

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/341189578>

Not Human After All: Exploring the Impact of Response Failure on User Perception of Anthropomorphic Conversational Service Agents

Conference Paper · June 2020

CITATIONS

10

READS

633

4 authors:



Stephan Diederich

Georg-August-Universität Göttingen

25 PUBLICATIONS 371 CITATIONS

[SEE PROFILE](#)



Tim-Benjamin Lembcke

Georg-August-Universität Göttingen

37 PUBLICATIONS 172 CITATIONS

[SEE PROFILE](#)



Alfred Benedikt Brendel

Technische Universität Dresden

90 PUBLICATIONS 851 CITATIONS

[SEE PROFILE](#)



Lutz Kolbe

Georg-August-Universität Göttingen

312 PUBLICATIONS 4,967 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



IS Support for Vehicle Supply and Demand Management in Shared Vehicle Services [View project](#)



Design Science and Design Thinking [View project](#)



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

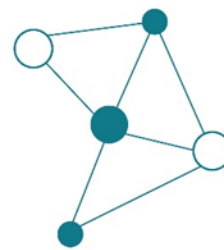
This is the author's version of a work that has been published in the following outlet:

Diederich, S.; Lembcke, T.-B.; Brendel, A.B.; Kolbe, L.M. (2020): Not Human After All: Exploring the Impact of Response Failure on User Perception of Anthropomorphic Conversational Service Agents, Proceedings of European Conference on Information Systems (ECIS), Marrakech, Morocco.



Chair of Information Management

Prof. Dr. Lutz M. Kolbe
Platz der Göttinger Sieben 5
37073 Göttingen – Germany
www.uni-goettingen.de/im



Smart
Mobility
Research
Group

Smart Mobility Research Group (SMRG)

Dr. Alfred Benedikt Brendel
Humboldtallee 3
37073 Göttingen – Germany
www.uni-goettingen.de/smrg

Please note: The copyright is owned by the author and / or the publisher.
Commercial use is not allowed.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

NOT HUMAN AFTER ALL: EXPLORING THE IMPACT OF RESPONSE FAILURE ON USER PERCEPTION OF ANTHROPOMORPHIC CONVERSATIONAL SERVICE AGENTS

Research paper

Diederich, Stephan, Chair of Information Management, University of Goettingen,
Goettingen, Germany, stephan.diederich@stud.uni-goettingen.de

Lembcke, Tim-Benjamin, Chair of Information Management, University of Goettingen,
Goettingen, Germany, tim-benjamin.lembcke@uni-goettingen.de

Brendel, Alfred Benedikt, Chair of Information Management, University of Goettingen,
Goettingen, Germany, abrendel@uni-goettingen.de

Kolbe, Lutz M., Chair of Information Management, University of Goettingen,
Goettingen, Germany, lkolbe@uni-goettingen.de

Abstract

Conversational agents (CAs) have attracted the interest of organizations due to their potential for automated service provision combined with the feeling of a human-like interaction. Emerging studies on CAs indicate a positive impact of humanness on customer perception and explore approaches for their anthropomorphic design, comprising both appearance and behavior of the agent. While these studies provide valuable knowledge on how to design human-like CAs, we still lack an understanding of the limited conversational capabilities of this technology and their impact on user perception. Oftentimes, these limitations lead to frustrated users and discontinued CAs in practice. We address this gap by investigating the impact of response failure, understood as the inability of a CA to provide a meaningful reply, in a service context drawing on Social Response Theory and the Theory of Uncanny Valley. By means of an experiment with 169 participants, we find that (1) response failure is detrimental to the perception of humanness and increases feelings of uncanniness, (2) humanness (uncanniness) positively (negatively) influences familiarity and service satisfaction, and (3) the negative impact of response failure on user perception is significant yet it does not lead to a sharp drop as posited by the Theory of Uncanny Valley.

Keywords: *Conversational Agent, Anthropomorphic Design, Social Response Theory, Theory of Uncanny Valley.*

1 Introduction

Conversational agents (CAs), defined as technological artifacts with which users interact through natural language (McTear et al., 2016), continue to gain interest in research (Maedche et al., 2019) and practice (Oracle, 2016) alike. Praised for their potential to provide a human-like interaction experience, CAs are increasingly used in private as well as professional life. From a theoretical perspective, such agents are a particular interesting phenomenon as humans show social responses to these agents (Pfeuffer et al., 2019). As posited by Social Response Theory (Reeves and Nass, 1996; Nass and Moon, 2000), the manifold social cues of CAs, such as the interaction via natural language, having a name and (human-like) avatar, or the expression of emotions through verbal and non-verbal communication, trigger social responses and lead users to anthropomorphize CAs (Seeger et al., 2018). Emerging design-oriented studies on human-like CAs provide valuable knowledge on the impact of social cues on humanness, understood as the degree to which users attribute actual human

properties (e.g. thoughtfulness) to the agent. Moreover, different research suggests further effects of anthropomorphism, such as on service satisfaction (Gnewuch et al., 2018; Diederich et al., 2019c), likability (Bickmore and Picard, 2005), or familiarity (de Visser et al., 2016). Thus, the perception of anthropomorphism can contribute to relevant context-specific variables. To make these artifacts appear as human-like as possible, the growing knowledge base for anthropomorphic CA design offers various social cues that can be incorporated in the design (Feine et al., 2019).

While these studies provide valuable knowledge for crafting CAs with a human-like appearance and behavior, the current debate for anthropomorphic design neglects the practical problem of limited conversational capabilities. As these studies were primarily carried out by means of experimental research where users were given a specific set of tasks or interacted with a human in a Wizard-of-Oz setting (Diederich et al., 2019a), the CAs were able to provide relevant responses to the users' requests. In practice, however, designing agents that continuously offer meaningful responses in an evolving dialogue represents a major challenge (Følstad and Brandtzæg, 2017). In fact, many CAs were discontinued particularly due to their inability to adequately respond to varying user input (Ben Mimoun et al., 2012). As anticipating user requests for natural language software is a challenging endeavor due to the unpredictability of such interactions, situations where a CA needs to provide some kind of fallback response are likely to occur and could remind users that they are still interacting with a machine that has limited capabilities (Ashktorab et al., 2019). Such failure to provide a meaningful reply might be detrimental to perception of humanness and further positive effects, thus diminishing the impact of social cues incorporated in the agent's design. In short, the impact of response failure on the perception of anthropomorphic CAs represents a substantial practical design problem for which we yet lack a solid understanding.

In our study, we seek to address this research gap by investigating the impact of response failure, understood as the inability of a CA to meaningfully respond to a valid user's request, on user perception with the following research question: *How does failure to provide a meaningful response influence user perception of anthropomorphic CAs in a service encounter?* Drawing on extant studies on anthropomorphic design of CAs, Social Response Theory and Theory of Uncanny Valley as two key theories on human perception of and interaction with human-like artifacts, we develop a research model comprising eight hypotheses and test it in a 2x2 experiment with $n = 169$ participants. Our research makes three main contributions: First, it advances our understanding of the influence of response failure due to limited conversational capabilities on the perception of humanness and uncanniness (i.e. the feeling of strangeness due to inhuman qualities of an anthropomorphic artifact) of an agent. Second, the study demonstrates the positive (negative) impact of humanness (uncanniness) on familiarity (i.e. the degree to which users feel acquainted with the agent) and service satisfaction (i.e. the satisfaction of users with the agent's service encounter and interaction). Third, our experiment allows to better understand the magnitude of the effect of modest response failure on user perception of the CA, particularly depending on the agent's machine- or human-like design.

We continue by outlining related work providing the theoretical background for our work. Afterwards, we derive eight hypotheses, introduce our research model, and describe the design of the experiment. We then present the results, discuss implications for the design of human-like CAs, and highlight limitations as well as directions for future research before closing with concluding remarks.

2 Related Work and Theoretical Background

The idea to interact with technology via natural language instead of graphical user interfaces emerged already in the 1960s (Weizenbaum, 1966). However, it regained interest just a few years ago when advances in natural language processing and machine learning substantially increased the (conversational) capabilities of such technologies (McTear, 2017). Today, conversational agents, defined as software with which users interact through natural language (McTear et al., 2016), are increasingly permeating our private and professional lives (Maedche et al., 2019) in various areas, such as customer service (Hu et al., 2018), marketing and sales (Vaccaro et al., 2018), human resources (Diederich et al., 2020), financial advisory (Dolata et al., 2019), or education (Crockett et

al., 2017). In addition to these application areas, different forms of CAs can be distinguished by their primary mode of communication and embodiment. In general, technology interaction through natural language can take place in spoken form, such as with Amazon's Alexa or Apple's Siri, or via written text like with chatbots on company websites or social media (Gnewuch et al., 2017). Furthermore, CAs can be physically embodied like service robots (Stock and Merkle, 2018; Stock et al., 2019), have a virtual static avatar (Wunderlich and Paluch, 2017), a virtual interactive avatar (Beer et al., 2015), or be disembodied, i.e. without any form of avatar at all (Araujo, 2018). In this study, we focus on a CA with which users communicate via written text (chatbot) and a static virtual avatar (image) in a customer service context.

2.1 Conversational Service Agents and their Responsiveness

Customer service is currently one of the most popular application areas for CAs in enterprises where such agents can fulfill requests like handling complaints or providing product information (Gnewuch et al., 2017; Diederich et al., 2019b). While current CAs primarily cover rather simple, frequent, and repetitive service requests, they are expected to support or even fully assume increasingly complex tasks currently performed by human service personnel (Verhagen et al., 2014; Marinova et al., 2017). As technological components of service systems, CAs are positioned between current service technology that is always available but lacks the feeling of a human interaction, such as online portals for self-service, and human service provision, offering a personal contact but with limited availability. In practice, different examples for CAs in a service context can be found across industries (Oracle, 2016). For example, the American railroad company Amtrak introduced the virtual agent "Julie" which now answers more than five million customer requests per year (NextIT, 2018). Similarly, the clothing brand H&M offers an artificial sales agent that provides individual product recommendations which can be directly purchased from the company's online store (Morana et al., 2017). This popularity is further underlined by the example of Facebook where more than 100.000 agents were deployed in the first year after opening the Messenger platform (Johnson, 2018).

Despite their popularity and success stories, many CAs fell behind high expectations in the past (Luger and Sellen, 2016) and were often discontinued because of flaws related to their design (Ben Mimoun et al., 2012). In an assessment of 80 conversational agents on French commercial websites, Ben Mimoun et al. (2012) identified inadequate appearance and a lack of interactivity as well as intelligence as reasons for CA failure. The authors argue that a mismatch between the human-like appearance of CAs and their actual service possibilities as well as competence leads to negative customer reactions due to unfulfilled, high expectations. Likewise, Luger and Sellen (2016, p. 5286) find "user expectations dramatically out of step with the operation of the systems, particular in terms of known machine intelligence, system capabilities and goals". Hence, both studies suggest a gap between user expectations and technical capabilities as a main reason for negative perception of CAs and suggest different design approaches to manage user expectations more adequately.

Against this background, Følstad and Brandtzæg (2017) emphasize that a natural language interface resembles a blank canvas where the capabilities of the system are mostly hidden from the user and that designers need to anticipate a much larger variety of input compared to graphical user interfaces. Consequently, the authors argue that fallbacks in a conversation are likely to occur (Følstad and Brandtzæg, 2017). Similarly, Go and Sundar (2019) highlight that equipping a CA with the ability to provide meaningful responses, contingent on what has already been communicated in a conversation, represents a substantial design issue.

Overall, we observe a growing popularity of CAs in a service context yet the implementation of sufficient conversational capabilities to respond to highly varying user input is a major design challenge and leaves room for response failures in the interaction.

2.2 Social Response Theory and the Uncanny Valley in the Context of CAs

A key theory underlying the design of and interaction with CAs is Social Response Theory (Reeves and Nass, 1996; Nass and Moon, 2000). Social Response Theory posits that humans apply social rules as well as expectations to technology that exhibits traits or behavior usually associated with humans (Nass and Moon, 2000). In a set of experiments, Nass and Moon (2000) discovered that humans overuse social categories, such as gender, and social behaviors, for example reciprocity, in an interaction with a human-like artifact. According to the researchers, the more human characteristics are present in a technological artifact, the stronger it leads to social responses (Nass and Moon, 2000). As CAs typically exhibit a variety of social cues (Feine et al., 2019), ranging from basic cues like the interaction via natural language and turn-taking in a conversation to more complex ones, such as understanding and expressing emotions, they trigger substantial social responses by humans. As suggested by Seeger et al. (2018), anthropomorphic design of CAs comprises social cues in three different dimensions: A human identity (e.g. age, gender, ethnicity), verbal cues (e.g. syntax and word variability or the use of self-references in a conversation), and non-verbal cues (e.g. response delays to indicate thinking or the use of emoticons to express emotions). In short, designers have various social cues at their disposal to make CAs seem as human-like as possible (Feine et al., 2019).

Recent experiments on the perception of anthropomorphic CAs have discovered mostly positive effects of a human-like design on aspects, such as social presence (Pereira, Prada and Paiva, 2014), trustworthiness (Araujo, 2018), persuasiveness (Diederich et al., 2019d), enjoyment (Liao et al., 2018), or service satisfaction (Gnewuch et al., 2018). However, some studies also indicate unintended negative effects. For example, Wunderlich and Paluch (2017) describe a risk of perceived uncertainty as to whether the user interacts with a machine or an actual human. Furthermore, Sohn (2019) discovers increased privacy concerns due to the mere presence of an anthropomorphic agent on an e-commerce website. In addition, Seeger et al. (2018) assume a negative effect on the perception of anthropomorphism in a CA when a CA includes social cues from all three aforementioned dimensions. Thus, the authors suggest to find an appealing combination of social cues instead of a “more is more” approach.

In this context, Mori (1970; 2012) hypothesized on the relationship between human-like objects and affinity (or familiarity, as it is often translated from the original Japanese manuscript) in the Theory of Uncanny Valley. The theory, originally from the field of robotics, posits that there is no linear relationship between human-likeness of an object and positive emotional responses by humans to it, but that a sharp drop in affinity or familiarity exists before the object becomes fully human-like. MacDorman et al. (2009, p.2) describe this as a shift of human attention from the human-like qualities to the aspects that seem to be inhuman by stating that “as something looks more human it looks also more agreeable, until it comes to look so human that we start to find its nonhuman imperfections unsettling”. This negative reaction termed as the Uncanny Valley comprises strong feelings of uncanniness due to the nonhuman imperfections of an object (MacDorman et al., 2009). Figure 1 depicts the Uncanny Valley as conceptualized by Mori et al. (2012).

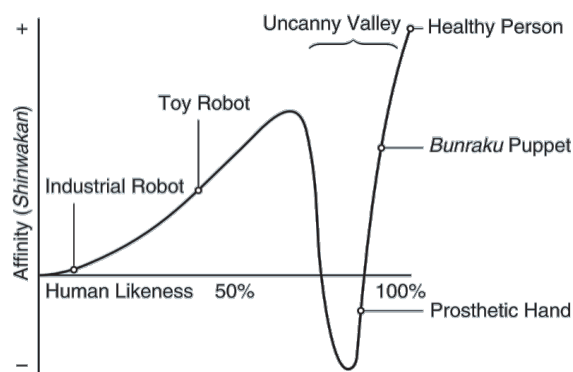


Figure 1. The Uncanny Valley (Mori et al., 2012)

In the remainder of this study, we consider the ideas of Social Response Theory and the Theory of Uncanny Valley to explore the effects of practical response failures on the perception of human-like CAs in a service context.

3 Hypotheses and Research Model

Our study aims for a better understanding of the impact of response failure by a human-like conversational agent in a natural language dialogue. For this purpose, we propose a research model comprising eight hypotheses.

Drawing on Social Response Theory (Reeves and Nass, 1996; Nass and Moon, 2000), a human-like appearance and behavior by a technological artifact triggers social responses in humans. As Nass and Moon (2000) argue, the more technological artifacts, such as a computer, exhibit human-like characteristics, the more likely they trigger social reactions. In the context of CAs, emerging studies on anthropomorphic design indicate that social cues lead to perception of humanness in the interaction (Wunderlich and Paluch, 2017; Feine et al., 2019). In line with these studies and Social Response Theory, we first hypothesize:

H1: Social cues have a positive impact on humanness of the agent.

Second, we propose that an appealing combination of different social cues, comprising a human identity of the agent as well as verbal and non-verbal cues (Seeger et al., 2018), reduces feelings of uncanniness when interacting with the agent. Different studies on anthropomorphic design indicate positive effects of social cues, such as on likability (Bickmore and Picard, 2005; Cowell and Stanney, 2005), trust (Nunamaker et al., 2011; de Visser et al., 2016) or enjoyment (Qiu and Benbasat, 2010). Thus, we suggest that social cues reduce feelings of uncanniness when interacting with an anthropomorphic agent.

H2: Social cues have a negative impact on uncanniness of the agent.

Against the background of the Theory of Uncanny Valley, Mori et al. (2012, p. 98) hypothesize that a “person’s response to a humanlike robot would abruptly shift from empathy to revulsion as it approached, but failed to attain, a lifelike appearance”. While there can be many reasons why a CA may not be able to sustain a human-like appearance in a dialogue (e.g. due to the imperfect representation an interactive, human-like avatar (Seymour et al., 2018), we argue that one of the most likely reasons is the inability to provide a meaningful response due to the complexity and unpredictability of user input in a natural language interaction (Følstad and Brandtzæg, 2017). As a result, this failure to respond would abruptly shift a user’s attention to the fact that she or he is in fact not interacting with an actual human, thereby decreasing the perception of humanness in the conversation:

H3: Response failure has a negative impact on humanness of the agent.

Similarly, we expect an agent’s inability to provide a meaningful response to induce feelings of uncanniness (Tinwell and Sloan, 2014) as it constitutes a strange situation that does not conform with the users expectations towards a human-like conversation (Luger and Sellen, 2016). Thus, we formulate our fourth hypothesis as follows:

H4: Response failure has a positive impact on uncanniness of the agent.

Furthermore, anthropomorphized artifacts have been known for their ability to induce feelings of familiarity (Epley et al., 2007) because social cues make it easier for the user to connect with the technology, potentially even on a personal level (Burgoon et al., 2000), and feel at ease with the artifacts form and function (Duffy, 2003). This relation is further conceptualized in the Theory of Uncanny Valley (Mori, 1970; Mori et al., 2012) where familiarity increases steadily until the valley is reached. Hence, we suggest that perception of humanness in an artifact positively impacts familiarity:

H5. Humanness has a positive effect on familiarity of the agent.

Next, we consider the relation between uncanniness and familiarity. Based on a similar reasoning as for the fifth hypothesis, we argue that feelings of uncanniness in the interaction are detrimental to familiarity in line with the Theory of Uncanny Valley (Mori, 1970; Mori et al., 2012). As uncanniness, manifested in feelings of for example strangeness or eeriness during the interaction with an anthropomorphic agent, contributes to a negative user perception (Tinwell and Sloan, 2014), it diminishes the perception of the agent as familiar:

H6: Uncanniness has a negative impact on familiarity of the agent.

Finally, different studies on CAs in a service context argue for a positive relation between the perception of humanness as well as related social responses as postulated in Social Response Theory (Nass and Moon, 2000), on service satisfaction, which is of particular importance in online service encounters. In this context, Gnewuch et al. (2018) for example found that dynamic response delays to indicate thinking and typing of an agent response (Gnewuch et al., 2018) lead to an increased feeling of humanness and service satisfaction. Similarly, Diederich et al. (2019c) find a positive impact of a sentiment-adaptive CA design to emulate human empathy on satisfaction of customers with a service encounter. Thus, we hypothesize that humanness in a service encounter positively impacts satisfaction and, similarly, that negative feelings of uncanniness are detrimental to service satisfaction:

H7: Humanness has a positive impact on service satisfaction.

H8: Uncanniness has a negative impact on service satisfaction.

Figure 2 summarizes our eight hypotheses and visualizes the research model for our study.

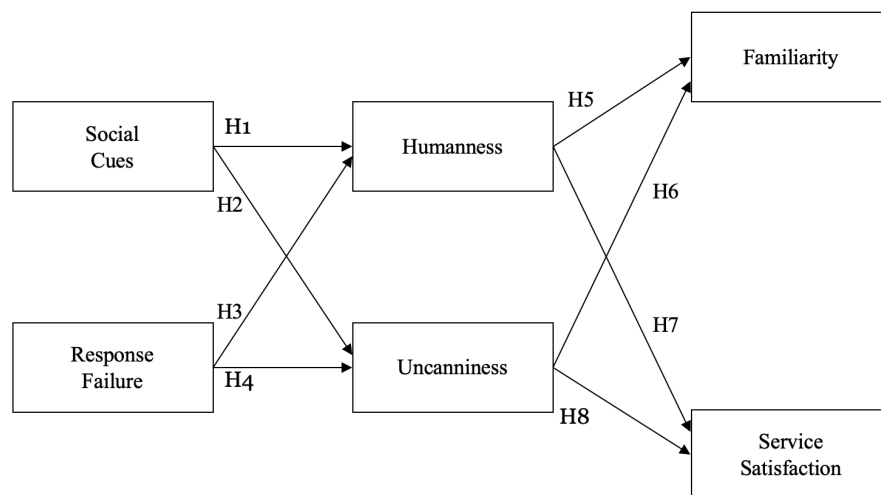


Figure 2. Research model

4 Research Design

We tested our hypotheses regarding anthropomorphic design and response failure in an online experiment with a text-based CA. To provide a familiar and understandable context and task, we selected a customer service setting with an online retailer. In the following, we describe the data collection procedure and sample, the four experimental conditions, the manipulation check, as well as the measures used in the post-experimental survey.

4.1 Data Collection Procedure and Sample

The participants of our experiment were asked to interact with a virtual service agent of a fictitious online retailer to track and cancel an existing order as well as to ask for a confirmation. Every participant received a link to a briefing website in which we described the context of the experiment (online retailer), the structure (interaction with a virtual customer service agent followed by a questionnaire) as well as the participant's tasks. The tasks comprised contacting the service agent and

finding out the current order status for a given identification number, authenticating with the agent, asking for order cancellation as well requesting a confirmation via e-mail. Similar to recent studies on CAs (e.g. Gnewuch et al. (2018), Diederich et al. (2019c)), we selected a rather specific set of tasks to enable a structured, comparable dialogue across the conditions and contribute to the responsiveness of the agent in the interaction. After successful completion of the last task, the CA provided a link to the questionnaire. Overall, participation in the experiment took around nine minutes per person. Our sample has a size of $n = 169$ with the participants' ages ranging from 19 to 59 years (mean: 27.8 years) and a share of 40.6% female persons. Four participants provided straight-line answers and were thus removed from the sample, decreasing the final sample size to $n = 165$. A monetary compensation was not provided for participation in the experiment. The participants were recruited from personal networks and comprised mainly students from a German university.

4.2 Experimental Conditions

For the experimental conditions, we designed four instances of a CA using the natural language platform Dialogflow by Google (2019). Dialogflow provides the technical capabilities to detect a user's intent from a natural language statement and formulate a response. All instances received the same set of training phrases. We varied the design and responsiveness as visualized in Figure 3.

		Social Cues	
		Few	Many
Response Failure	With	Condition 1 Agent with few social cues and with response failure	Condition 2 Agent with many social cues and with response failure
	Without	Condition 3 Agent with few social cues and without response failure	Condition 4 Agent with many social cues and without response failure

Figure 3. Experimental conditions

The two instances with a human-like design received a set of social cues to make the agents appear human-like (Figure 4). Against the background of the three anthropomorphic design dimensions proposed by Seeger et al. (2018), we provided the agent with a comic-like avatar of a female customer service employee (Gong, 2008), gave it a human name (Cowell and Stanney, 2005) and a gender (Nunamaker et al., 2011) to establish a human identity.

The screenshot shows a chat window with the following text:

- Agent: All right, the zip code and city are correct for order #1290. 😊 How can I help you with this order?
- User: I would like to cancel the order, if possible.
- Agent: All right, I cancelled the order in our system. Do you need further support?
- User: Please confirm the cancellation of the order via e-mail.
- Agent: Sure, I will send you an e-mail right away to confirm the cancellation of order #1290.
- Agent: Thank you for contacting our customer service today. 🙏

Labels on the left side of the image:

- Avatar, name, gender: Points to the agent's profile picture and name 'Laura (Customer Service)'.
- Emotion expression with emoticon: Points to the smiley face emoji in the first agent message.
- Self-reference: Points to the agent's use of 'I' and 'myself' in the messages.
- Exemplary social cues: Points to the thank-you message at the end of the conversation.

Labels on the right side of the image:

- Agent without response failure: Points to the entire chat window.
- Immediate meaningful reply for cancellation request: Points to the agent's response 'All right, I cancelled the order in our system...'.

Figure 4. CA with human-like design and without response failure (condition 4)

We further integrated self-references (Sah and Peng, 2015), self-disclosure (Schuetzler et al., 2018), a personal introduction and greeting (Cafaro et al., 2016), and variability in syntax as well as word choice for the agent’s responses (Seeger et al., 2018) in terms of verbal cues. With regard to non-verbal cues, we added dynamic response delays to indicate thinking and typing of replies (Gnewuch et al., 2018) in combination with blinking dots (de Visser et al., 2016) as well as the use of emoticons to express emotions (Wang et al., 2008).

With regard to the second dimension, response failure, we designed the agent in conditions 1 and 2 to indicate a lack of understanding at one point in the interaction. When the participants in those conditions requested to cancel the given order, the agent politely responded that it did not understand the user’s input and asked for reformulation of the request for two times. After the participant entered the request for order cancellation a third time, the agent provided a meaningful response and confirmed the cancellation. Table 1 shows the agent’s responses in the two conditions with low responsiveness.

Iteration	Condition 1 (few social cues)	Condition 2 (many social cues)
1	“Unfortunately, I do not understand your request.”	“Unfortunately, I do not understand your request. Can you please reformulate it?”
2	“Unfortunately, I do not understand your request.”	“I am so sorry, but I do not understand what you are saying. Can you please formulate it differently?”
3	“Your order is now cancelled.”	“All right, I cancelled the order in our system. Do you need further support?”

Table 1. Agent statements with response failure (translated to English)

4.3 Manipulation Check

To check whether the manipulation of the responsiveness dimension was successful in the sense that participants only received meaningless responses as intended without a multitude of further fallback replies, we analyzed the conversation data provided by Google Dialogflow. Reviewing the interactions with the agents in the four conditions showed that in most cases the agents demonstrated the communication behavior as intended with an overall average of additional conversational fallbacks of around 1.5 responses per interaction. All conditions showed similar average fallbacks between 1.7 and 1.4 messages, comprising situations in a dialogue where the agent did not understand a user’s intent. Thus, only a minimal number of fallbacks, similar across all groups, existed in the interactions.

4.4 Measures

Every participant was asked to complete a survey to measure perceptions of humanness, uncanniness, familiarity, and service satisfaction. We used established measurement instruments for the four constructs. Humanness and familiarity were both measured on a 9-point semantic differential scale with items from Holtgraves and Han (2007) and MacDorman (2006) respectively. To measure feelings of uncanniness, we adapted a 7-point Likert scale based on the studies by MacDorman et al. (2009) and Tinwell and Sloan (2014). Similarly, service satisfaction was measured on a 7-point Likert scale using items from Verhagen et al. (2014). Furthermore, we collected demographic information (age, gender) and information on the frequency of digital assistant use (e.g. Siri, Alexa, and chatbots). Finally, we asked for free form feedback on the perception of the agent. We added attention checks by inverting two items in the survey.

Constructs and items	Loadings	Scale and source
Humanness ($\alpha = .904$, CR = .927, AVE = .680) Extremely inhuman-like – extremely human-like Extremely unskilled – extremely skilled Extremely unthoughtful – extremely thoughtful Extremely impolite – extremely polite Extremely unresponsive – extremely responsive Extremely unengaging – extremely engaging	.887 .882 .853 .671 .821 .816	9-point semantic differential scale (Holtgraves and Han 2007)
Uncanniness ($\alpha = .911$, CR = .932, AVE = .698) I perceived the agent as eerie. I perceived the agent as inhuman-like. I perceived the agent as strange. I perceived the agent as unappealing. I perceived the agent as inclement. I perceived the agent as unpleasant.	.665 .784 .880 .909 .853 .895	7-point Likert scale (MacDorman et al., 2009; Tinwell and Sloan, 2014)
Familiarity Extremely strange – extremely familiar	n/a	9-point semantic differential scale (MacDorman, 2006)
Service Satisfaction ($\alpha = .888$, CR = .931, AVE = .819) How satisfied are you with the agent ‘s advice? ...the way the agent treated you? ...the overall interaction with the agent?	.914 .854 .944	7-point Likert scale (Verhagen et al. 2014)

Table 2. Constructs, items, and factor loadings

Table 2 shows the constructs, items, and factor loadings as well as Cronbach’s α , composite reliability (CR) and average variance extracted (AVE). Following the suggestions by Gefen and Straub (2005), items with loadings larger than .60 were used in the analysis. Humanness, uncanniness, and service satisfaction showed sufficient values for CR (larger than .80), Cronbach’s α (larger than .80) as well as AVE (larger than .50) considering the levels proposed by Urbach and Ahlemann (2010).

5 Results

We tested our hypotheses on the impact of response failure in combination with human-like CA design in a service context using partial least squares (PLS). The correlations between the latent variables are shown in Table 3 with square roots for AVE displayed in the diagonal.

	Familiarity	Humanness	Service Satisfaction	Uncanniness
Familiarity	1,000			
Humanness	0,816	0,825		
Service Satisfaction	0,751	0,891	0,905	
Uncanniness	-0,769	-0,777	-0,746	0,835

Table 3. Correlation matrix

We calculated the significance of path coefficients with a bootstrapping resampling approach with 5,000 samples (Chin, 1998). The resulting path coefficients, R^2 values for the dependent variables as well as significance levels are shown in Figure 5. All analyses were carried out using SmartPLS 3.

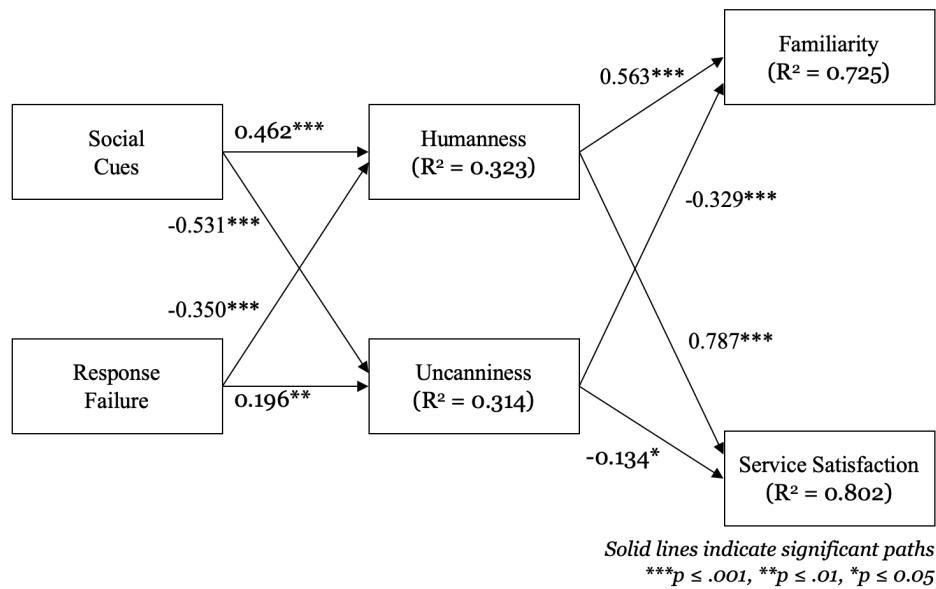


Figure 5. PLS structural model (n = 165)

The paths between social cues and humanness as well as uncanniness show significant relationships. In line with Social Response Theory, we find empirical evidence that social cues positively impact humanness of an agent (Social Cues → Humanness, $\beta = 0.462$, $p \leq .001$), thus confirming our first hypothesis. Furthermore, a human-like design with social cues has a negative impact on uncanniness, providing support for the second hypothesis (Social Cues → Uncanniness, $\beta = -0.531$, $p \leq .001$). With regard to response failure, we observe a negative impact on humanness (Response Failure → Humanness, $\beta = -0.350$, $p \leq .001$) and a positive impact on uncanniness (Response Failure → Uncanniness, $\beta = 0.196$, $p = .004$), as stated by hypotheses three and four. Our data further indicates a positive impact of humanness on familiarity with the agent (Humanness → Familiarity, $\beta = 0.563$, $p \leq .001$) and a negative impact of uncanniness (Uncanniness → Familiarity, $\beta = -0.329$, $p \leq .001$) as proposed in hypotheses five and six. Finally, we find support for the impact of humanness and uncanniness on service satisfaction: Humanness positively contributes to satisfaction (Humanness → Service Satisfaction, $\beta = 0.787$, $p \leq .001$) while uncanniness has a detrimental influence on service satisfaction (Uncanniness → Service Satisfaction, $\beta = -0.134$, $p = .017$).

To complement the results for our model, we analyzed the effect of the control variables that comprised demographic information of the participants (age, gender) as well as prior experience with digital assistants and chatbots. However, the variables did not exhibit significant paths to the latent variables. Furthermore, we assessed the size of direct effects with the f^2 values. Using the levels by Cohen (1988), we interpreted values of 0.02, 0.15, and 0.35 as small, medium, and large sizes respectively. All significant relationships had effect sizes that exceeded the small effect threshold.

Small effect ($f^2 \geq 0.02$)	Medium effect ($f^2 \geq 0.15$)	Large effect ($f^2 \geq 0.35$)
Response Failure → Uncanniness	Response Failure → Humanness	Humanness → Familiarity
Uncanniness → Service Satisfaction	Uncanniness → Familiarity	Humanness → Service Satisfaction
	Social Cues → Humanness	Social Cues → Uncanniness

Table 4. Effect sizes for significant paths according to the levels by Cohen (1988)

Direct effects with a small size comprise the paths between response failure and uncanniness ($f^2 = 0.054$) as well as uncanniness and service satisfaction ($f^2 = 0.036$). Medium-sized direct effects include social cues ($f^2 = 0.0313$) and response failure ($f^2 = 0.175$) to humanness as well as uncanniness to familiarity ($f^2 = 0.155$). Large effect sizes can be observed from social cues to uncanniness ($f^2 = 0.407$) and from humanness to familiarity ($f^2 = 0.453$) as well as to service satisfaction ($f^2 = 1.24$).

Finally, we indicatively compared the differences of the means for the latent variables to investigate whether a strong negative emotional response as postulated in the Uncanny Valley Theory can be observed. Our dataset indicates that participants who received a design with many social cues showed substantially stronger negative reactions ($\Delta = 1.59$) to response failure with regard to familiarity than participants that interacted with a machine-like CA with few social cues ($\Delta = 0.42$) while the differences between the means for humanness ($\Delta = 1.44$, $\Delta = 1.15$), uncanniness ($\Delta = -0.66$, $\Delta = -0.28$) service satisfaction ($\Delta = 1.11$, $\Delta = 1.97$) were comparatively smaller.

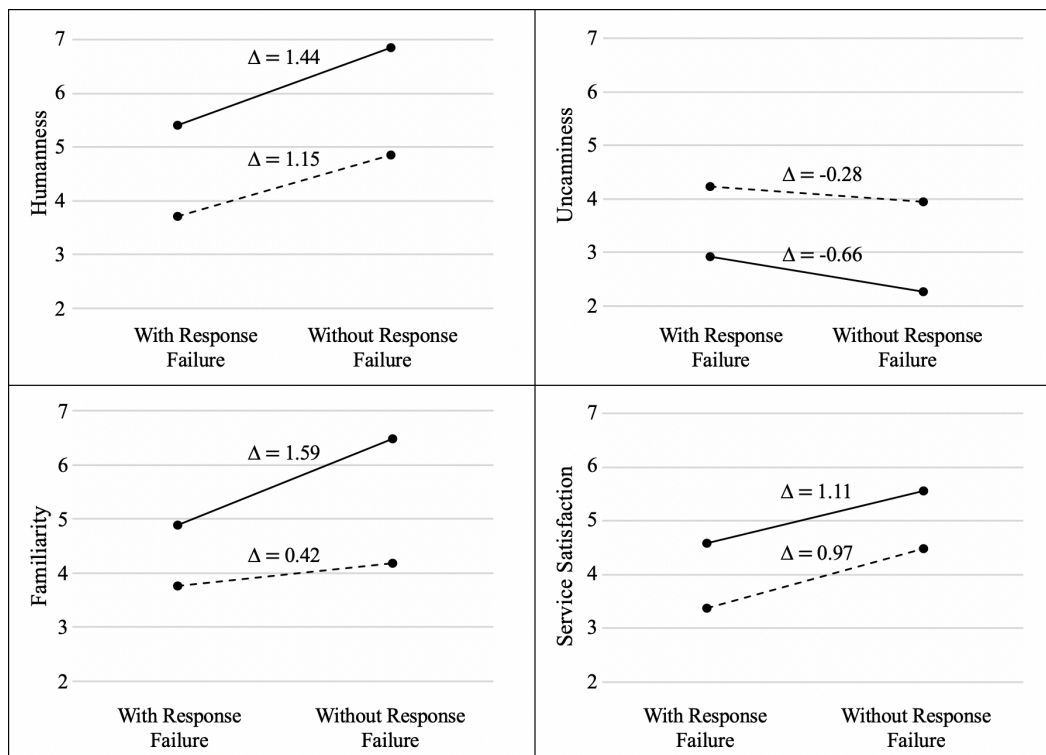


Figure 6. Mean values differentiated by social cues (dotted lines indicate conditions 1 and 3)

6 Discussion

Our study provides empirical evidence for the negative impact of response failure of CAs in a service context on the perception of humanness as well as a positive influence on unintended uncanniness of an agent. Furthermore, the main results show a positive impact of humanness in a service encounter on familiarity and satisfaction as well as a detrimental impact of uncanniness on these variables. In the following, we discuss implications of our results for research on anthropomorphic CA design as well as CA design in practice, indicate limitations, and suggest opportunities for future research.

6.1 Implications for Research on Anthropomorphic CAs

The results of our experiment emphasize a substantial negative impact of response failure of a CA on user perception in the context of a service encounter and are thus in line with recent research that highlights the importance of sufficient conversational capabilities of CAs (Følstad and Brandtzæg, 2017; Gnewuch et al., 2017; Schuetzler et al., 2018). Participants that interacted with a CA that

exhibited response failure indicated a lower perception of humanness, familiarity, and service satisfaction as well as increased uncanniness of the agent. Even if the response failure in conditions 1 and 2 can be considered rather modest and the CA in all cases was ultimately able to complete the user's service request, the negative impact on user perception was substantial. The qualitative, free-form feedback on the CA's design underlines this effect. For example, participants that interacted with a CA with response failure stated that they perceived it as "incomplete" or criticized that they had to "ask for order cancellation a thousand times". Furthermore, one participant commented that "What is the purpose of the nice design if the computer does not understand me?". Thus, the CA's failure to respond was immediately recognized by the participants in the respective experimental conditions.

Considering the results from our analysis of direct effect sizes (Table 4), even a rather modest failure to respond in the conversation led to a medium-sized detrimental effect on the perception of humanness of the agent. Interestingly, the large variety of social cues incorporated in the design of the human-like agent (conditions 2 and 4) exhibited a comparable effect size on humanness. Our experimental data indicates that even small response failure leads to a substantial negative effect on humanness with an effect size comparable to impact of the rich social cues on humanness of the agent. Our experiment therefore underlines the importance of managing and matching (high) user expectations when designing (anthropomorphic) CAs as suggested by Luger and Sellen (2016).

With regard to a potential Uncanny Valley-effect (Figure 1), our data does not exhibit a sharp drop in familiarity or increase in uncanniness of the magnitude posited by the original theory. While a substantial negative impact on user perception of response failure can be observed in our data, the mean values for familiarity (uncanniness) are still higher (lower) in the conditions with a human-like design with many social cues than in those with a machine like design with few social cues (Figure 6). Interestingly, however, the difference in familiarity depending on response failure seems to be larger for the human-like CAs ($\Delta = 1.59$) than the difference for the CAs with a machine-like design ($\Delta = 0.42$). This could be in line with a small decrease in familiarity at the beginning of the Valley's sharp drop. Drawing on the, admittedly conceptual, idea of Mori's (1970) theory, our data could indicate that current anthropomorphic designs of CAs may achieve a level of human-likeness close to the beginning of the Uncanny Valley yet not reaching it. Alternatively, there could be a differently structured curve that describes the relationship between familiarity and human-likeness for CAs as MacDorman (2006), for example, report in their analysis of human reactions to robot video clips.

6.2 Implications for CA Design in Practice

Three practical implications for the design of anthropomorphic CAs, particularly in a service context, can be drawn from these results: First, the perception of humanness of a conversational service agent, enabled by a rich combination of social cues, positively contributes to familiarity and, in particular, service satisfaction. These results are in line with the findings of for example Gnewuch et al. (2018) or Diederich et al. (2019c). According to our data, crafting CAs with an appealing human-like representation and behavior is thus generally desirable in a service context.

Second, equipping the CA with sufficient conversational capabilities to mitigate and adequately handle response failures should be a key consideration when designing an agent due to the substantial negative impact of even modest response failures. While the designer's task of anticipating a wide variety of user input is admittedly a challenging one (Følstad and Brandtzæg, 2017), treating conversations as the core object of a CA's design is essential to build agents that fulfill user expectations and are able to maintain a human-like behavior (Clark et al., 2019). Furthermore, designers should carefully reflect on and select coping strategies to handle unanticipated situations in a conversation (see for example Ashktorab et al. (2019) for an overview and initial evaluation).

Third, due to the absence of very strong negative effects as posited by the Theory of the Uncanny Valley, a human-like design is favorable even if small response failure may take place. Accordingly, with response failures still occurring in practice, equipping the CA with various social cues seems to be nonetheless advantageous, at least for a modest number of conversational fallbacks

6.3 Limitations and Opportunities for Future Research

Our research exhibits different limitations and indicates opportunities for future research on anthropomorphic CA design. The experimental setting offered the benefit of control yet lacked realism (Dennis and Valacich, 2001). Similar to other studies on CA design, we provided the participants with a set of rather specific tasks. Hence, we were able to create a setting in which the agent consistently failed to provide a meaningful response for around two times, which allowed us to better understand the impact on user perception of the agent by comparing the experimental groups. However, in a practical interaction with a CA, response failures are likely to occur with different frequency depending on the agent's design. In addition, other reasons for response failure exist like spelling errors by a user or out-of-context questions, which were not in the scope of this experiment yet represent a worthwhile opportunity for future research. A further limitation exists with regard to our measurement of familiarity, which was based on a single item as done in other studies, such as (MacDorman, 2006). In addition, our experiment was conducted in a specific context (customer service of an online retailer) with users expecting the agent to be able to fulfill their rather trivial service request (order cancellation). Therefore, we suggest that future studies explore the impact of response failure in different, potentially more complex (service) contexts and, in particular, in situations where the agent is not able to ultimately fulfill a customer's request.

Furthermore, our findings concerning the rather strong detrimental impact of agent response failures in a conversation offer two directions for future studies. First, design-oriented research can be conducted to investigate how response failure can be mitigated, such as by providing a more transparent structure with the agent or having the CA suggest answers in the interaction as frequently done in practice by means of quick reply buttons, thereby leading the conversation in a direction where the agent is able to provide relevant responses again. Second, different approaches in a service context to react to unexpected input can be conceptualized and empirically tested, such as polite and personal context-specific fallback responses or offering the possibility to contact a human service employee.

7 Concluding Remarks

CAs in organizational contexts promise to provide automated service that is always available and resembles the feeling of a human interaction. However, the limited capabilities of current agents often lead to situations in which agents fail to provide meaningful replies in a service encounter. As such response failures may negatively impact the perception of anthropomorphic CAs and are neglected in current research, we conducted an experiment to better understand the relationship between response failure and user perception of CAs.

Our findings provide evidence for the detrimental impact of modest response failures on the perception of humanness and uncanniness and highlight the positive (negative) impact of humanness (uncanniness) on familiarity and service satisfaction. Furthermore, our data does not indicate a very strong negative emotional reaction to response failure as proposed by the Uncanny Valley, but highlights a comparatively moderate negative effect. The findings from our experiment have implications for research, especially regarding the Uncanny Valley-effect in the context of anthropomorphic CAs. Our results further provide practical insights by confirming the positive impact of human-like CA design for innovative service provision and by emphasizing the need to mitigate response failures in natural language interactions.

8 Acknowledgments

This research was partly supported by "Mobility Opportunities Valuable to Everyone" (MOVE), an Interreg project supported by the North Sea Program of the European Regional Development Fund of the European Union.

References

- 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.
- Ashktorab, Z., M. Jain, V. Q. Liao and J. D. Weisz. (2019). "Resilient chatbots: Repair strategy preferences for conversational breakdowns." In: *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (pp. 1–12). Glasgow, Scotland.
- Beer, J. M., C. A. Smarr, A. D. Fisk and W. A. Rogers. (2015). "Younger and older users' recognition of virtual agent facial expressions." *International Journal of Human Computer Studies*, 75, 1–20.
- Ben Mimoun, M. S., I. Poncin and M. Garnier. (2012). "Case study-Embodied virtual agents: An analysis on reasons for failure." *Journal of Retailing and Consumer Services*, 19(6), 605–612.
- Bickmore, T. W. and R. W. Picard. (2005). "Establishing and Maintaining Long-term Human-computer Relationships." *ACM Transactions on Computer-Human Interaction*, 12(2), 293–327.
- Burgoon, J. K., J. A. Bonito, B. Bengtsson, C. Cederberg, M. Lundeborg and L. Allspach. (2000). "Interactivity in human-computer interaction: A study of credibility, understanding, and influence." *Computers in Human Behavior*, 16(6), 553–574.
- Cafaro, A., H. H. Vilhjalmsson and T. Bickmore. (2016). "First impressions in human-agent virtual encounters." *ACM Transactions on Computer-Human Interaction*, 24(4), 1–40.
- Chin, W. W. (1998). "The Partial Least Squares Approach for Structural Equation Modeling." In: L. Erlbaum (Ed.), *Modern Methods for Business Research* (pp. 295–336). Hillsdale, New Jersey.
- Clark, L., N. Pantidi, O. Cooney, P. Doyle, D. Garaialde, J. Edwards, ... B. R. Cowan. (2019). "What makes a good conversation? Challenges in designing truly conversational agents." In: *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (pp. 1–12). Glasgow, Scotland.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences* (2nd Editio). Taylor & Francis Inc.
- Cowell, A. J. and K. M. Stanney. (2005). "Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility." *International Journal of Human Computer Studies*, 62(2), 281–306.
- Crockett, K., A. Latham and N. Whitton. (2017). "On predicting learning styles in conversational intelligent tutoring systems using fuzzy decision trees." *International Journal of Human Computer Studies*, 97, 98–115.
- de Visser, E. J., S. S. Monfort, R. McKendrick, M. A. B. Smith, P. E. McKnight, F. Krueger and R. Parasuraman. (2016). "Almost human: Anthropomorphism increases trust resilience in cognitive agents." *Journal of Experimental Psychology: Applied*, 22(3), 331–349.
- Dennis, A. R. and J. S. Valacich. (2001). "Conducting Experimental Research in Information Systems." *Communications of the Association for Information Systems*, 7(5), 1–41.
- Diederich, S., A. B. Brendel and L. M. Kolbe. (2019a). "On Conversational Agents in Information Systems Research: Analyzing the Past to Guide Future Work." In: *Proceedings of the International Conference on Wirtschaftsinformatik* (pp. 1550–1564).
- Diederich, S., A. B. Brendel and L. M. Kolbe. (2020). "Designing Anthropomorphic Enterprise Conversational Agents." *Business & Information Systems Engineering*.
- Diederich, S., A. B. Brendel, S. Lichtenberg and L. M. Kolbe. (2019b). "Design for Fast Request Fulfillment or Natural Interaction? Insights from an Online Experiment with a Conversational Agent." In: *Proceedings of the European Conference on Information Systems (ECIS)* (pp. 1–17). Stockholm, Sweden.
- Diederich, S., M. Janßen-Müller, A. B. Brendel and S. Morana. (2019c). "Emulating Empathetic Behavior in Online Service Encounters with Sentiment-Adaptive Responses: Insights from an Experiment with a Conversational Agent." In: *Proceedings of the International Conference on Information Systems (ICIS)* (pp. 1–17). Munich, Germany.
- Diederich, S., S. Lichtenberg, A. B. Brendel and S. Trang. (2019d). "Promoting Sustainable Mobility Beliefs with Persuasive and Anthropomorphic Design: Insights from an Experiment with a

- Conversational Agent.” In: *Proceedings of the International Conference on Information Systems (ICIS)* (pp. 1–17). Munich, Germany.
- Dolata, M., M. Kilic and G. Schwabe. (2019). “When a computer speaks institutional talk: Exploring challenges and potentials of virtual assistants in face-to-face advisory services.” In: *Proceedings of the Hawaii International Conference on System Sciences (HICSS)* (pp. 105–114). Grand Wailea, Maui.
- Duffy, B. R. (2003). “Anthropomorphism and the social robot.” *Robotics and Autonomous Systems*, 42(3–4), 177–190.
- Epley, N., A. Waytz and J. T. Cacioppo. (2007). “On Seeing Human: A Three-Factor Theory of Anthropomorphism.” *Psychological Review*, 114(4), 864–886.
- Feine, J., U. Gnewuch, S. Morana and A. Maedche. (2019). “A Taxonomy of Social Cues for Conversational Agents.” *International Journal of Human-Computer Studies*, 132(December), 138–161.
- Følstad, A. and P. B. Brandtzæg. (2017). “Chatbots and the new world of HCI.” *Interactions*, 24(4), 38–42.
- Gefen, D. and D. Straub. (2005). “A practical guide to factorial validity using PLS-graph: tutorial and annotated example.” *Communications of the Association for Information Systems*, 91–109.
- Gnewuch, U., S. Morana, M. T. P. Adam and A. Maedche. (2018). “Faster Is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction.” In: *Proceedings of the European Conference on Information Systems (ECIS)* (pp. 1–17). Portsmouth, United Kingdom.
- Gnewuch, U., S. Morana and A. Maedche. (2017). “Towards Designing Cooperative and Social Conversational Agents for Customer Service.” In: *Proceedings of the International Conference on Information Systems (ICIS)* (pp. 1–13). Seoul, Korea.
- Go, E. and S. S. Sundar. (2019). “Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions.” *Computers in Human Behavior*, 97, 304–316.
- Gong, L. (2008). “How social is social responses to computers? The function of the degree of anthropomorphism in computer representations.” *Computers in Human Behavior*, 24(4), 1494–1509.
- Google. (2019). “Google Dialogflow.” Retrieved from <https://dialogflow.com>
- Holtgraves, T. and T. L. Han. (2007). “A procedure for studying online conversational processing using a chat bot.” *Behavior Research Methods*, 39(1), 156–163.
- Hu, T., A. Xu, Z. Liu, Q. You, Y. Guo, V. Sinha, ... R. Akkiraju. (2018). “Touch Your Heart: A Tone-aware Chatbot for Customer Care on Social Media.” In: *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (pp. 1–12). Montréal, Canada.
- Johnson, K. (2018). “Facebook Messenger hits 100,000 bots.” Retrieved from <https://venturebeat.com/2017/04/18/facebook-messenger-hits-100000-bots/>
- Liao, Q. V., M. M. Hussain, P. Chandar, M. Davis, M. Crasso, D. Wang, ... W. Geyer. (2018). “All Work and no Play? Conversations with a Question-and-Answer Chatbot in the Wild.” In: *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (pp. 1–13). Montréal, Canada.
- Luger, E. and A. Sellen. (2016). ““Like Having a Really Bad PA”: The Gulf between User Expectation and Experience of Conversational Agents.” In: *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (pp. 5286–5297). San José, USA.
- MacDorman, K. F. (2006). “Subjective Ratings of Robot Video Clips for Human Likeness, Familiarity, and Eeriness: An Exploration of the Uncanny Valley.” In: *Proceedings of the ICCS/CogSci-2006 long symposium: Toward social mechanisms of android science* (pp. 1–4). Canada.
- MacDorman, K. F., R. D. Green, C. C. Ho and C. T. Koch. (2009). “Too real for comfort? Uncanny responses to computer generated faces.” *Computers in Human Behavior*, 25(3), 695–710.
- Maedche, A., C. Legner, A. Benlian, B. Berger, H. Gimpel, T. Hess, ... M. Söllner. (2019). “AI-Based Digital Assistants.” *Business & Information Systems Engineering*, (4), 1–28.
- Marinova, D., K. de Ruyter, M. H. Huang, M. L. Meuter and G. Challagalla. (2017). “Getting Smart:

- Learning From Technology-Empowered Frontline Interactions.” *Journal of Service Research*, 20(1), 29–42.
- McTear, M. (2017). “The rise of the conversational interface: A new kid on the block?” *Lecture Notes in Computer Science*.
- McTear, M., Z. Callejas and D. Griol. (2016). *The Conversational Interface: Talking to Smart Devices*. Basel, Switzerland: Springer Publishing Company.
- Morana, S., C. Friemel, U. Gnewuch, A. Maedche and J. Pfeiffer. (2017). “Interaktion mit smarten Systemen – Aktueller Stand und zukünftige Entwicklungen im Bereich der Nutzerassistenz.” *Wirtschaftsinformatik & Management*, 5, 42–51.
- Mori, M. (1970). “The Uncanny Valley.” *Energy*.
- Mori, M., K. F. MacDorman and N. Kageki. (2012). “The Uncanny Valley.” *IEEE Robotics and Automation Magazine*, 19(2), 98–100.
- Nass, C. and Y. Moon. (2000). “Machines and mindlessness: Social responses to computers.” *Journal of Social Issues*, 56(1), 81–103.
- NextIT. (2018). “Helping a railroad service conduct business.” Retrieved from <http://nextit.com/case-studies/amtrak>
- Numamaker, J. F., D. C. Derrick, A. C. Elkins, J. K. Burgoon and M. W. Patton. (2011). “Embodied Conversational Agent-Based Kiosk for Automated Interviewing.” *Journal of Management Information Systems*, 28(1), 17–48.
- Oracle. (2016). “Can Virtual Experiences Replace Reality? The future role for humans in delivering customer experience,” 19.
- Pereira, A. T., R. Prada and A. Paiva. (2014). “Improving social presence in human-agent interaction.” In: *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (pp. 1449–1458). Toronto, Canada.
- Pfeuffer, N., A. Benlian, H. Gimpel and O. Hinz. (2019). “Anthropomorphic Information Systems.” *Business & Information Systems Engineering*, 1–16.
- Qiu, L. and I. Benbasat. (2010). “A study of demographic embodiments of product recommendation agents in electronic commerce.” *International Journal of Human Computer Studies*, 68(10), 669–688.
- Reeves, B. and C. Nass. (1996). *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*. The Center for the Study of Language and Information Publications.
- Sah, Y. J. and W. Peng. (2015). “Effects of visual and linguistic anthropomorphic cues on social perception, self-awareness, and information disclosure in a health website.” *Computers in Human Behavior*, 45, 392–401.
- Schuetzler, R. M., G. M. Grimes and J. S. Giboney. (2018). “An Investigation of Conversational Agent Relevance, Presence, and Engagement.” In: *Proceedings of the Americas Conference on Information Systems (AMCIS)* (pp. 1–10). New Orleans, USA.
- Seeger, A.-M., J. Pfeiffer and A. Heinzl. (2018). “Designing Anthropomorphic Conversational Agents: Development and Empirical Evaluation of a Design Framework.” In: *Proceedings of the International Conference on Information Systems (ICIS)* (pp. 1–17). San Francisco, USA.
- Seymour, M., K. Riemer and J. Kay. (2018). “Actors, Avatars and Agents: Potentials and Implications of Natural Face Technology for the Creation of Realistic Visual Presence.” *Journal of the Association for Information Systems*, 19, 953–981.
- Sohn, S. (2019). “Can Conversational User Interfaces Be Harmful? The Undesirable Effects on Privacy Concern.” In: *Proceedings of the International Conference on Information Systems* (pp. 1–9). Munich, Germany.
- Stock, R. M. and M. Merkle. (2018). “Customer Responses to Robotic Innovative Behavior Cues During the Service Encounter.” In: *Proceedings of the International Conference on Information Systems (ICIS)* (pp. 1–17). San Francisco, USA.
- Stock, R., M. Merkle, D. Eidens, M. Hannig, P. Heineck, M. A. Nguyen and J. Völker. (2019). “Understanding Employee Trust in Assistive Robots When Robots Enter Our Workplace: Understanding Employee Trust in Assistive Robots.” In: *Proceedings of the International*

- Conference on Information Systems* (pp. 1–9). Munich, Germany.
- Tinwell, A. and R. J. S. Sloan. (2014). “Children’s perception of uncanny human-like virtual characters.” *Computers in Human Behavior*, 36, 286–296.
- Urbach, N. and F. Ahlemann. (2010). “Structural Equation Modeling in Information Systems Research Using Partial Least Squares.” *Journal of Information Technology Theory and Application (JITTA)*, 11(2), 5–40.
- Vaccaro, K., T. Agarwalla, S. Shivakumar and R. Kumar. (2018). “Designing the Future of Personal Fashion Experiences Online.” In: *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (pp. 1–11). Montréal, Canada.
- Verhagen, T., J. van Nes, F. Feldberg and W. van Dolen. (2014). “Virtual customer service agents: Using social presence and personalization to shape online service encounters.” *Journal of Computer-Mediated Communication*, 19(3), 529–545.
- Wang, N., W. L. Johnson, R. E. Mayer, P. Rizzo, E. Shaw and H. Collins. (2008). “The politeness effect: Pedagogical agents and learning outcomes.” *International Journal of Human Computer Studies*, 66(2), 98–112.
- 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.
- Wunderlich, N. V. and S. Paluch. (2017). “A Nice and Friendly Chat With a Bot: User Perceptions of AI-based Service Agents.” In: *Proceedings of the International Conference on Information Systems (ICIS)* (pp. 1–11). Seoul, Korea.