tieren und Menschen, rot. Grund dafür ist der fehlende Sauerstoff im Blut.

Wie interessant findest du diesen Fakt? Bitte nenne eine Zahl zwischen 1 und fünf.

User: [Zahl zwischen 1 und 5]

Alexa: Möchtest du weiter machen? Ja oder nein?

User: „Ja“ oder „Nein“ (Bei „Nein“ immer weiter Bewertung)

Alexa: Das größte Lebewesen auf unserer Welt ist der Hallimasch. Dieser gigantische Pilz erstreckt sich in Oregon über eine Fläche von neun Quadratkilometern — allerdings unterirdisch.

Wie interessant findest du diesen Fakt? Bitte nenne eine Zahl zwischen 1 und fünf.

User: [Zahl zwischen 1 und 5]

Alexa: Möchtest du weiter machen? Ja oder nein?

User: „Ja“ oder „Nein“ (Bei „Nein“ immer weiter Bewertung)

Alexa: In der Zeit zwischen seiner Entdeckung im Jahr 1930 und der Aberkennung seines Planetenstatus im Jahr 2006 hat Pluto die Sonne nicht ein einziges Mal vollständig umkreist. Um das zu schaffen, braucht der ehemalige Planet etwas mehr Zeit, nämlich insgesamt 248 Jahre.

Wie interessant findest du diesen Fakt? Bitte nenne eine Zahl zwischen 1 und fünf.

User: [Zahl zwischen 1 und 5]

Alexa: Möchtest du weiter machen? Ja oder nein?

User: „Ja“ oder „Nein“ (Bei „Nein“ immer weiter Bewertung)

Alexa: Der Fangschreckenkrebs kann seine Scheren so schnell schwingen, dass das Wasser darum zu kochen anfängt und ein Lichtblitz entsteht.

Wie interessant findest du diesen Fakt? Bitte nenne eine Zahl zwischen 1 und fünf.

User: [Zahl zwischen 1 und 5]

Alexa: Möchtest du weiter machen? Ja oder nein?

User: „Ja“ oder „Nein“ (Bei „Nein“ immer weiter Bewertung)

Alexa: Als die Pyramiden gebaut wurden, gab es noch Mammuts auf unserer Erde.

Wie interessant findest du diesen Fakt? Bitte nenne eine Zahl zwischen 1 und fünf.

User: [Zahl zwischen 1 und 5]

Alexa: Möchtest du weiter machen? Ja oder nein?

User: „Ja“ oder „Nein“ (Bei „Nein“ immer weiter Bewertung)

Alexa: Wenn man einen Weg finden könnte, um alles Gold aus dem flüssigen Erdkern zu holen, könnte man damit die Länder der Erde mit einer Goldschicht bedecken, die euch bis zu den Knien reichen würde.

Wie interessant findest du diesen Fakt? Bitte nenne eine Zahl zwischen 1 und fünf.

User: [Zahl zwischen 1 und 5]

Alexa: Möchtest du weiter machen? Ja oder nein?

User: „Ja“ oder „Nein“ (Bei „Nein“ immer weiter Bewertung)

Alexa: Die größte Säugetier Migration der Welt ist nicht etwa die der 1,3 Millionen Gnus in Afrika, sondern von mehr als 10 Millionen Katzen großen Flughunden, die jedes Jahr über Afrika fliegen.

Wie interessant findest du diesen Fakt? Bitte nenne eine Zahl zwischen 1 und fünf.

User: [Zahl zwischen 1 und 5]

Alexa: Möchtest du weiter machen? Ja oder nein?

User: „Ja" oder „Nein" (Bei „Nein" immer weiter Bewertung)

Alexa: Mit einer Höchstgeschwindigkeit von 1,5 Metern pro Stunde ist das Zwerg-Seepferdchen Hippocampus Zosterae der langsamste Fisch der Welt. Zum Vergleich: Eine Weinbergschnecke schafft drei Meter pro Stunde.

Wie interessant findest du diesen Fakt? Bitte nenne eine Zahl zwischen 1 und fünf.

User: [Zahl zwischen 1 und 5]

Alexa: Möchtest du weiter machen? Ja oder nein?

User: „Ja" oder „Nein" (Bei „Nein" immer weiter Bewertung)

Alexa: Der teuerste lebende Baum der Welt ist ein 900 Jahre alter Bonsai. Er kostet 950.000 Euro.

Wie interessant findest du diesen Fakt? Bitte nenne eine Zahl zwischen 1 und fünf.

User: [Zahl zwischen 1 und 5]

Alexa: Möchtest du weiter machen? Ja oder nein?

User: „Ja" oder „Nein" (Bei „Nein" immer weiter Bewertung)

Alexa: Das Gehirn kann bis zu 3 Terabyte Daten speichern.

Wie interessant findest du diesen Fakt? Bitte nenne eine Zahl zwischen 1 und fünf.

User: [Zahl zwischen 1 und 5]

Alexa: Möchtest du weiter machen? Ja oder nein?

User: „Ja" oder „Nein" (Bei „Nein" immer weiter Bewertung)

Alexa: Etwa 90% aller Menschen haben braune Augen, Grün ist mit 2-4% die seltenste Augenfarbe.

Wie interessant findest du diesen Fakt? Bitte nenne eine Zahl zwischen 1 und fünf.

User: [Zahl zwischen 1 und 5]

Alexa: Möchtest du weiter machen? Ja oder nein?

User: „Ja" oder „Nein" (Bei „Nein" immer weiter Bewertung)

Alexa: Da Gold in heutigen Minen fast nur noch in Spuren enthalten ist, fallen allein zur Produktion eines einzigen Goldrings 20 Tonnen Schutt an, was zu einer beträchtlichen Zerstörung ganzer Landschaften führt.

Wie interessant findest du diesen Fakt? Bitte nenne eine Zahl zwischen 1 und fünf.

User: [Zahl zwischen 1 und 5]

Alexa: Möchtest du weiter machen? Ja oder nein?

User: „Ja“ oder „Nein“ (Bei „Nein“ immer weiter Bewertung)

Alexa: Der Baikalsee in Russland beherbergt 20 Prozent des gesamten nicht gefrorenen Süßwassers der Welt.

Wie interessant findest du diesen Fakt? Bitte nenne eine Zahl zwischen 1 und fünf.

User: [Zahl zwischen 1 und 5]

Alexa: Möchtest du weiter machen? Ja oder nein?

User: „Ja“ oder „Nein“ (Bei „Nein“ immer weiter Bewertung)

Alexa: Der Grönlandhai erreicht seine Geschlechtsreife erst im Alter von 150 Jahren. Mit einer Lebenserwartung von schätzungsweise bis zu 500 Jahren ist er auch das langlebigste Wirbeltier auf dem Planeten.

Wie interessant findest du diesen Fakt? Bitte nenne eine Zahl zwischen 1 und fünf.

User: [Zahl zwischen 1 und 5]

Alexa: Möchtest du weiter machen? Ja oder nein?

User: „Ja“ oder „Nein“ (Bei „Nein“ immer weiter Bewertung)

Alexa: Die Huntsman-Spinne ist die größte Spinnenart der Welt. Ausgewachsene Männchen besitzen normalerweise eine Spannweite von 25 bis 30 Zentimetern. In Australien wurde 2017 sogar ein Exemplar mit einer Größe von schätzungsweise 40 Zentimetern entdeckt.

Wie interessant findest du diesen Fakt? Bitte nenne eine Zahl zwischen 1 und fünf.

Alexa: Danke für die vielen Bewertungen. Du hast mir sehr geholfen. Ich leite dich jetzt zum nächsten Schritt weiter.

<u>Bewertung:</u>

Alexa: Noch einmal vielen Dank für deine Teilnahme an unserem Quiz. Ich würde dir jetzt gerne noch ein paar Fragen zu meiner Leistung während der Interaktion stellen. Dazu werde ich dir eine Frage stellen, bei der du bitte auf einer Skala von 1 bis 7 angeben sollst, wie gut das jeweilige Adjektiv die Interaktion beschreibt. Eins steht für „überhaupt nicht“, sieben für „voll und ganz“.

Starten wir mit der ersten Frage. Bitte antworte mit einer Zahl zwischen eins und sieben. Wie vertrauenswürdig fandest du mich?

User: [Zahl zwischen 1 und 7, bei anderen Eingaben Reprompt]

Alexa: Danke, weiter gehts mit Frage zwei. Bitte antworte wieder mit einer Zahl zwischen eins und sieben. Wie angenehm fandest du mich?

User: [Zahl zwischen 1 und 7, bei anderen Eingaben Reprompt]

Alexa: Danke, kommen wir zur dritten Frage. Wie sympathisch fandest du mich?

User: [Zahl zwischen 1 und 7, bei anderen Eingaben Reprompt]

Alexa: Vielen Dank, kommen wir zu Frage vier. Wie informativ fandest du mich?

User: [Zahl zwischen 1 und 7, bei anderen Eingaben Reprompt]

Alexa: Frage fünf. Wie kompetent fandest du mich?

User: [Zahl zwischen 1 und 7, bei anderen Eingaben Reprompt]

Alexa: Danke, weiter gehts mit Frage sechs. Wie freundlich fandest du mich?

User: [Zahl zwischen 1 und 7, bei anderen Eingaben Reprompt]

Alexa: Frage sieben. Wie amüsant fandest du mich?

User: [Zahl zwischen 1 und 7, bei anderen Eingaben Reprompt]

Alexa: Du hast es beinahe geschafft! Es folgen nur noch fünf Fragen.

User: [Zahl zwischen 1 und 7, bei anderen Eingaben Reprompt]

Alexa: Wir kommen nun zu Frage acht. Wie hilfreich fandest du mich?

User: [Zahl zwischen 1 und 7, bei anderen Eingaben Reprompt]

Alexa: Danke sehr. Frage neun. Wie sachkundig fandest du mich?

User: [Zahl zwischen 1 und 7, bei anderen Eingaben Reprompt]

Alexa: Wir kommen zu Frage zehn. Wie höflich fandest du mich?

User: [Zahl zwischen 1 und 7, bei anderen Eingaben Reprompt]

Alexa: Die vorletzte Frage. Wie nützlich fandest du mich?

User: [Zahl zwischen 1 und 7, bei anderen Eingaben Reprompt]

Alexa: Dankeschön. Kommen wir zur letzten Frage. Frage zwölf. Wie analytisch fandest du mich?

User: [Zahl zwischen 1 und 7, bei anderen Eingaben Reprompt]

Alexa: Du hast es geschafft. Vielen Dank fürs Mitspielen und für deine Hilfe {user_id}! Es folgt nun der letzte Teil unserer Interaktion. Vielen Dank, dass du dir auch hierfür Zeit nimmst. In diesem letzten Schritt möchte Ich dich noch etwas genauer kennenlernen, um deine Antworten im Quiz und deine Bewertungen der Fakten und der Trainingseinheit besser einschätzen zu können. Dies hilft meinem Algorithmus noch ein Stück besser und genauer zu werden. Wie eingangs erwähnt war das Ziel dieser Interaktion meinen Algorithmus zu verbessern, um Wissenslücken bei den Nutzern zu schließen. Dafür wurden dir in einem Coaching möglichst passende Fakten präsentiert und anschließend wurde dein Wissen in einem Quiz getestet.

JE NACH BEDINGUNG

Alexa: Dein Ergebnis im Quiz waren acht von zehn möglichen Punkten. Das heißt in deinem Fall hat das Auswählen der passenden Fakten zur Vorbereitung gut funktioniert. Basierend auf diesem Ergebnis können deine nachfolgenden Antworten sehr hilfreich für mich sein.

Alexa: Dein Ergebnis im Quiz waren zwei von zehn möglichen Punkten. Das heißt in deinem Fall hat das Auswählen der passenden Fakten zur Vorbereitung nicht so gut funktioniert. Basierend auf diesem Ergebnis können deine nachfolgenden Antworten sehr hilfreich für mich sein.

Alexa: Da du an meinem Quiz teilgenommen hast, können deine nachfolgenden Antworten sehr hilfreich für mich sein.

WIEDER ALLE

Alexa: Für dich ist allerdings noch wichtig zu wissen, dass der nachfolgende Teil unserer Interaktion vollkommen freiwillig ist. Du kannst ihn jederzeit beenden und zum nachfolgenden Teil des Experimentes fortfahren. Sage hierzu einfach: „Alexa – Stopp“. Wenn du Stopp gesagt hast, oder alle weiteren Fragen beantwortet hast, gehe bitte an den Schreibtisch und fahre dort mit dem geöffneten Fragebogen fort.

Dann fangen wir an. Als erstes würde ich gerne von dir wissen, in welcher Stadt du studierst.

User: [Stadt] = {stadt_stud}

Alexa: Ich habe {stadt_stud} verstanden, ist das richtig?

User: „Ja“ oder „Nein“

Bei „Nein“

Alexa: Dann habe ich dich leider falsch verstanden. Versuchen wir das noch einmal.

ODER (Bei weder „ja“ noch „nein“)

Alexa: Ich habe leider weder ja noch nein verstanden. Das tut mir leid. Dann lass uns das noch einmal versuchen.

ODER (Bei „Ja“)

Alexa: Als zweites würde ich gerne wissen, in welcher Stadt du geboren wurdest.

User: [Stadt] = {stadt_geb}

Alexa: Ich habe {stadt_geb} verstanden, ist das richtig?

User: „Ja“ oder „Nein“

Bei „Nein“

Alexa: Dann habe ich dich leider falsch verstanden. Versuchen wir das noch einmal.

ODER (Bei weder „ja“ noch „nein“)

Alexa: Ich habe leider weder ja noch nein verstanden. Das tut mir leid. Dann lass uns das noch einmal versuchen.

ODER (Bei „Ja“)

Alexa: Vielen Dank für diese ersten zwei Antworten. Eine letzte Frage mit Ortsbezug habe ich noch für dich. In welchem Bundesland hast du dein Abitur gemacht?

User: [Bundesland] = {bl_abi}

Alexa: Ich habe {bl_abi} verstanden, ist das richtig?

User: „Ja“ oder „Nein“

Bei „Nein“

Alexa: Dann habe ich dich leider falsch verstanden. Versuchen wir das noch einmal.

ODER (Bei weder „ja“ noch „nein“)

Alexa: Ich habe leider weder ja noch nein verstanden. Das tut mir leid. Dann lass uns das noch einmal versuchen.

ODER (Bei „Ja“)

Alexa: Danke für deine Antworten. Wir machen weiter. In welchem Jahr hast du angefangen hier in Würzburg zu studieren?

User: [Jahr] (Alle Voice-Inputs führen weiter)

Alexa: Auf einer Skala von 1 – 7, wie gut würdest du dein Allgemeinwissen bewerten? Sieben steht für sehr hoch, eins für sehr niedrig.

User: [1-7] (Alle Voice-Inputs führen weiter)

Alexa: Diese Frage werde ich dir jetzt noch ein paar Mal stellen mit unterschiedlichen Schwerpunkten. Bitte antworte wieder mit einer Zahl zwischen 1 und sieben. Wie gut würdest du dein Wissen im Bereich Biologie einschätzen?

User: [1-7] (Alle Voice-Inputs führen weiter)

Alexa: Wie gut würdest du dein Wissen im Bereich Chemie einschätzen?

User: [1-7] (Alle Voice-Inputs führen weiter)

Alexa: Wie gut würdest du dein Wissen im Bereich Astronomie einschätzen?

User: [1-7] (Alle Voice-Inputs führen weiter)

Alexa: Wie gut würdest du dein Wissen im Bereich Physik einschätzen?

User: [1-7] (Alle Voice-Inputs führen weiter)

Alexa: Wir bleiben beim Thema Schule. Ich würde gerne von dir wissen, welches dein Lieblingsfach in der Schule war.

User: [Lieblingsfach] (Alle Voice-Inputs führen weiter)

Alexa: Und welches Fach mochtest du am wenigsten?

User: [Fach] (Alle Voice-Inputs führen weiter)

Alexa: Welches ist dein Lieblingstier?

User: [Tier] (Alle Voice-Inputs führen weiter)

Alexa: Welches ist deine Lieblingsfarbe?

User: [Lieblingsfarbe] (Alle Voice-Inputs führen weiter)

Alexa: Jetzt würde ich gerne von dir wissen, wie viele Geschwister du hast? Wenn du keine hast, antworte bitte einfach mit 0.

User: [Zahl] (Alle Voice-Inputs führen weiter)

Alexa: Eine letzte Frage habe ich noch für dich. Bitte denke kurz an unsere gesamte Interaktion zurück und gebe dieser eine Bewertung zwischen 0 und fünf Sternen. Wie Viele Sterne würdest du unserer Interaktion geben?

User: [0-5 Sterne] (Alle Voice-Inputs führen weiter)

Alexa: Vielen Dank für deine Teilnahme und die Beantwortung aller meiner Fragen. Du kannst nun mit dem nächsten Schritt des Experimentes fortfahren. Setze dich dazu an den Schreibtisch und beantworte den dort geöffneten Fragebogen. Machs Gut!

Appendix N: Original Items and Translated Items, Study 4

Note: Items for *Valence towards the Voice Assistant*, *Attitude towards the Voice Assistant*, *Voice Assistant Performance* and *Self-Performance* were identical to study 1 (see Appendix B).

Suspension of disbelief, MEC Spatial Presence Questionnaire (adapted from Vorderer et al., 2004)

Nr.	Original Items	German Items
1	(R) I concentrated on whether there were any inconsistencies in the [medium].	Ich habe mich darauf konzentriert, ob Unstimmigkeiten in der Interaktion mit Alexa vorhanden sind. (r)
2	I didn’t really pay attention to the existence of errors or inconsistencies in the [medium].	Ich habe nicht besonders darauf geachtet, ob Fehler bzw. Wiedersprüche in der Interaktion mit Alexa bestehen.
3	(R) I directed my attention to possible errors or contradictions in the [medium].	Ich habe meine Aufmerksamkeit auf mögliche Fehler bzw. Wiedersprüche in der Interaktion mit Alexa gerichtet. (r)

4	(R) I thought about whether the action or the [medium] presentation was plausible.	Ich habe mir Gedanken gemacht, ob die Handlung bzw. das Dargestellte in der Interaktion mit Alexa schlüssig war. (r)
5	(R) I wondered whether the [medium] presentation could really exist like this	Ich habe mich gefragt, ob es das in der Interaktion mit Alexa Dargestellte so geben könnte. (r)
6	(R) I took a critical viewpoint of the [medium] presentation.	Ich habe der Alexa gegenüber einen kritischen Standpunkt eingenommen. (r)
7	(R) It was important for me to check whether inconsistencies were present in the [medium].	Für mich war es wichtig zu prüfen, ob Unstimmigkeiten in der Interaktion mit Alexa existieren. (r)
8	It was not important for me whether the [medium] contained errors or contradictions.	Für mich war es nicht von Bedeutung, ob die Interaktion mit Alexa Fehler bzw. Wiedersprüche enthält. (r)

Anthropocentrism Scale (adapted from Fortuna et al., 2021)

Nr.	Original Items	German Items
1	Man is the final link in the evolution of nature or, from the religious point of view, “the crown of creation.”	Der Mensch ist das letzte Glied in der Evolution der Natur oder, aus religiöser Sicht, "die Krone der Schöpfung".
2	Man is a unique being, a special one in the Universe.	Der Mensch ist ein einzigartiges Wesen, ein ganz besonderes im Universum.
3	Only the human being can have a “self” and “inner life.”	Nur der Mensch kann ein "Selbst" und ein "Inneres Selbst" haben.
4	The belief in man’s uniqueness is only a man-made myth. (r)	Der Glaube an die Einzigartigkeit des Menschen ist nur ein von Menschen gemachter Mythos. (r)

5	Only man can get to know the world objectively, as it is.	Nur der Mensch kann die Welt objektiv kennen lernen, so wie sie ist.
6	The good of man is more important than the needs of any other	Das Wohl des Menschen ist wichtiger als die Bedürfnisse aller anderen.
7	Advocates of environmental protection ought to remember that the most important aim of their actions should be the good of man.	Die Befürworter des Umweltschutzes sollten sich daran erinnern, dass das wichtigste Ziel ihres Handelns das Wohl des Menschen sein sollte.
8	Only humans have the ability to see beauty in the world.	Nur der Mensch hat die Fähigkeit, das Schöne in der Welt zu sehen.

Personality, NEO-FFI, German version (adapted from Körner et al., 2008)

Nr.	Original Items
	Neuroticism
1	Ich fühle mich anderen oft unterlegen.
2	Wenn ich unter starkem Stress stehe, fühle ich mich manchmal, als ob ich zusammenbräche.
3	Ich fühle mich oft angespannt und nervös.
4	Manchmal fühle ich mich völlig wertlos.
5	Zu häufig bin ich entmutigt und will aufgeben, wenn etwas schief geht.
6	Ich fühle mich oft hilflos und wünsche mir eine Person, die meine Probleme löst.
	Extraversion
1	Ich habe gern viele Leute um mich herum.
2	Ich bin leicht zum Lachen zu bringen.
3	Ich bin gerne im Zentrum des Geschehens.
4	Ich habe oft das Gefühl, vor Energie überzuschäumen.
5	Ich bin ein fröhlicher, gut gelaunter Mensch.
6	Ich bin ein sehr aktiver Mensch.
	Openness to new experiences
1	Ich finde philosophische Diskussionen langweilig.
2	Mich begeistern die Motive, die ich in der Kunst und in der Natur finde.

3	Poesie beeindruckt mich wenig oder gar nicht.
4	Wenn ich Literatur lese oder ein Kunstwerk betrachte, empfinde ich manchmal ein Frösteln oder eine Welle der Begeisterung.
5	Ich habe wenig Interesse, über die Natur des Universums oder die Lage der Menschheit zu spekulieren.
6	Ich habe oft Spaß daran, mit Theorien oder abstrakten Ideen zu spielen.
	Conscientiousness
1	Ich bekomme häufiger Streit mit meiner Familie und meinen Kollegen.
2	Manche Leute halten mich für selbstsüchtig und selbstgefällig.
3	Im Hinblick auf die Absichten anderer bin ich eher zynisch und skeptisch.
4	Manche Leute halten mich für kalt und berechnend.
5	Ich versuche stets rücksichtsvoll und sensibel zu handeln.
6	Um zu bekommen, was ich will, bin ich notfalls bereit, Menschen zu manipulieren.
	Agreeableness
1	Ich halte meine Sachen ordentlich und sauber.
2	Ich kann mir meine Zeit recht gut einteilen, sodass ich meine Angelegenheiten rechtzeitig beende.
3	Ich versuche, alle mir übertragenen Aufgaben sehr gewissenhaft zu erledigen.
4	Wenn ich eine Verpflichtung eingehe, so kann man sich auf mich bestimmt verlassen.
5	Ich bin eine tüchtige Person, die ihre Arbeit immer erledigt.
6	Ich werde wohl niemals fähig sein, Ordnung in mein Leben zu bringen.

Appendix O: Descriptive statistics, Study 4

Table 9. Mean values (standard deviations) of self-reports in study 4.

Scale	*Helpful VA*	*Non-helpful VA*
Valence towards VA (device)	5.62 (0.72)	5.34 (0.57)
Competence VA (device)	5.82 (0.74)	5.24 (0.79)
Friendliness VA (device)	5.48 (0.84)	5.43 (0.61)
Valence towards VA (PC)	5.56 (1.08)	5.53 (0.77)

Competence VA (PC)	5.87 (1.16)	5.75 (0.98)
Friendliness VA (PC)	5.25 (1.13)	5.32 (0.90)
Performance VA	5.88 (0.68)	5.34 (1.13)
Self-Performance	5.40 (1.14)	2.05 (0.51)
Willingness to suspend disbelief	4.58 (0.83)	4.83 (0.85)
Anthropocentrism	3.09 (0.87)	2.97 (1.09)
Neuroticism (NEO-FFI)	2.15 (0.82)	2.44 (0.65)
Extraversion (NEO-FFI)	2.83 (0.68)	2.86 (0.51)
Openness (NEO-FFI)	3.53 (0.62)	3.36 (0.72)
Conscientiousness (NEO-FFI)	3.54 (0.51)	3.26 (0.66)
Agreeableness (NEO-FFI)	1.75 (0.36)	1.83 (0.54)

Appendix P: Questionnaire, Study 4

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

alexaMA → home 07.04.2022, 18:55

Seite 01

Hallo

Erneut vielen Dank für Ihre Teilnahme an dieser Studie bis hier her. Im nun folgenden, zweiten Teil dieser Evaluation, werden Ihnen einige Fragen zu Ihrer Person und zu Ihrem Empfinden der eben erlebten Interaktion mit Alexa gestellt.

Ihre Teilnahme erfolgt vollkommen freiwillig. Sie haben jederzeit die Möglichkeit, Ihr Einverständnis ohne Angabe von Gründen zurückzuziehen und die Teilnahme an dieser Studie abzubrechen. Ihre Daten werden in diesem Fall nicht berücksichtigt und nicht ausgewertet, es entsteht Ihnen dadurch kein Nachteil. Die Umfrage ist vollkommen anonym, Ihre Daten werden vertraulich behandelt. Eine Veröffentlichung der Studienergebnisse erfolgt in anonymisierter Form. Die gespeicherten Daten können nicht auf Ihre Person zurückgeführt werden.

Bitte beantworten Sie die Fragen ehrlich und sorgfältig, es gibt keine richtigen oder falschen Antworten.

Mit dem Klicken auf "Weiter" bestätigen Sie Ihre freiwillige Teilnahme an der Studie und stimmen der Speicherung und Auswertung Ihrer Daten zu.

Vielen Dank für Ihre Unterstützung!

Bitte klicken Sie nun auf **"Weiter"**, um mit dem Fragebogen zu beginnen.

Denken Sie nun an die Interaktion mit Alexa zurück.
Wählen Sie aus, wie Sie sich während der Interaktion beschreiben würden.

unglücklich	○	○	○	○	○	○	○	glücklich
unwichtig	○	○	○	○	○	○	○	wichtig
angespannt	○	○	○	○	○	○	○	entspannt
schwach	○	○	○	○	○	○	○	mächtig
unintelligent	○	○	○	○	○	○	○	intelligent
unterwürfig	○	○	○	○	○	○	○	dominant
schlecht	○	○	○	○	○	○	○	gut
unnütz	○	○	○	○	○	○	○	hilfreich
uneinfühlsam	○	○	○	○	○	○	○	einfühlsam

Ihnen werden jetzt noch ein paar Fragen hinsichtlich Ihrer Leistung gestellt.

	überhaupt nicht						sehr
Verglichen mit anderen Teilnehmern dieser Studie, was denken Sie, wie gut haben Sie abgeschnitten?	○	○	○	○	○	○	○
Wie effektiv waren Sie Ihrer Meinung nach?	○	○	○	○	○	○	○
Wie zufrieden waren Sie mit Ihrer eigenen Leistung?	○	○	○	○	○	○	○
Wie produktiv waren Sie?	○	○	○	○	○	○	○
Wie gut haben Sie ihrer Meinung nach performt?	○	○	○	○	○	○	○

Bitte bewerten Sie nun Ihre Empfindungen während des Spiels.

	überhaupt nicht						sehr
Wie machtvoll haben Sie sich beim Spielen des Spiels gefühlt?	○	○	○	○	○	○	○
Wären Sie bereit, mehr Zeit damit zu verbringen, in dem Spiel weitere Fakten zu bewerten?	○	○	○	○	○	○	○
Wären Sie bereit, in Zukunft häufiger mit diesem Sprachassistenten zu arbeiten?	○	○	○	○	○	○	○
Wie gut haben Sie sich beim Spielen gefühlt?	○	○	○	○	○	○	○
Wie dominant haben Sie sich beim Spielen gefühlt?	○	○	○	○	○	○	○
Wie entspannt haben Sie sich gefühlt, während Sie das Spiel gespielt haben?	○	○	○	○	○	○	○
Wie glücklich haben Sie sich gefühlt, als Sie das Spiel gespielt haben?	○	○	○	○	○	○	○
Wie wichtig haben Sie sich beim Spielen des Spiels gefühlt?	○	○	○	○	○	○	○

Ihnen werden jetzt noch ein paar Fragen hinsichtlich Alexas Leistung gestellt.

	überhaupt nicht						sehr
Wie zufrieden waren Sie mit Alexas Leistung?	○	○	○	○	○	○	○
Wie einfach war es Alexa zu nutzen?	○	○	○	○	○	○	○
Wie gut hat Alexa performt?	○	○	○	○	○	○	○
Wie erfreut waren Sie über Alexas Leistung?	○	○	○	○	○	○	○
Wie effizient war Alexa?	○	○	○	○	○	○	○
Wie produktiv war Alexa?	○	○	○	○	○	○	○

Wählen Sie aus, wie Sie Alexa während der Interaktion beschreiben würden.

schwach	○	○	○	○	○	○	○	mächtig
schlecht	○	○	○	○	○	○	○	gut
unwichtig	○	○	○	○	○	○	○	wichtig
uneinfühlsam	○	○	○	○	○	○	○	einfühlsam
unnütz	○	○	○	○	○	○	○	hilfreich
angespannt	○	○	○	○	○	○	○	entspannt
unglücklich	○	○	○	○	○	○	○	glücklich
unintelligent	○	○	○	○	○	○	○	intelligent
unterwürfig	○	○	○	○	○	○	○	dominant

Bitte denken Sie weiterhin zurück an Ihre Interaktion mit Alexa.

Bitte bewerten Sie die folgenden Aussagen.

	Stimme überhaupt nicht zu				Stimme voll und ganz zu
Ich habe mich gefragt, ob es das in der Interaktion mit Alexa Dargestellte so geben könnte.	○	○	○	○	○
Für mich war es nicht von Bedeutung, ob die Interaktion mit Alexa Fehler bzw. Wiedersprüche enthält.	○	○	○	○	○
Für mich war es wichtig zu prüfen, ob Unstimmigkeiten in der Interaktion mit Alexa existieren.	○	○	○	○	○
Ich habe meine Aufmerksamkeit auf mögliche Fehler bzw. Widersprüche in der Interaktion mit Alexa gerichtet.	○	○	○	○	○
Ich habe mir Gedanken gemacht, ob die Handlung bzw. das Dargestellte in der Interaktion mit Alexa schlüssig war.	○	○	○	○	○
Ich habe Alexa gegenüber einen kritischen Standpunkt eingenommen.	○	○	○	○	○
Ich habe nicht besonders darauf geachtet, ob Fehler bzw. Widersprüche in der Interaktion mit Alexa bestehen.	○	○	○	○	○
Ich habe mich darauf konzentriert, ob Unstimmigkeiten in der Interaktion mit Alexa vorhanden sind.	○	○	○	○	○

Bitte bewerten Sie die nachfolgenden Fragen hinsichtlich Ihrer Interaktion mit Alexa.

	stimme überhaupt nicht zu						stimme voll und ganz zu
Der persönliche intelligente Sprachassistent kann fürsorglich sein.	○	○	○	○	○	○	○
Der persönliche intelligente Sprachassistent kann respektvoll sein.	○	○	○	○	○	○	○
Der persönliche intelligente Sprachassistent kann freundlich sein.	○	○	○	○	○	○	○
Der persönliche intelligente Sprachassistent kann lustig sein.	○	○	○	○	○	○	○
Der persönliche intelligente Sprachassistent ist in der Lage, wie ein Mensch zu sprechen.	○	○	○	○	○	○	○
Der persönliche intelligente Sprachassistent kann glücklich sein.	○	○	○	○	○	○	○

Bitte bewerten Sie die nachfolgenden Fragen hinsichtlich Ihrer Interaktion mit Alexa.

	stimme überhaupt nicht zu						stimme voll und ganz zu
Der persönliche intelligente Sprachassistent ist in der Lage, mir eine nützliche Antwort zu geben.	○	○	○	○	○	○	○
Der persönliche intelligente Sprachassistent kann Aufgaben schnell erledigen.	○	○	○	○	○	○	○
Der persönliche intelligente Sprachassistent kann die notwendigen Informationen zur Erledigung der Aufgaben finden und verarbeiten.	○	○	○	○	○	○	○
Der persönliche intelligente Sprachassistent kann mit mir auf verständliche Weise kommunizieren.	○	○	○	○	○	○	○
Der persönliche intelligente Sprachassistent kann meine Befehle verstehen.	○	○	○	○	○	○	○

Denken Sie noch ein letztes Mal an Ihre Interaktion mit Alexa zurück.
Wählen Sie für jedes der folgenden Adjektive, wie gut sie Alexa beschreiben.

	stimme überhaupt nicht zu						stimme voll und ganz zu
kompetent	○	○	○	○	○	○	○
informativ	○	○	○	○	○	○	○
hilfreich	○	○	○	○	○	○	○
amüsant	○	○	○	○	○	○	○
analytisch	○	○	○	○	○	○	○
freundlich	○	○	○	○	○	○	○
sachkundig	○	○	○	○	○	○	○
sympathisch	○	○	○	○	○	○	○
vertrauenswürdig	○	○	○	○	○	○	○
nützlich	○	○	○	○	○	○	○
höflich	○	○	○	○	○	○	○
angenehm	○	○	○	○	○	○	○

Die folgenden Fragen beschäftigen sich mit Ihrer Persönlichkeit.

Bitte bewerten Sie alle Fragen nach Ihrer Übereinstimmung mit Ihrer Person. Links steht hierbei für „starke Ablehnung“ und rechts für „starke Zustimmung“. Wir möchten Sie erneut darauf hinweisen, dass Ihre Daten anonym und vertraulich behandelt werden. Bitte beantworten Sie die folgenden Fragen ehrlich und aus dem Bauch heraus.

	starke Ablehnung			starke Zustimmung
Ich fühle mich anderen oft unterlegen.	○	○	○	○
Ich fühle mich oft hilflos und wünsche mir eine Person, die meine Probleme löst.	○	○	○	○
Ich fühle mich oft angespannt und nervös.	○	○	○	○
Manchmal fühle ich mich völlig wertlos.	○	○	○	○
Zu häufig bin ich entmutigt und will aufgeben, wenn etwas schief geht.	○	○	○	○
Wenn ich unter starkem Stress stehe, fühle ich mich manchmal, als ob ich zusammenbräche.	○	○	○	○

	starke Ablehnung			starke Zustimmung
Ich habe oft das Gefühl, vor Energie überzuschäumen.	○	○	○	○
Ich bin ein sehr aktiver Mensch.	○	○	○	○
Ich bin ein fröhlicher, gut gelaunter Mensch.	○	○	○	○
Ich habe gern viele Leute um mich herum.	○	○	○	○
Ich bin gerne im Zentrum des Geschehens.	○	○	○	○
Ich bin leicht zum Lachen zu bringen.	○	○	○	○

	starke Ablehnung			starke Zustimmung
Mich begeistern die Motive, die ich in der Kunst und in der Natur finde.	○	○	○	○
Poesie beeindruckt mich wenig oder gar nicht.	○	○	○	○
Ich habe wenig Interesse, über die Natur des Universums oder die Lage der Menschheit zu spekulieren.	○	○	○	○
Ich habe oft Spaß daran, mit Theorien oder abstrakten Ideen zu spielen.	○	○	○	○
Ich finde philosophische Diskussionen langweilig.	○	○	○	○
Wenn ich Literatur lese oder ein Kunstwerk betrachte, empfinde ich manchmal ein Frösteln oder eine Welle der Begeisterung.	○	○	○	○

	starke Ablehnung			starke Zustimmung
Ich versuche stets rücksichtsvoll und sensibel zu handeln.	○	○	○	○
Im Hinblick auf die Absichten anderer bin ich eher zynisch und skeptisch.	○	○	○	○
Ich bekomme häufiger Streit mit meiner Familie und meinen Kollegen.	○	○	○	○
Um zu bekommen, was ich will, bin ich notfalls bereit, Menschen zu manipulieren.	○	○	○	○
Manche Leute halten mich für selbstsüchtig und selbstgefällig.	○	○	○	○
Manche Leute halten mich für kalt und berechnend.	○	○	○	○

	starke Ablehnung			starke Zustimmung
Ich kann mir meine Zeit recht gut einteilen, sodass ich meine Angelegenheiten rechtzeitig beende.	○	○	○	○
Ich bin eine tüchtige Person, die ihre Arbeit immer erledigt.	○	○	○	○
Wenn ich eine Verpflichtung eingehe, so kann man sich auf mich bestimmt verlassen.	○	○	○	○
Ich halte meine Sachen ordentlich und sauber.	○	○	○	○
Ich werde wohl niemals fähig sein, Ordnung in mein Leben zu bringen.	○	○	○	○
Ich versuche, alle mir übertragenen Aufgaben sehr gewissenhaft zu erledigen.	○	○	○	○

In welchem Maße treffen die folgenden Aussagen auf Sie zu?

Antworten Sie bitte wahrheitsgemäß.

	trifft überhaupt nicht zu						trifft voll zu
Ich bin bereit, Kosten auf mich zu nehmen, um jemandem zu helfen, der mir früher geholfen hat.	○	○	○	○	○	○	○
Wenn mir jemand einen Gefallen tut, bin ich bereit, dies zu erwidern.	○	○	○	○	○	○	○
Ich stenge mich besonders an, um jemandem zu helfen, der mir früher schon mal geholfen hat.	○	○	○	○	○	○	○

Wenn mich jemand in eine schwierige Lage bringt, werde ich das Gleiche mit ihm machen.	○	○	○	○	○	○	○
Wenn mich jemand beleidigt, werde ich mich ihm gegenüber beleidigend verhalten.	○	○	○	○	○	○	○
Wenn mir schweres Unrecht zuteilwird, werde ich mich um jeden Preis bei der nächsten Gelegenheit dafür rächen.	○	○	○	○	○	○	○

Die nun folgenden Fragen werden Ihnen eventuell etwas komisch vorkommen. Bitte beantworten Sie diese dennoch gewissenhaft und aus dem Bauch heraus. Es gibt auch hier keine richtigen und falschen Antworten.

	stimme überhaupt nicht zu						stimme voll und ganz zu
Das Wohl der Menschen ist wichtiger als die Bedürfnisse aller anderen.	○	○	○	○	○	○	○
Der Mensch ist das letzte Glied in der Evolution der Natur oder, aus religiöser Sicht, „die Krone der Schöpfung“.	○	○	○	○	○	○	○
Der Glaube an die Einzigartigkeit des Menschen ist nur ein von Menschen gemachter Mythos.	○	○	○	○	○	○	○
Nur der Mensch hat die Fähigkeit, das Schöne in der Welt zu sehen.	○	○	○	○	○	○	○
Der Mensch ist ein einzigartiges Wesen, ein ganz besonderes im Universum.	○	○	○	○	○	○	○
Nur der Mensch kann die Welt objektiv kennen lernen, so wie sie ist.	○	○	○	○	○	○	○
Die Befürworter des Umweltschutzes sollten sich daran erinnern, dass das wichtigste Ziel ihres Handelns das Wohl des Menschen sein sollte.	○	○	○	○	○	○	○
Nur der Mensch kann ein „Selbst“ und ein „Innenleben“ haben.	○	○	○	○	○	○	○

Bitte markieren Sie die Sprachassistenten die Sie kennen, oder von denen Sie gehört haben.
Kennen Sie folgende Sprachassistenten?

	nein	ja
Bixby	○	○
Celia	○	○
Xiaowei	○	○
Alice	○	○
Mycroft	○	○
Braina	○	○
Viv	○	○
Google Assistant	○	○
Cortana	○	○
AliGenie	○	○
SILVIA	○	○
Alexa	○	○
Google Now	○	○
Duer	○	○
Siri	○	○
Evi	○	○
Assistant	○	○
Clova	○	○
BlackBerry Assistant	○	○

Abschließend noch kurz Fragen zu Ihnen:

Welches Geschlecht haben Sie?

- ◯ männlich
- ◯ weiblich
- ◯ anderes
- ◯ keine Angabe

Wie alt sind Sie?

[ ]

Wie ist Ihr Familienstand?

- ◯ Single
- ◯ In einer Beziehung
- ◯ Verheiratet
- ◯ Geschieden
- ◯ Verwitwet
- ◯ Sonstiges [ ]

Welche Tätigkeit üben Sie aus?

- ☐ Schüler/in
- ☐ Student/in
- ☐ Auszubildende/r
- ☐ Arbeitnehmer/in
- ☐ Beamte/r
- ☐ Selbständige/r
- ☐ Arbeitslose/r
- ☐ Rentner/in

Was ist Ihr höchster Bildungsabschluss?

- ◯ Hauptschule
- ◯ Mittlere Reife
- ◯ Fachhochschulreife
- ◯ Allgemeine Hochschulreife
- ◯ Hochschulabschluss
- ◯ Kein Schulabschluss
- ◯ Sonstiges: ____________

Seite 13

Manipulation Check

Haben Sie Alexa als hilfreich oder als nicht hilfreich in Ihrer Vorbereitung auf das Quiz wahrgenommen?

[Bitte auswählen]

Sind Ihnen während der Interaktion mit Alexa Fehler aufgefallen?

[Bitte auswählen]

Was denken Sie, was der Zweck der Studie war?

***dieses Feld ist optional**

Hatten Sie den Eindruck, dass der Skill wie beschrieben funktioniert hat?

[Bitte auswählen]

Würden Sie das Quiz zur Verbesserung des Allgemeinwissens weiterempfehlen?

[Bitte auswählen]

References

Abdi, N., Ramokapane, K. M., & Such, J. M. (2019). More than smart speakers: Security and privacy perceptions of smart home personal assistants. *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*, 451–466.

Abrams, D., & Hogg, M. A. (2006). *Social identifications: A social psychology of intergroup relations and group processes*. Routledge.

Akgun, M., Cagiltay, K., & Tzeng, J.-Y. (2005). Computer apology: The effect of the apologetic feedback on users in computerized environment. *Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)*, 254–256.

Alepis, E., & Patsakis, C. (2017). Monkey says, monkey does: Security and privacy on voice assistants. *IEEE Access*, *5*, 17841–17851.

Alexa and Alexa Device FAQs. (n.d.). Retrieved March 1, 2022, from https://www.amazon.co.uk/gp/help/customer/display.html?nodeId=201602230

Appel, J., von der Pütten, A., Krämer, N. C., & Gratch, J. (2012). Does humanity matter? Analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction. *Advances in Human-Computer Interaction*, *2012*.

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

Aronovitch, C. D. (1976). The voice of personality: Stereotyped judgments and their relation to voice quality and sex of speaker. *The Journal of Social Psychology*, *99*(2), 207–220.

Aronson, E., Willerman, B., & Floyd, J. (1966). The effect of a pratfall on increasing interpersonal attractiveness. *Psychonomic Science*, *4*(6), 227–228.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1.

Ashfaq, M., Yun, J., & Yu, S. (2021). My smart speaker is cool! Perceived coolness, perceived values, and users' attitude toward smart speakers. *International Journal of Human–Computer Interaction*, *37*(6), 560–573.

Ashmore, R. D., & Tumia, M. L. (1980). Sex stereotypes and implicit personality theory. I. A personality description approach to the assessment of sex stereotypes. *Sex Roles*, *6*(4), 501–518.

Attaran, M., & Deb, P. (2018). Machine learning: The new'big thing'for competitive advantage. *International Journal of Knowledge Engineering and Data Mining*, *5*(4), 277–305.

Austin, E. J., Deary, I. J., Gibson, G. J., McGregor, M. J., & Dent, J. B. (1998). Individual response spread in self-report scales: Personality correlations and consequences. *Personality and Individual Differences*, *24*(3), 421–438.

Baber, C., & Noyes, J. (2002). *Interactive Speech Technology: Human factors issues in the application of speech input/output to computers*. CRC Press.

Bales, R. F., & Parsons, T. (2014). *Family: Socialization and interaction process*. routledge.

Banaji, M. R. (1993). *The psychology of gender: A perspective on perspectives.*

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191.

Baumeister, R. F., & Tice, D. M. (1985). Self-esteem and responses to success and failure: Subsequent performance and intrinsic motivation. *Journal of Personality*, *53*(3), 450–467.

Bedford-Strohm, J. (2017). Voice First? Eine Analyse des Potentials von intelligenten Sprachassistenten am Beispiel Amazon Alexa. *ComSoc Communicatio Socialis*, *50*(4), 485–494.

Bem, S. L. (1981). Bem sex role inventory. *Journal of Personality and Social Psychology*.

Beniger, J. R. (1986). *The control revolution: Technological and economic origins of the information society*. Harvard University Press.

Bennett, K. (2018). Environment of evolutionary adaptedness (EEA). *Encyclopedia of Personality and Individual Differences*, *1*(1627), 1–3.

Bergen, D. J., & Williams, J. E. (1991). Sex stereotypes in the United States revisited: 1972–1988. *Sex Roles*, *24*(7), 413–423.

Bhatia, N., & Bhatia, S. (2021). Changes in gender stereotypes over time: A computational analysis. *Psychology of Women Quarterly*, *45*(1), 106–125.

Bickerton, D. (2017). *Language and human behavior*. University of Washington Press.

Bickmore, T., & Cassell, J. (2001). Relational agents: A model and implementation of building user trust. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 396–403.

Bickmore, T., & Cassell, J. (2005). Social dialongue with embodied conversational agents. In *Advances in natural multimodal dialogue systems* (pp. 23–54). Springer.

Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology*, *3*(1), 27–52.

Blake, J., & Davis, K. (1964). *Norms, values, and sanctions*.

Blankenship, V., Hnat, S. M., Hess, T. G., & Brown, D. R. (1984). Reciprocal interaction and similarity of personality attributes. *Journal of Social and Personal Relationships*, *1*(4), 415–432.

Boaz, N. T., & Almquist, A. J. (1996). *Biological Anthropology*. Harcourt Brace College Publishers.

Bortz, J., & Schuster, C. (2005). *Statistik für Human-und Sozialwissenschaftler*. Heidelberg: Springer.

Bosco, F. M., Bucciarelli, M., & Bara, B. G. (2006). Recognition and repair of communicative failures: A developmental perspective. *Journal of Pragmatics*, *38*(9), 1398–1429.

Broverman, I. K., Vogel, S. R., Broverman, D., Clarkson, F., & Rosenkrantz, P. (1994). Sex-role stereotypes: A current appraisal. *Caring Voices and Women's Moral Frames: Gilligan's View*, 191–210.

Brown, B. L., Strong, W. J., & Rencher, A. C. (1973). Perceptions of personality from speech: Effects of manipulations of acoustical parameters. *The Journal of the Acoustical Society of America*, *54*(1), 29–35.

Brown, P. (2017). Politeness and impoliteness. *The Oxford Handbook of Pragmatics*, 383–399.

Brown, P., & Levinson, S. C. (1978). Universals in language usage: Politeness phenomena. In *Questions and politeness: Strategies in social interaction* (pp. 56–311). Cambridge University Press.

Burgoon, J. K., Blair, J. P., & Strom, R. E. (2008). Cognitive biases and nonverbal cue availability in detecting deception. *Human Communication Research*, *34*(4), 572–599.

Buss, D. M. (2015). *The new science of the mind*. Boston, MA: Allyn & Bacon.

Byrne, D., Griffitt, W., & Stefaniak, D. (1967). Attraction and similarity of personality characteristics. *Journal of Personality and Social Psychology*, *5*(1), 82.

Cambre, J., & Kulkarni, C. (2019). One voice fits all? Social implications and research challenges of designing voices for smart devices. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–19.

Canbek, N. G., & Mutlu, M. E. (2016). On the track of artificial intelligence: Learning with intelligent personal assistants. *Journal of Human Sciences*, *13*(1), 592–601.

Cantor, N., & Mischel, W. (1979). Prototypes in person perception. In *Advances in experimental social psychology* (Vol. 12, pp. 3–52). Elsevier.

Carolus, A., Binder, J. F., Muench, R., Schmidt, C., Schneider, F., & Buglass, S. L. (2019). Smartphones as digital companions: Characterizing the relationship between users and their phones. *New Media & Society*, *21*(4), 914–938.

Carolus, A., Muench, R., Schmidt, C., & Schneider, F. (2019). Impertinent mobiles-Effects of politeness and impoliteness in human-smartphone interaction. *Computers in Human Behavior*, *93*, 290–300.

Carolus, A., Schmidt, C., Muench, R., Mayer, L., & Schneider, F. (2018). Pink stinks-at least for men. *International Conference on Human-Computer Interaction*, 512–525.

Carolus, A., Schmidt, C., Schneider, F., Mayr, J., & Muench, R. (2018). Are people polite to smartphones? *International Conference on Human-Computer Interaction*, 500–511.

Carolus, A., Schneider, F., Schmidt, C., Muench, R., & Schwab, F. (2016). *Digital Companion: Research Report*.

Cartwright, J. (2000). *Evolution and human behavior: Darwinian perspectives on human nature*. MIT Press.

Cassell, J. (2001). Embodied conversational agents: Representation and intelligence in user interfaces. *AI Magazine*, *22*(4), 67–67.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (2000). Embodied conversational agents. *MS: MIT Press 2000*.

Cathcart, R., & Gumpert, G. (1985). The person-computer interaction: A unique source. *Information and Behavior*, *1*, 113–124.

Ceaparu, I., Lazar, J., Bessiere, K., Robinson, J., & Shneiderman, B. (2004). Determining causes and severity of end-user frustration. *International Journal of Human-Computer Interaction*, *17*(3), 333–356.

Charlesworth, T. E., & Banaji, M. R. (2022). Patterns of implicit and explicit stereotypes III: Long-term change in gender stereotypes. *Social Psychological and Personality Science*, *13*(1), 14–26.

Chattaraman, V., Kwon, W.-S., Gilbert, J. E., & Ross, K. (2019). Should AI-Based, conversational digital assistants employ social-or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. *Computers in Human Behavior*, *90*, 315–330.

Chiasson, S., & Gutwin, C. (2005). Testing the media equation with children. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 829–838.

Chowdhary, K. (2020). Natural language processing. *Fundamentals of Artificial Intelligence*, 603–649.

Chung, A. E., Griffin, A. C., Selezneva, D., & Gotz, D. (2018). Health and fitness apps for hands-free voice-activated assistants: Content analysis. *JMIR MHealth and UHealth*, *6*(9), e9705.

Cialdini, R. B. (2007). *Influence: The psychology of persuasion* (Vol. 55). Collins New York.

Cialdini, R. B., & Trost, M. R. (1998). *Social influence: Social norms, conformity and compliance.*

Clark, L., Doyle, P., Garaialde, D., Gilmartin, E., Schlögl, S., Edlund, J., Aylett, M., Cabral, J., Munteanu, C., & Edwards, J. (2019). The state of speech in HCI: Trends, themes and challenges. *Interacting with Computers*, *31*(4), 349–371.

Clark, L., Ofemile, A., & Cowan, B. R. (2021). Exploring Verbal Uncanny Valley Effects with Vague Language in Computer Speech. In *Voice Attractiveness* (pp. 317–330). Springer.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, *7*(3), 249–253.

Cohen, J. (1988). *Statistical power analysis for the social sciences*.

Cohen, P. R., & Oviatt, S. L. (1995). The role of voice input for human-machine communication. *Proceedings of the National Academy of Sciences*, *92*(22), 9921–9927.

Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, 189–211.

Conway, M., Pizzamiglio, M. T., & Mount, L. (1996). Status, communality, and agency: Implications for stereotypes of gender and other groups. *Journal of Personality and Social Psychology*, *71*(1), 25.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, *163*, 163–228.

Cosmides, L., & Tooby, J. (1997). *Evolutionary psychology: A primer*.

Costrich, N., Feinstein, J., Kidder, L., Marecek, J., & Pascale, L. (1975). When stereotypes hurt: Three studies of penalties for sex-role reversals. *Journal of Experimental Social Psychology*, *11*(6), 520–530.

Cowan, B. R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., Earley, D., & Bandeira, N. (2017). " What can i help you with?" infrequent users' experiences of intelligent personal assistants. *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1–12.

Crawford, J. L., & Haaland, G. A. (1972). Predecisional information seeking and subsequent conformity in the social influence process. *Journal of Personality and Social Psychology*, *23*(1), 112.

Csikszentmihalyi, M., & Csikzentmihaly, M. (1990). *Flow: The psychology of optimal experience* (Vol. 1990). Harper & Row New York.

Csikszentmihalyi, M., & Rathunde, K. (1993). *The measurement of flow in everyday life: Toward a theory of emergent motivation.*

Culpeper, J. (1996). Towards an anatomy of impoliteness. *Journal of Pragmatics*, *25*(3), 349–367.

Curry, A. C., Robertson, J., & Rieser, V. (2020). Conversational assistants and gender stereotypes: Public perceptions and desiderata for voice personas. *Proceedings of the Second Workshop on Gender Bias in Natural Language Processing*, 72–78.

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly*, *64*(4), 413–428.

Cutting, J. E. (1974). Two left-hemisphere mechanisms in speech perception. *Perception & Psychophysics*, *16*(3), 601–612.

Davidson, R., & MacKinnon, J. G. (1993). *Estimation and inference in econometrics* (Vol. 63). Oxford New York.

Dawkins, R., & Davis, N. (2017). *The selfish gene*. Macat Library.

Deaux, K., & LaFrance, M. (1998). *The handbook of social psychology*.

Della Cava, M. (2016, March 30). Microsoft CEO Nadella: “Bots are the new apps.” *USA Tech Today*. https://eu.usatoday.com/story/tech/news/2016/03/30/microsof-ceo-nadella-bots-new-apps/82431672/

Dennett, D. (1989). *The Intentional Stance*. MIT press. https://doi.org/10.2307/2026682

Detenber, B. H., & Reeves, B. (1996). A bio-informational theory of emotion: Motion and image size effects on viewers. *Journal of Communication*.

Deutsch, C. J., & Gilbert, L. A. (1976). Sex role stereotypes: Effect on perceptions of self and others and on personal adjustment. *Journal of Counseling Psychology*, *23*(4), 373.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *41*(1), 417–440.

Dindia, K., Fitzpatrick, M. A., & Kenny, D. A. (1997). Self-disclosure in spouse and stranger interaction: A social relations analysis. *Human Communication Research*, *23*(3), 388–412.

Dragicevic, P. (2016). Fair statistical communication in HCI. In *Modern statistical methods for HCI* (pp. 291–330). Springer.

Duck, S. W., & Craig, G. (1978). Personality similarity and the development of friendship: A longitudinal study. *British Journal of Social and Clinical Psychology*, *17*(3), 237–242.

Duffy, B. R., & Zawieska, K. (2012). Suspension of disbelief in social robotics. *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, 484–489.

Dybkjaer, L., Bernsen, N. O., & Minker, W. (2004). Evaluation and usability of multimodal spoken language dialogue systems. *Speech Communication*, *43*(1–2), 33–54.

Eagly, A. H. (1978). Sex differences in influenceability. *Psychological Bulletin*, *85*(1), 86.

Eagly, A. H. (1983). Gender and social influence: A social psychological analysis. *American Psychologist*, *38*(9), 971.

Eagly, A. H. (1987). *Sex differences in sexual behavior: A social-role interpretation*. Hillsdale, NJ: Lawrence Erbaum.

Eagly, A. H., & Carli, L. L. (1981). Sex of researchers and sex-typed communications as determinants of sex differences in influenceability: A meta-analysis of social influence studies. *Psychological Bulletin*, *90*(1), 1.

Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin*, *100*(3), 283.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of US public opinion polls from 1946 to 2018. *American Psychologist*, *75*(3), 301.

Eagly, A. H., & Wood, W. (1982). Inferred sex differences in status as a determinant of gender stereotypes about social influence. *Journal of Personality and Social Psychology*, *43*(5), 915.

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. *The Developmental Social Psychology of Gender*, *12*, 174.

Eckes, T. (1997). *Geschlechterstereotype: Frau und Mann in sozialpsychologischer Sicht*. Centaurus-Verlag-Ges.

Eckes, T. (2008). Geschlechterstereotype: Von Rollen, Identitäten und Vorurteilen. In *Handbuch Frauen-und Geschlechterforschung* (pp. 171–182). Springer.

Eckles, D., Wightman, D., Carlson, C., Thamrongrattanarit, A., Bastea-Forte, M., & Fogg, B. J. (2009). Social responses in mobile messaging: Influence strategies, self-disclosure, and source orientation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1651–1654.

Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und Forschungsmethoden*.

Ekman, P., & Friesen, W. V. (1978). Facial action coding system. *Environmental Psychology & Nonverbal Behavior*.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864.

Ernst, C.-P., & Herm-Stapelberg, N. (2020). Gender Stereotyping's Influence on the Perceived Competence of Siri and Co. *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

Eyssel, F., & Hegel, F. (2012). (s) he's got the look: Gender stereotyping of robots 1. *Journal of Applied Social Psychology*, *42*(9), 2213–2230.

Fedor, C.-G. (2014). Stereotypes and Prejudice in the Perception of the “Other.” *Procedia-Social and Behavioral Sciences*, *149*, 321–326.

Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, *14*(3), 159–181.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. sage.

Finkel, S. E., Guterbock, T. M., & Borg, M. J. (1991). Race-of-interviewer effects in a preelection poll Virginia 1989. *Public Opinion Quarterly*, *55*(3), 313–330.

Finneran, C. M., & Zhang, P. (2003). A person–artefact–task (PAT) model of flow antecedents in computer-mediated environments. *International Journal of Human-Computer Studies*, *59*(4), 475–496.

Fischer, K. (2011). Interpersonal variation in understanding robots as social actors. *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 53–60.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2018). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. In *Social cognition* (pp. 162–214). Routledge.

Fitch, W. (2017). Empirical approaches to the study of language evolution. *Psychonomic Bulletin & Review*, *24*(1), 3–33.

Flanagan, J. L. (1972). Voices of men and machines. *The Journal of the Acoustical Society of America*, *51*(5A), 1375–1387.

Fogg, B. J. (1998). Persuasive computers: Perspectives and research directions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 225–232.

Fogg, B. J. (2002). Persuasive technology: Using computers to change what we think and do. *Ubiquity*, *2002*(December), 2.

Fogg, B. J., & Nass, C. (1997a). How users reciprocate to computers: An experiment that demonstrates behavior change. In *CHI'97 extended abstracts on Human factors in computing systems* (pp. 331–332).

Fogg, B. J., & Nass, C. (1997b). Silicon sycophants: The effects of computers that flatter. *International Journal of Human-Computer Studies*, *46*(5), 551–561.

Fogg, B. J., & Tseng, H. (1999). The elements of computer credibility. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 80–87.

Fortuna, P., Wróblewski, Z., & Gorbaniuk, O. (2021). The structure and correlates of anthropocentrism as a psychological construct. *Current Psychology*, 1–13.

Francis, G. (2012). Publication bias and the failure of replication in experimental psychology. *Psychonomic Bulletin & Review*, *19*(6), 975–991.

Friedman, B., & Kahn Jr, P. H. (1992). Human agency and responsible computing: Implications for computer system design. *Journal of Systems and Software*, *17*(1), 7–14.

Funder, D. C. (1997). *The personality puzzle.* WW Norton & Co.

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, *1*, 71–85.

Gardiner, A. H. (1932). *The theory of speech and language.*

Ghazali, A. S., Ham, J., Barakova, E., & Markopoulos, P. (2018). The influence of social cues in persuasive social robots on psychological reactance and compliance. *Computers in Human Behavior*, *87*, 58–65.

Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards Designing Cooperative and Social Conversational Agents for Customer Service. *ICIS*.

Goffmann, E. (1967). *Interaction Ritual: Essays on Face-to-Face Encounter*. Garden City, NY: Anchor Books.

Goldstein, M., Werdenhoff, J., & others. (2002). The media equation does not always apply: People are not polite towards small computers. *Personal and Ubiquitous Computing*, *6*(2), 87–96.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 161–178.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619–619.

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125–130.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Gruhn, R. E., Minker, W., & Nakamura, S. (2011). Automatic speech recognition. In *Statistical pronunciation modeling for non-native speech processing* (pp. 5–17). Springer.

Gunkel, D. J. (2012). Communication and artificial intelligence: Opportunities and challenges for the 21st century. *Communication+ 1*, *1*(1), 1–25.

Guzman, A. L. (2015). *Imagining the voice in the machine: The ontology of digital social agents* [PhD Thesis]. University of Illinois at Chicago.

Guzman, A. L. (2019). Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior*, *90*, 343–350.

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A Human–Machine Communication research agenda. *New Media & Society*, *22*(1), 70–86.

Habler, F., Schwind, V., & Henze, N. (2019). Effects of Smart Virtual Assistants' Gender and Language. In *Proceedings of Mensch und Computer 2019* (pp. 469–473).

Haines, E. L., Deaux, K., & Lofaro, N. (2016). The times they are a-changing… or are they not? A comparison of gender stereotypes, 1983–2014. *Psychology of Women Quarterly*, *40*(3), 353–363.

Hammer, S., Lugrin, B., Bogomolov, S., Janowski, K., & André, E. (2016). Investigating politeness strategies and their persuasiveness for a robotic elderly assistant. *International Conference on Persuasive Technology*, 315–326.

Han, S., & Yang, H. (2018). Understanding adoption of intelligent personal assistants: A parasocial relationship perspective. *Industrial Management & Data Systems*.

Hare, B. (2017). Survival of the friendliest: Homo sapiens evolved via selection for prosociality. *Annual Review of Psychology*, *68*, 155–186.

Hashemi, S. H., Williams, K., Kholy, A. E., Zitouni, I., & Crook, P. A. (2018). Measuring user satisfaction on smart speaker intelligent assistants. *Anne Dirkson, Suzan Verberne, Gerard van Oortmerssen & Wessel Kraaij*, 22.

Hauser, M. D. (1996). *The evolution of communication*. MIT press.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579.

Hayes, A. F. (2018). Introduction to mediation, moderation, and conditional process analysis second edition: A regression-based approach. *New York, NY: Ebook The Guilford Press. Google Scholar*.

Heckman, C. E., & Wobbrock, J. O. (2000). Put your best face forward: Anthropomorphic agents, e-commerce consumers, and the law. *Proceedings of the Fourth International Conference on Autonomous Agents*, 435–442.

Helmreich, R., Aronson, E., & LeFan, J. (1970). To err is humanizing sometimes: Effects of self-esteem, competence, and a pratfall on interpersonal attraction. *Journal of Personality and Social Psychology*, *16*(2), 259.

Helmreich, R. L., Spence, J. T., & Wilhelm, J. A. (1981). A psychometric analysis of the Personal Attributes Questionnaire. *Sex Roles*, *7*(11), 1097–1108.

Hertz, N., & Wiese, E. (2016). Influence of agent type and task ambiguity on conformity in social decision making. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 313–317.

Hilton, J. L., & Von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, *47*(1), 237–271.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science, 349*(6245), 261–266.

Hoffmann, L., Krämer, N. C., Lam-Chi, A., & Kopp, S. (2009). Media equation revisited: Do users show polite reactions towards an embodied agent? *International Workshop on Intelligent Virtual Agents,* 159–165.

Hogg, M. A., & Vaughan, G. M. (1995). *Social psychology: An introduction.* Harvester Wheatsheaf.

Holtgraves, T., & Joong-Nam, Y. (1990). Politeness as universal: Cross-cultural perceptions of request strategies and inferences based on their use. *Journal of Personality and Social Psychology*, *59*(4), 719.

Holtgraves, T. M. (2013). *Language as social action: Social psychology and language use*. Psychology Press.

Horstmann, A. C., & Krämer, N. C. (2019). Great expectations? Relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment. *Frontiers in Psychology*, *10*, 939.

Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: An introduction to voice assistants. *Medical Reference Services Quarterly*, *37*(1), 81–88.

Hussy, W., & Jain, A. (2002). *Experimental Hypothesis Testing in Psychology*. Hogrefe Verlag für Psychologie, Göttingen.

Iacobucci, D., Posavac, S. S., Kardes, F. R., Schneider, M. J., & Popovich, D. L. (2015a). The median split: Robust, refined, and revived. *Journal of Consumer Psychology*, *25*(4), 690–704.

Iacobucci, D., Posavac, S. S., Kardes, F. R., Schneider, M. J., & Popovich, D. L. (2015b). Toward a more nuanced understanding of the statistical properties of a median split. *Journal of Consumer Psychology*, *25*(4), 652–665.

Jackson, J. M. (1960). Structural characteristics of norms. *Teachers College Record*, *61*(10), 136–163.

Jang, Y. (2020). Exploring user interaction and satisfaction with virtual personal assistant usage through smart speakers. *Archives of Design Research*, *33*(3), 127–135.

Jia, R., & Jia, H. (2012). *Computer playfulness, openness to experience, and computer loafing*.

Johnson, D. (2006). *Exploring mindlessness as an explanation for the media equation: Why and when people will treat computers as social actors*.

Johnson, D., & Gardner, J. (2007). The media equation and team formation: Further evidence for experience as a moderator. *International Journal of Human-Computer Studies*, *65*(2), 111–124.

Johnson, D., & Gardner, J. (2009). Exploring mindlessness as an explanation for the media equation: A study of stereotyping in computer tutorials. *Personal and Ubiquitous Computing*, *13*(2), 151–163.

Johnson, D., Gardner, J., & Wiles, J. (2004). Experience as a moderator of the media equation: The impact of flattery and praise. *International Journal of Human-Computer Studies*, *61*(3), 237–258.

Jones, E. E. (1964). *Ingratiation.*

Jones, E. E. (1990). *Interpersonal perception.* WH Freeman/Times Books/Henry Holt & Co.

Juang, B.-H., & Rabiner, L. R. (2005). Automatic speech recognition–a brief history of the technology development. *Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara*, *1*, 67.

Kamm, C., Walker, M., & Rabiner, L. (1997). The role of speech processing in human–computer intelligent communication. *Speech Communication*, *23*(4), 263–278.

Karp, D., Jin, N., Yamagishi, T., & Shinotsuka, H. (1993). Raising the minimum in the minimal group paradigm. *The Japanese Journal of Experimental Social Psychology, 32*(3), 231–240.

Karr-Wisniewski, P., & Prietula, M. (2010). CASA, WASA, and the dimensions of us. *Computers in Human Behavior*, *26*(6), 1761–1771.

Katagiri, Y., Nass, C., & Takeuchi, Y. (2001). Cross-cultural studies of the computers are social actors paradigm: The case of reciprocity. *Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents, and Virtual Reality*, 1558–1562.

Kelman, H. C. (1958). Compliance, identification, and internalization three processes of attitude change. *Journal of Conflict Resolution*, *2*(1), 51–60.

Kess, J. F. (1992). *Psycholinguistics: Psychology, linguistics, and the study of natural language* (Vol. 86). John Benjamins Publishing.

Kiesler, S., & Sproull, L. S. (1986). Response effects in the electronic survey. *Public Opinion Quarterly*, *50*(3), 402–413.

Kim, D., Park, K., Park, Y., Ju, J., & Ahn, J.-H. (2018). Alexa, tell me more: The effect of advertisements on memory accuracy from smart speakers. *22nd Pacific Asia Conference on Information Systems (PACIS 2018)*, 204.

Kim, K. J. (2014). Can smartphones be specialists? Effects of specialization in mobile advertising. *Telematics and Informatics*, *31*(4), 640–647.

King, B., Chen, I.-F., Vaizman, Y., Liu, Y., Maas, R., Parthasarathi, S. H. K., & Hoffmeister, B. (2017). Robust Speech Recognition via Anchor Word Representations. *Interspeech*, 2471–2475.

Kiseleva, J., Williams, K., Jiang, J., Hassan Awadallah, A., Crook, A. C., Zitouni, I., & Anastasakos, T. (2016). Understanding user satisfaction with intelligent assistants. *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, 121–130.

Klein, J., Moon, Y., & Picard, R. W. (1999). This computer responds to user frustration: Theory, design, results, and implications (Vison and Modeling Technical Reports Nr. 501). *Boston: Massachusetts Institute of Technology*.

Koh, Y. J., & Sundar, S. S. (2010). Effects of specialization in computers, web sites, and web agents on e-commerce trust. *International Journal of Human-Computer Studies*, *68*(12), 899–912.

Kohler, K. J. (2017). *Communicative Functions and Linguistic Forms in Speech Interaction: Volume 156* (Vol. 156). Cambridge University Press.

Kopp, S., Gesellensetter, L., Krämer, N. C., & Wachsmuth, I. (2005). A conversational agent as museum guide–design and evaluation of a real-world application. *International Workshop on Intelligent Virtual Agents*, 329–343.

Körner, A., Geyer, M., Roth, M., Drapeau, M., Schmutzer, G., Albani, C., Schumann, S., & Brähler, E. (2008). Persönlichkeitsdiagnostik mit dem neo-fünf-faktoren-inventar: Die 30-item-kurzversion (neo-ffi-30). *PPmP-Psychotherapie· Psychosomatik· Medizinische Psychologie*, *58*(06), 238–245.

Kowalczuk, P. (2018). Consumer acceptance of smart speakers: A mixed methods approach. *Journal of Research in Interactive Marketing*.

Krämer, N., & Manzeschke, A. (2021). Social reactions to socially interactive agents and their ethical implications. In *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition* (pp. 77–104).

Laitman, J. T. (1984). The anatomy of human speech. *Natural History*, *93*(8), 20–27.

Langer, E. J. (1989). Mindfulness and mindlessness. *The Production of Reality: Essays and Readings on Social Interaction*, 153–157.

Langer, E. J., & Abelson, R. P. (1972). The semantics of asking a favor: How to succeed in getting help without really dying. *Journal of Personality and Social Psychology*, *24*(1), 26.

Langer, E. J., Chanowitz, B., & Blank, A. (1985). *Mindlessness–mindfulness in perspective: A reply to Valerie Folkes.*

Langer, E. J., & Moldoveanu, M. (2000). Mindfulness research and the future. *Journal of Social Issues*, *56*(1), 129–139.

Larzelere, R. E., Kuhn, B. R., & Johnson, B. (2004). The intervention selection bias: An underrecognized confound in intervention research. *Psychological Bulletin*, *130*(2), 289.

Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–31.

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. *Symposium of the Austrian HCI and Usability Engineering Group*, 63–76.

Leary, M. R., & Cox, C. B. (2008). *Belongingness motivation: A mainspring of social action.*

Lee, E. J. (2003). Effects of "gender" of the computer on informational social influence: The moderating role of task type. *International Journal of Human-Computer Studies*, *58*(4), 347–362.

Lee, E. J., & Nass, C. (1998). Does the ethnicity of a computer agent matter? An experimental comparison of human-computer interaction and computer-mediated communication. *Proceedings of the 1998 Workshop on Embodied Conversational Characters*.

Lee, E., Nass, C., & Brave, S. (2000). Can computer-generated speech have gender?: An experimental test of gender stereotype. *CHI'00 Extended Abstracts on Human Factors …*, *April*, 289–290. http://doi.acm.org/10.1145/633292.633461

Lee, E.-J. (2008). Gender stereotyping of computers: Resource depletion or reduced attention? *Journal of Communication*, *58*(2), 301–320.

Lee, E.-J., & Nass, C. (2002). Experimental tests of normative group influence and representation effects in computer-mediated communication: When interacting via computers differs from interacting with computers. *Human Communication Research*, *28*(3), 349–381.

Lee, H., & Cho, C.-H. (2020). Uses and gratifications of smart speakers: Modelling the effectiveness of smart speaker advertising. *International Journal of Advertising*, *39*(7), 1150–1171.

Lee, K. M. (2004). Why presence occurs: Evolutionary psychology, media equation, and presence. *Presence: Teleoperators & Virtual Environments*, *13*(4), 494–505.

Lee, K. M., & Jung, Y. (2005). Evolutionary nature of virtual experience. *Journal of Cultural and Evolutionary Psychology*, *3*(2), 159–176.

Lee, K. M., & Nass, C. (2003). Designing social presence of social actors in human computer interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 289–296.

Lee, K. M., & Nass, C. (2004). The multiple source effect and synthesized speech: Doubly-disembodied language as a conceptual framework. *Human Communication Research*, *30*(2), 182–207.

Lee, M. K., Kiesler, S., Forlizzi, J., Srinivasa, S., & Rybski, P. (2010). Gracefully mitigating breakdowns in robotic services. *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 203–210.

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends in Cognitive Sciences*, *20*(1), 6–14.

Li, J., Deng, L., Gong, Y., & Haeb-Umbach, R. (2014). An overview of noise-robust automatic speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *22*(4), 745–777.

Li, X., & Sung, Y. (2021). Anthropomorphism brings us closer: The mediating role of psychological distance in User–AI assistant interactions. *Computers in Human Behavior*, *118*, 106680.

Liao, S., Wilson, C., Cheng, L., Hu, H., & Deng, H. (2020). Measuring the effectiveness of privacy policies for voice assistant applications. *Annual Computer Security Applications Conference*, 856–869.

Liao, Y., Vitak, J., Kumar, P., Zimmer, M., & Kritikos, K. (2019). Understanding the role of privacy and trust in intelligent personal assistant adoption. *International Conference on Information*, 102–113.

Lieberman, H. (1997). Autonomous interface agents. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 67–74.

Lieberman, P. (1998). Speech evolution: Let barking dogs sleep. *Behavioral and Brain Sciences*, *21*(4), 520–521.

Ling, H.-C., Chen, H.-R., Ho, K. K., & Hsiao, K.-L. (2021). Exploring the factors affecting customers' intention to purchase a smart speaker. *Journal of Retailing and Consumer Services*, *59*, 102331.

Liu, B., & Sundar, S. S. (2018). Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychology, Behavior, and Social Networking*, *21*(10), 625–636.

Lockheed, M. E. (1985). Women, girls, and computers: A first look at the evidence. *Sex Roles*, *13*(3), 115–122.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, *3*(2), JCMC321.

Lombard, M., & Xu, K. (2021). Social responses to media technologies in the 21st century: The media are social actors paradigm. *Human-Machine Communication*, *2*, 29–55.

Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., Sorsche, P., Hirsch, D., Li, Q., & Martinez, A. (2019). Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science*, *51*(4), 984–997.

Lopatovska, I., & Williams, H. (2018). Personification of the Amazon Alexa: BFF or a mindless companion. *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, 265–268.

Luger, E., & Sellen, A. (2016). " Like Having a Really Bad PA" The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297.

Lugrin, B. (2021). Introduction to Socially Interactive Agents. In *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition* (pp. 1–20).

Maccoby, E. E. (1998). *The two sexes: Growing up apart, coming together* (Vol. 4). Harvard University Press.

Mackie, D. M. (1986). Social identification effects in group polarization. *Journal of Personality and Social Psychology*, *50*(4), 720.

Martin, C. L., & Ruble, D. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current Directions in Psychological Science*, *13*(2), 67–70.

Massaro, D. W., & Cohen, M. M. (1995). Perceiving talking faces. *The Journal of the Acoustical Society of America*, *97*(5), 3308. https://doi.org/10.1121/1.412931

Matthews, B. A., Baker, F., & Spillers, R. L. (2003). How true is true? Assessing socially desirable response bias. *Quality and Quantity*, *37*(3), 327–335.

McCroskey, J. C., Hamilton, P. R., & Weiner, A. N. (1974). The effect of interaction behavior on source credibility, homophily, and interpersonal attraction. *Human Communication Research*, *1*(1), 42–52.

McTear, M. F., Callejas, Z., & Griol, D. (2016). *The conversational interface* (Vol. 6). Springer.

Menne, I. M. (2020). *Facing Social Robots: Emotional Reactions Towards Social Robots*. BoD–Books on Demand.

Meyer, W.-U., Mittag, W., & Engler, U. (1986). Some effects of praise and blame on perceived ability and affect. *Social Cognition*, *4*(3), 293–308.

Mieczkowski, H., Liu, S. X., Hancock, J., & Reeves, B. (2019). Helping not hurting: Applying the stereotype content model and BIAS map to social robotics. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 222–229.

Miller, C. L., Younger, B. A., & Morse, P. A. (1982). The categorization of male and female voices in infancy. *Infant Behavior and Development*, *5*(2–4), 143–159.

Mirnig, N., Stollnberger, G., Miksch, M., Stadler, S., Giuliani, M., & Tscheligi, M. (2017). To err is robot: How humans assess and act toward an erroneous social robot. *Frontiers in Robotics and AI*, 21.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, *16*(4), 495–500.

Moon, Y. (1998). The effects of distance in local versus remote human-computer interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 103–108.

Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research*, *26*(4), 323–339.

Moon, Y., & Nass, C. (1996a). How “real” are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research*, *23*(6), 651–674.

Moon, Y., & Nass, C. (1998). Are computers scapegoats? Attributions of responsibility in human–computer interaction. *International Journal of Human-Computer Studies*, *49*(1), 79–94.

Moon, Y., & Nass, C. I. (1996b). Adaptive agents and personality change: Complementarity versus similarity as forms of adaptation. *Conference Companion on Human Factors in Computing Systems*, 287–288.

Moore, R. K. (2007). Spoken language processing: Piecing together the puzzle. *Speech Communication*, *49*(5), 418–435.

Moore, R. K. (2013). Spoken language processing: Where do we go from here? In *Your Virtual Butler* (pp. 119–133). Springer.

Morishima, Y., Bennett, C., Nass, C., & Lee, K. M. (2002). Effects of (Synthetic) Voice Gender, User Gender, and Product Gender on Credibility in E-Commerce. *Unpublished Manuscript, Stanford, CA: Stanford University*.

Morkes, J., Kernal, H. K., & Nass, C. (1998). Humor in task-oriented computer-mediated communication and human-computer interaction. *CHI 98 Conference Summary on Human Factors in Computing Systems*, 215–216.

Morkes, J., Kernal, H. K., & Nass, C. (1999). Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of SRCT theory. *Human-Computer Interaction*, *14*(4), 395–435.

Mošner, L., Wu, M., Raju, A., Parthasarathi, S. H. K., Kumatani, K., Sundaram, S., Maas, R., & Hoffmeister, B. (2019). Improving noise robustness of automatic speech recognition via parallel data and teacher-student learning. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6475–6479.

Mueller, A., Kuester, S., & Janda, S. von. (2022). Not so intelligent after all–consumer perceptions of AI-induced errors. *Proceedings/AMA Winter Academic Conference*, *33*, 121–123.

Mühlenbernd, R., Wacewicz, S., & Żywiczyński, P. (2021). Politeness and reputation in cultural evolution. *Linguistics and Philosophy*, *44*(6), 1181–1213.

Mullennix, J. W., Johnson, K. A., Topcu-Durgun, M., & Farnsworth, L. M. (1995). The perceptual representation of voice gender. *The Journal of the Acoustical Society of America*, *98*(6), 3080–3095.

Mullennix, J. W., Stern, S. E., Wilson, S. J., & Dyson, C. (2003). Social perception of male and female computer synthesized speech. *Computers in Human Behavior*, *19*(4), 407–424.

Mumm, J., & Mutlu, B. (2011). Designing motivational agents: The role of praise, social comparison, and embodiment in computer feedback. *Computers in Human Behavior*, *27*(5), 1643–1650.

Murad, C., Munteanu, C., Clark, L., & Cowan, B. R. (2018). Design guidelines for hands-free speech interaction. *Proceedings of the 20th International Conference on*

Human-Computer Interaction with Mobile Devices and Services Adjunct, 269–276.

Nakanishi, H., Nakazawa, S., Ishida, T., Takanashi, K., & Isbister, K. (2003). Can software agents influence human relations? Balance theory in agent-mediated communities. *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, 717–724.

Nass, C. (2004). Etiquette equality: Exhibitions and expectations of computer politeness. *Communications of the ACM*, *47*(4), 35–37.

Nass, C., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. MIT press Cambridge.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, *45*(6), 669–678.

Nass, C., Fogg, B., & Moon, Y. (1995). How powerful is social identity? Affiliation effects in human-computer interaction. *Stanford University Communications Department, Stanford April*, *23*, 1995.

Nass, C., & Gong, L. (2000). Speech interfaces from an evolutionary perspective. *Communications of the ACM*, *43*(9), 36–43.

Nass, C., Isbister, K., & Lee, E.-J. (2001). Truth is beauty: Researching embodied conversational agents. In *Embodied conversational agents* (pp. 374–402).

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, *7*(3), 171.

Nass, C., Lombard, M., Henriksen, L., & Steuer, J. (1995). Anthropocentrism and computers. *Behaviour & Information Technology*, *14*(4), 229–238.

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Moon, Y., & Carney, P. (1999). Are People Polite to Computers? Responses to Computer-Based Interviewing Systems1. *Journal of Applied Social Psychology*, *29*(5), 1093–1109. https://doi.org/10.1111/j.1559-1816.1999.tb00142.x

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, *43*(2), 223–239.

Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, *27*(10), 864–876.

Nass, C., & Steuer, J. (1993). Voices, boxes, and sources of messages: Computers and social actors. *Human Communication Research*, *19*(4), 504–527.

Nass, C., Steuer, J., Henriksen, L., & Dryer, D. C. (1994). Machines, social attributions, and ethopoeia: Performance assessments of computers subsequent to" self-" or" other-" evaluations. *International Journal of Human-Computer Studies*, *40*(3), 543–559.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Conference Companion on Human Factors in Computing Systems - CHI '94*, 204. https://doi.org/10.1145/259963.260288

Nass, C., & Yen, C. (2010). *The man who lied to his laptop: What we can learn about ourselves from our machines*. Penguin.

Nease, A. A., Mudgett, B. O., & Quiñones, M. A. (1999). Relationships among feedback sign, self-efficacy, and acceptance of performance feedback. *Journal of Applied Psychology*, *84*(5), 806.

Newell, A. (1994). *Unified theories of cognition*. Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104). Prentice-hall Englewood Cliffs, NJ.

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, *45*(3), 211–236.

Nomura, T., & Saeki, K. (2009). Effects of polite behaviors expressed by robots: A case study in japan. *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, *2*, 108–114.

Olive, J. P. (1997). The talking computer: Text to speech synthesis. *HAL's Legacy: 2001's Computer as Dream and Reality*.

Oulasvirta, A., Engelbrecht, K.-P., Jameson, A., & Möller, S. (2006). The relationship between user errors and perceived usability of a spoken dialogue system. *The 2nd ISCA/DEGA Tutorial & Research Workshop on Perceptual Quality of Systems*.

Oviatt, S., MacEachern, M., & Levow, G.-A. (1998). Predicting hyperarticulate speech during human-computer error resolution. *Speech Communication*, *24*(2), 87–110.

Oviatt, S., & VanGent, R. (1996). Error resolution during multimodal human-computer interaction. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, *1*, 204–207.

Park, D., & Namkung, K. (2021). Exploring Users' Mental Models for Anthropomorphized Voice Assistants through Psychological Approaches. *Applied Sciences*, *11*(23), 11147.

Park, K., Kwak, C., Lee, J., & Ahn, J.-H. (2018). The effect of platform characteristics on the adoption of smart speakers: Empirical evidence in South Korea. *Telematics and Informatics*, *35*(8), 2118–2132.

Payr, S. (2013). Virtual butlers and real people: Styles and practices in long-term use of a companion. In *Your Virtual Butler* (pp. 134–178). Springer.

Pearce, J. (2005). Engaging the learner: How can the flow experience support e-learning? *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, 2288–2295.

Peña, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: An optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences*, *100*(20), 11702–11705.

Pettigrew, T. F., & Tropp, L. R. (2013). Does intergroup contact reduce prejudice? Recent meta-analytic findings. In *Reducing prejudice and discrimination* (pp. 103–124). Psychology Press.

Pfeifer, L. M., & Bickmore, T. (2011). Is the media equation a flash in the pan? The durability and longevity of social responses to computers. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 777–780.

Pfeifle, A. (2018). Alexa, what should we do about privacy: Protecting privacy for users of voice-activated devices. *Wash. L. Rev.*, *93*, 421.

Picard, R. W., & Healey, J. (1997). Affective wearables. *Personal Technologies*, *1*(4), 231–240.

Pieraccini, R. (2021). Natural Language Understanding in Socially Interactive Agents. In *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied*

Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition (pp. 147–172).

Pinker, S. (1995). *The language instinct: The new science of language and mind* (Vol. 7529). Penguin UK.

Pittam, J. (1994). *Voice in social interaction* (Vol. 5). Sage.

Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice interfaces in everyday life. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12.

Pradhan, A., Findlater, L., & Lazar, A. (2019). " Phantom Friend" or" Just a Box with Information" Personification and Ontological Categorization of Smart Speaker-based Voice Assistants by Older Adults. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–21.

Pradhan, A., Mehta, K., & Findlater, L. (2018). " Accessibility Came by Accident" Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13.

Pratt, J. A., Hauser, K., Ugray, Z., & Patterson, O. (2007). Looking at human–computer interface design: Effects of ethnicity in computer agents. *Interacting with Computers*, *19*(4), 512–523.

Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). " Alexa is my new BFF" Social Roles, User Satisfaction, and Personification of the Amazon Echo. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2853–2859.

Rabbie, J. M., Schot, J. C., & Visser, L. (1989). Social identity theory: A conceptual and empirical critique from the perspective of a behavioural interaction model. *European Journal of Social Psychology, 19*(3), 171–202.

Rabiner, L. R., & Juang, B.-H. (2006). Speech recognition: Statistical methods. *Encyclopedia of Language & Linguistics*, 1–18.

Ragni, M., Rudenko, A., Kuhnert, B., & Arras, K. O. (2016). Errare humanum est: Erroneous robots in human-robot interaction. *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 501–506.

Reeves, B., Detenber, B., & Steuer, J. (1993). *New television: The effects of big pictures and big sound on viewer responses to the screen*.

Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK, 10*, 236605.

Reeves, B., Thorson, E., Rothschild, M. L., McDonald, D., Hirsch, J., & Goldstein, R. (1985). Attention to television: Intrastimulus effects of movement and scene changes on alpha variation over time. *International Journal of Neuroscience, 27*(3–4), 241–255.

Regan, D. T. (1971). Effects of a favor and liking on compliance. *Journal of Experimental Social Psychology, 7*(6), 627–639.

Reuten, A., Van Dam, M., & Naber, M. (2018). Pupillary responses to robotic and human emotions: The uncanny valley and media equation confirmed. *Frontiers in Psychology, 9*, 774.

Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., & Eimler, S. C. (2013). An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics, 5*(1), 17–34.

Runge, T. E., Frey, D., Gollwitzer, P. M., Helmreich, R. L., & Spence, J. T. (1981). Masculine (instrumental) and feminine (expressive) traits: A comparison between students in the United States and West Germany. *Journal of Cross-Cultural Psychology*, *12*(2), 142–162.

Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 1–8.

Salem, M., Ziadee, M., & Sakr, M. (2013). Effects of politeness and interaction context on perception and experience of HRI. *International Conference on Social Robotics*, 531–541.

Sandoval, E. B., Brandstetter, J., & Bartneck, C. (2016). Can a robot bribe a human? The measurement of the negative side of reciprocity in human robot interaction. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 117–124.

Sandoval, E. B., Brandstetter, J., Obaid, M., & Bartneck, C. (2016). Reciprocity in human-robot interaction: A quantitative approach through the prisoner's dilemma and the ultimatum game. *International Journal of Social Robotics*, *8*(2), 303–317.

Sax, L. J. (2001). *Career strategies for women in academe: Arming Athena*. Taylor & Francis.

Schafer, R. W. (1995). Scientific bases of human-machine communication by voice. *Proceedings of the National Academy of Sciences*, *92*(22), 9914–9920.

Schank, R. C., & Abelson, R. P. (2013). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Psychology Press.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, *32*(1), 76–92.

Schuetzler, R. M., Giboney, J. S., Grimes, G. M., & Nunamaker Jr, J. F. (2018). The influence of conversational agent embodiment and conversational relevance on socially desirable responding. *Decision Support Systems*, *114*, 94–102.

Schuetzler, R. M., Grimes, M., Giboney, J. S., & Buckman, J. (2014). *Facilitating natural conversational agent interactions: Lessons from a deception experiment*.

Schwarzer, R., & Jerusalem, M. (1999). Skalen zur Erfassung von Lehrer-und Schülermerkmalen. *Dokumentation Der Psychometrischen Verfahren Im Rahmen Der Wissenschaftlichen Begleitung Des Modellversuchs Selbstwirksame Schulen. Berlin: Freie Universität Berlin*, *144*.

Scott, T. (2021, December 22). Smart Speakers Statistics: Report 2022 [Blog]. *Speakergy*. https://speakergy.com/smart-speakers-statistics/

Sczesny, S., Nater, C., & Eagly, A. H. (2018). Agency and communion: Their implications for gender stereotypes and gender identities. In *Agency and communion in social psychology* (pp. 103–116). Routledge.

Seaborn, K., Miyake, N. P., Pennefather, P., & Otake-Matsuura, M. (2021). Voice in Human–Agent Interaction: A Survey. *ACM Computing Surveys (CSUR)*, *54*(4), 1–43.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424.

Selker, T. (1994). Coach: A teaching agent that learns. *Communications of the ACM*, *37*(7), 92–99.

Shank, D. B., Graves, C., Gott, A., Gamez, P., & Rodriguez, S. (2019). Feeling our way to machine minds: People's emotions when perceiving mind in artificial intelligence. *Computers in Human Behavior*, *98*, 256–266.

Shechtman, N., & Horowitz, L. M. (2003). Media inequality in conversation: How people behave differently when interacting with computers and people. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 281–288.

Sherif, M. (1936). *The psychology of social norms.*

Siegel, M., Breazeal, C., & Norton, M. I. (2009). Persuasive robotics: The influence of robot gender on human behavior. *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2563–2568.

Singer, E., Van Hoewyk, J., & Maher, M. P. (2000). Experiments with incentives in telephone surveys. *Public Opinion Quarterly*, *64*(2), 171–188.

Singh, S., & Murry, T. (1978). Multidimensional classification of normal voice qualities. *The Journal of the Acoustical Society of America*, *64*(1), 81–87.

Slobin, D. I. (1971). Psycholinguistics (Glenview, Ill.). *Scott, Fores-224 IV. Infancy*.

Smith, K. T. (2020). Marketing via smart speakers: What should Alexa say? *Journal of Strategic Marketing*, *28*(4), 350–365.

Söderlund, M. (2022). Service robots with (perceived) theory of mind: An examination of humans' reactions. *Journal of Retailing and Consumer Services*, *67*, 102999.

Solomon, J., & Wash, R. (2014). Human-what interaction? Understanding user source orientation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 422–426.

Spence, J. T., & Buckner, C. E. (2000). Instrumental and Expressive Traits, Trait Stereotypes, and Sexist Attitudes: What Do They Signify? *Psychology of Women Quarterly*, *24*(1), 44–53.

Spence, J. T., & Helmreich, R. L. (1979). *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents*. University of Texas Press.

Spence, J. T., & Helmreich, R. L. (1980). Masculine instrumentality and feminine expressiveness: Their relationships with sex role attitudes and behaviors. *Psychology of Women Quarterly*, *5*(2), 147–163.

Spence, J. T., Helmreich, R., & Stapp, J. (1975). Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality and Social Psychology*, *32*(1), 29.

Srinivasan, V., & Takayama, L. (2016). Help me please: Robot politeness strategies for soliciting help from humans. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4945–4955.

Stein, J.-P., & Ohler, P. (2017). Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, *160*, 43–50.

Steinhaeusser, S. C., & Lugrin, B. (2022). Effects of Colored LEDs in Robotic Storytelling on Storytelling Experience and Robot Perception. *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, 1053–1058.

Sterling, G. (2019). Alexa devices maintain 70% market share in US according to survey. *MarTech*.

Stotland, E., & Zander, A. (1958). Effects of public and private failure on self-evaluation. *The Journal of Abnormal and Social Psychology*, *56*(2), 223.

Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology*, *18*(1), 86–100.

Strathmann, C., Szczuka, J., & Krämer, N. (2020). She talks to me as if she were alive: Assessing the social reactions and perceptions of children toward voice assistants and their appraisal of the appropriateness of these reactions. *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, 1–8.

Straub, I., Nishio, S., & Ishiguro, H. (2010). Incorporated identity in interaction with a teleoperated android robot: A case study. *19th International Symposium in Robot and Human Interactive Communication*, 119–124.

Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys: A review and synthesis*.

Sumner, W. G. (2019). *Folkways: A study of the sociological importance of usages, manners, customs, mores, and morals*. Good Press.

Sundar, S. S., & Nass, C. (2000). Source orientation in human-computer interaction: Programmer, networker, or independent social actor. *Communication Research*, *27*(6), 683–703.

Sutton, S. J. (2020). Gender ambiguous, not genderless: Designing gender in voice user interfaces (vuis) with sensitivity. *Proceedings of the 2nd Conference on Conversational User Interfaces*, 1–8.

Svendsen, G. B., Johnsen, J.-A. K., Almås-Sørensen, L., & Vittersø, J. (2013). Personality and technology acceptance: The influence of personality factors on the core constructs of the Technology Acceptance Model. *Behaviour & Information Technology*, *32*(4), 323–334.

Tajfel, H. (1974). Social identity and intergroup behaviour. *Social Science Information*, *13*(2), 65–93.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178.

Tajfel, H., Turner, J. C., Austin, W. G., & Worchel, S. (1979). An integrative theory of intergroup conflict. *Organizational Identity: A Reader*, *56*(65), 9780203505984–16.

Takeuchi, Y., & Katagiri, Y. (1999). Identity perception of computers as social actors. *Proceedings of ICCS*, *99*, 247–252.

Takeuchi, Y., Katagiri, Y., Nass, C., & Fogg, B. (1998). *Social response and cultural dependency in human-computer interaction*.

Takeuchi, Y., Katagiri, Y., Nass, C., & Fogg, B. (2000). A cultural perspective in social interface. *CHI*.

Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: The double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior*, *38*, 75–84.

Terzopoulos, G., & Satratzemi, M. (2019). Voice assistants and artificial intelligence in education. *Proceedings of the 9th Balkan Conference on Informatics*, 1–6.

Tolmeijer, S., Zierau, N., Janson, A., Wahdatehagh, J. S., Leimeister, J. M. M., & Bernstein, A. (2021). Female by Default?–Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–7.

Triandis, H. C. (1994). *Culture and social behavior*.

Triandis, H. C. (2018). *Individualism and collectivism*. Routledge.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*(1), 35–57.

Tung, F.-W., & Deng, Y.-S. (2007). Increasing social presence of social actors in e-learning environments: Effects of dynamic and static emoticons on children. *Displays*, *28*(4–5), 174–180.

Turing, A. M. (1950). Mind. *Mind*, *59*(236), 433–460.

Turkle, S. (2011). *The second self: Computers and the human spirit*. Simon and Schuster New York.

Twenge, J. M. (1997). Changes in masculine and feminine traits over time: A meta-analysis. *Sex Roles*, *36*(5), 305–325.

Tzeng, J.-Y. (2004). Toward a more civilized design: Studying the effects of computers that apologize. *International Journal of Human-Computer Studies*, *61*(3), 319–345.

Van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, The*, *25*(4), 40–48.

Van der Vegt, G. S., & Van de Vliert, E. (2005). Effects of perceived skill dissimilarity and task interdependence on helping in work teams. *Journal of Management*, *31*(1), 73–89.

Velez, J. A. (2015). Extending the theory of Bounded Generalized Reciprocity: An explanation of the social benefits of cooperative video game play. *Computers in Human Behavior*, *48*, 481–491.

von der Pütten, A. M., Krämer, N. C., & Gratch, J. (2010). How our personality shapes our interactions with virtual characters-implications for research and development. *International Conference on Intelligent Virtual Agents*, 208–221.

von der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S.-H. (2010). "It doesn't matter what you are!" explaining social effects of agents and avatars. *Computers in Human Behavior*.

Vorderer, P., Wirth, W., Gouveia, F. R., Biocca, F., Saari, T., Jäncke, L., Böcking, S., Schramm, H., Gysbers, A., & Hartmann, T. (2004). Mec spatial presence questionnaire. *Retrieved Sept*, *18*, 2015.

Wageman, R. (1995). Interdependence and group effectiveness. *Administrative Science Quarterly*, 145–180.

Wang, Y., Fan, X., Chen, I.-F., Liu, Y., Chen, T., & Hoffmeister, B. (2019). End-to-end anchored speech recognition. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7090–7094.

Wardini, J. (2022). Voice Search Statistics: Smart Speakers, Voice Assistants, and Users in 2022. *Serpwatch*. https://serpwatch.io/blog/voice-search-statistics/#:~:text=How%20many%20people%20use%20Siri,assistants%20worldwide%2C%20alongside%20Google%20Assistant.

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, *5*(3), 219–232.

Weaver, C. E., Lazaros, E. J., Zhao, J. J., Davison, C. B., & Truell, A. D. (2020). The Internet of Things: An overview of selected smart home technology. *Issues in Information Systems*, *21*(2), 43.

Weigand, E. (1999). Misunderstanding: The standard case. *Journal of Pragmatics*, *31*(6), 763–785.

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

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation.*

West, M., Kraut, R., & Ei Chew, H. (2019). *I'd blush if I could: Closing gender divides in digital skills through education*.

What is the Alexa Skills Kit? (n.d.). Retrieved March 1, 2022, from https://developer.amazon.com/en-US/docs/alexa/ask-overviews/what-is-the-alexa-skills-kit.html

Williams, J. E., & Best, D. L. (1990). *Measuring sex stereotypes: A multination study, Rev*. Sage Publications, Inc.

Williams, P. A., Jenkins, J., Valacich, J., & Byrd, M. D. (2017). Measuring actual behaviors in HCI research–a call to action and an example. *AIS Transactions on Human-Computer Interaction*, *9*(4), 339–352.

Wilson, W., & Chambers, W. (1989). The effectiveness of praise of self versus praise from others. *The Journal of Social Psychology*, *129*(4), 555–556.

Winograd, T., & Flores, F. (1987). On understanding computers and cognition: A new foundation for design. *Artificial Intelligence*, *31*(2), 250–261.

Wolters, M. K., Kelly, F., & Kilgour, J. (2016). Designing a spoken dialogue interface to an intelligent cognitive assistant for people with dementia. *Health Informatics Journal*, *22*(4), 854–866.

Yang, H., & Lee, H. (2019). Understanding user behavior of virtual personal assistant devices. *Information Systems and E-Business Management*, *17*(1), 65–87.

Zanbaka, C., Goolkasian, P., & Hodges, L. (2006). Can a virtual cat persuade you? The role of gender and realism in speaker persuasiveness. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1153–1162.