Who’s Bad? – The Influence of Perceived Humanness on Users’ Intention to Complain about Conversational Agent Errors to Others

Fabian Hildebrandt
Technische Universität Dresden, fabian.hildebrandt@tu-dresden.de

Sascha Lichtenberg
Technisch Universität Dresden, sascha.lichtenberg@tu-dresden.de

Benedikt Brendel
Technisch Universität Dresden, Alfred_benedikt.brendel@tu-dresden.de

Elisa Landmann
Technische Universität Dresden, elisa.landmann@tu-dresden.de

Who's Bad? – The Influence of Perceived Humanness on Users' Intention to Complain about Conversational Agent Errors to Others

Completed Research Paper

Fabian Hildebrandt
TUD Dresden University of Technology
Dresden, Germany
fabian.hildebrandt@tu-dresden.de

Sascha Lichtenberg
TUD Dresden University of Technology
Dresden, Germany
sascha.lichtenberg@tu-dresden.de

Alfred Benedikt Brendel
TUD Dresden University of Technology
Dresden, Germany
alfred_benedikt.brendel@tu-dresden.de

Elisa Landmann
TUD Dresden University of Technology
Dresden, Germany
elisa.landmann@tu-dresden.de

Abstract

The perception of humanness in a conversational agent (CA) has been shown to strongly impact users' processing and reaction to it. However, it is largely unclear how this perception of humanness influences users' processing of errors and subsequent intention for negative word-of-mouth (WoM). In this context, we propose two pathways between perceived humanness and negative WoM: a cognitive pathway and an affective pathway. In a 2x2 online experiment with chatbots, we manipulated both the occurrence of errors and the degree of humanlike design. Our findings indicate that perceived humanness effects users' intentions towards negative WoM through the cognitive pathway: users' confirmation of expectations is increased by perceived humanness, reducing negative WoM intentions. However, it has no effect on users' anger and frustration and does not interact with the effects of errors. For practice, our results indicate that adding humanlike design elements can be a means to reduce negative WoM.

Keywords: Conversational Agents, Errors, Anthropomorphism, Perceived Humanness, Anger, Frustration, Negative Word of Mouth

Introduction

Within the last few years, many companies have begun to implement conversational agents (CAs) due to the recent technological advancements in this area (Nicolescu & Tudorache, 2022), which have led to the widespread availability of development tools and platforms (e.g., Google's Dialogflow, IBM's Watson, and ChatGPT) (Diederich et al., 2019; Hughes et al., 2023). CAs are defined as "software-based systems designed to interact with humans using natural language" (Feine, Gnewuch, et al. 2019, p. 1.). Common examples are the chatbots used by e-commerce platforms (e.g., Amazon's chat assistant), smart home assistants (e.g., Amazon's Alexa), and voice assistants in smartphones (e.g., Apple's Siri) (McTear, 2017; Moussawi et al., 2022; Nicolescu & Tudorache, 2022). Many companies now offer CAs as interfaces for customers, to provide services that were previously restricted to human employees, such as sales (Adam et al., 2022) or financial advising (Back et al., 2023). The benefits of CAs are the easy-to-learn interaction via natural language, independence of time and place, and overall convenience (Hughes et al., 2023; Nicolescu & Tudorache, 2022; Verhagen et al., 2014). Many examples show that the use of CAs leads to cost savings and increased revenue (Feine, Morana, et al., 2019; McTear et al., 2016).

Despite the efforts of designers, developers, and technological advancements, however, CAs are prone to producing errors due to the complexity of natural language interactions (Brandtzæg & Følstad, 2018), the increasing expectations of users (Hughes et al., 2023), and the constantly evolving nature of human language (Christiansen & Kirby, 2003). In the past, many CAs have been discontinued, as they were ineffective in serving the majority of customers: either the dialogue was not efficient, or the responses were not consistently meaningful (Ben Mimoun et al., 2012; Janssen et al., 2021). Errors in the interactions between CAs and their users can negatively impact perceptions of CAs and the associated service: for example, when a chatbot does not understand the user input and responds with a fallback message (e.g., "Sorry, can you rephrase your request?"), this can result in dissatisfied and angry customers, leading to negative word of mouth (WoM). In the context of CAs, this means that a user speaks negatively about the CAs or the operating company to others, which can have serious consequences such as loss of potential customers and revenue (East et al., 2008; Lau & Ng, 2001; Williams & Buttle, 2014). Since negative WoM has a stronger impact than positive WoM (Arndt, 1967), it is of vital importance to understand how the design of CA interacts with users' intention towards negative WoM when errors occur.

CAs have a key difference from software with traditional graphical interfaces: they can be designed to appear humanlike, which is common in practice (Seeger et al., 2018). The term humanlike design refers to equipping a CA with humanlike elements (called social cues), such as a human name and avatar, using emoticons, and greeting users (Feine, Gnewuch, et al., 2019). This humanlike design induces a perception of humanness in users (Gnewuch et al., 2017; Nass & Moon, 2000), which influences their perceived service satisfaction and intention to use (Diederich et al., 2020; Gnewuch, Morana, et al., 2018). However, research on the influence of perceived humanness on users' intention toward negative WoM is limited (van Pinxteren et al., 2020). Building upon existing theory and evidence from studies on related phenomena, we derive how cognitive and affective pathways could explain how the intention towards negative WoM is formed under the influence of perceived humanness when an error occurs. Against this background, this study aims to answer the following research question:

RQ: What is the influence of perceived humanness on user intention towards negative WoM?

To answer this question, we conducted a 2x2 treatment design within an online experiment involving 179 participants, in which we varied the humanlike design and the occurrence of errors. Based on users' responses, we analyzed the relations between the occurrence of an error, perceived humanness, anger and frustration, expectation confirmation, and intention towards negative WoM. Our results show that perceived humanness directly increases confirmation of expectations but has no moderating effect on the influence of the error on expectation confirmation. Similarly, perceived humanness does not influence whether confirmation of expectations leads to intention towards negative WoM. In terms of the affective pathway, perceived humanness plays no role: it does not influence the effect of an error on emotions, and has no effect on the relation between anger and frustration and users' intention towards negative WoM.

Research Background and Related Work

The complexity of natural language interactions often leads to errors by CAs (Brandtzæg & Følstad, 2018), and their effectiveness relies heavily on the developers and technology employed (Brandtzæg & Følstad, 2018; Verhagen et al., 2014). The use of limited amounts of training data and vocabulary often cause errors, and poorly developed and trained CAs struggle to process user requests, resulting in errors (Brandtzæg & Følstad, 2018; Zemčík, 2021). In addition, human language is constantly evolving, with the introduction of new words and phrases or alterations in the use of existing ones (Christiansen & Kirby, 2003).

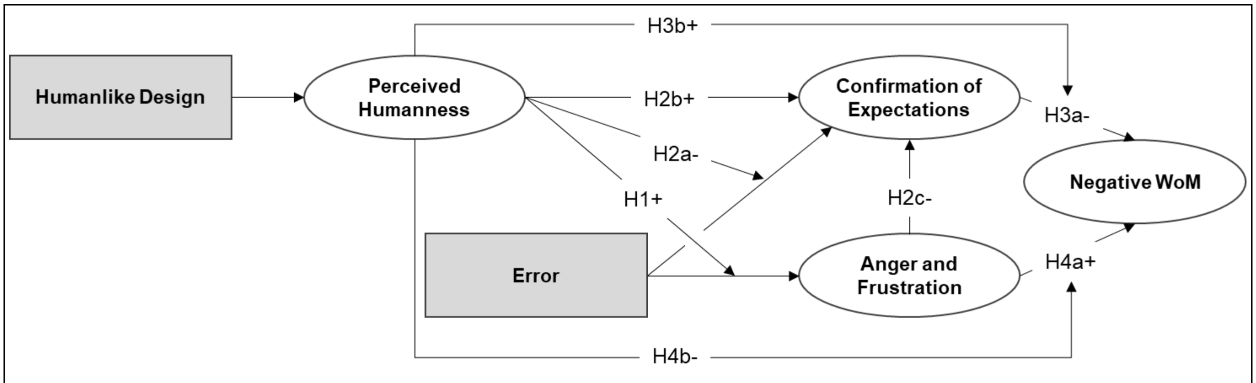
In general, users form expectations about a service interaction, and its outcome that are subsequently compared with the actual service (Oliver, 1981), and a CA may meet, exceed, or fail to meet these expectations (Crolic et al., 2021; Diederich et al., 2022). Errors are unexpected, and therefore violate users' expectations (Ben Mimoun et al., 2012; Diederich et al., 2021; Oliver, 1981). The yield-shift theory of satisfaction (Briggs et al., 2008) suggests that the evaluation of whether expectations were met has both cognitive and emotional aspects: users deliberately and cognitively evaluate the service and its outcome, but services also affect individuals' emotions during and after the interaction. For instance, when the goal of the service is not met, anger and frustration are common emotional reactions (Briggs et al., 2008; Clore & Centerbar, 2004; Crolic et al., 2021; Spexard et al., 2008).

Anger and frustration are understood as emotional reactions to a failure to reach a desired goal (Berkowitz, 1964; Dollard et al., 1939). Campos et al. (1994) found that anger and frustration are affective reactions to threats and obstacles that need to be overcome quickly to reach a goal, often leading to aggression (Bodenhausen et al., 1994; Martin et al., 2000). In the context of CAs, this means that when an interaction with a CA does not help the user to achieve their goals, anger and frustration are the expected affective reactions of users. The resulting anger and frustration and disconfirmation of expectations are drivers of customers' negative WoM (Anderson, 1998; Richins, 1983). In a business context, WoM is a post-purchase/service behavior that involves interpersonal communication about a company's services and goods or the company itself, which influences and shapes the behaviors and attitudes of others towards the company (Duhan et al., 1997; Katz & Lazarsfeld, 1955; Litvin et al., 2008). In the context of CAs, this means that a CA user speaks negatively about the CA itself or the provider of the CA.

Research on negative WoM in the context of CAs remains scarce. Based on the recent comprehensive literature review of 262 research articles on CA carried out by Diederich et al. (2022), we were able to identify only a very limited set of studies. In their study, Lambert de Diesbach & Bagozzi (2022) showed that the use of a more human-like 3D avatar improves users' WoM. Seeger and Heinzl (2021) showed that a humanlike design can prevent users from losing trust in the CA, which can prevent negative WoM when an error occurs. Furthermore, in their literature review of 61 articles on the communication behavior of CAs, van Pinxteren et al. (2020) propose that WoM should be considered when studying CAs, but, based on their assessment, researchers have yet to engage with this topic. In this context, they highlight that there is strong evidence from human-to-human service interactions that should be used to develop hypotheses for interactions between users and CAs.

Research Model and Hypotheses

This study investigates the influence of perceived humanness of CAs on users' intention towards negative WoM regarding an error during the interaction (i.e., complaining about the CA and its error to others). The corresponding research model is illustrated in Figure 1. In the following sections, we will present the reasoning behind our hypothesis in detail.



Grey boxes represent treatments

Figure 1. Research Model

Perceived Humanness

As mentioned above, CAs can be equipped with social cues, such as a human name and an avatar, to make them appear more similar to humans, which is known as humanlike design (Seeger et al., 2018). These social cues trigger anthropomorphism in onlookers (Dacey, 2017). According to the CASA paradigm, the perception of humanness in computers, such as CAs equipped with social cues, is an automatic response (Nass & Moon, 2000): despite knowing that they are interacting with a computer rather than a human, users will still perceive some degree of humanness (Nass & Moon, 2000).

However, not all users perceive the same level of humanness in the same CA; in other words, a CA that is perceived as highly human by one user may not be viewed as such by another (Epley et al., 2007; Spatola &

Wudarczyk, 2021). Hence, despite the clear intentions of the CA designers, the perception of humanness cannot be forced, and is highly individual. We therefore implemented two different CAs, with the intention of producing different levels of perceived humanness, to test our hypotheses. In our theorizing, we will focus on the effect of perceived humanness.

Errors by Conversational Agents

Errors made by CAs (e.g., frequent failure to understand users' inputs or to provide the expected service) (Diederich et al., 2021; Riquel, Brendel, Hildebrandt, Greve, & Dennis, 2021) interfere with the users' goals (e.g., completing a service request) (Spexard et al., 2008). Individuals react with anger and frustration when they are unable to achieve their desired goals and outcomes (Berkowitz, 1964; Wetzer et al., 2007). Hence, in the context of CA service encounters, unexpected errors that prevent users from achieving their goals are likely to lead to anger and frustration. For example, Han et al. (2021) reported that users reacted with anger when a chatbot service was interrupted, and Goetsu and Sakai (2020) showed that errors by voice assistants frustrated users.

Errors also interfere with users' expectations, influencing the cognitive evaluation of expectations (i.e., users expect that the service is not interrupted by errors) (Jiang et al., 2002; Oliver, 1981). In the end, errors reduce users' confirmation of expectations. Some support for this view can be found in the literature; for instance, Lee et al. (2010) reported that errors by robots violate users' expectations, and similar results were reported by Ashktorab et al. (2019) in the context of chatbots. Overall, in the context of this study, it can be expected that a chatbot that is unable to complete the service will lead to anger and frustration as well as a reduction in users' confirmation of expectations.

Anger and Frustration

Although errors are already frustrating and anger-inducing in themselves (Riquel, Brendel, Hildebrandt, Greve, & Dennis, 2021), when individuals perceive intention behind errors, they become even angrier and/or more frustrated (González-Gómez et al., 2021). In the literature, various studies have reported on this effect in the context of human-to-human interaction. For instance, Harrison-Walker (2019) showed that customers typically feel the main emotions of frustration, anger, and regret after experiencing a service failure. Similarly, the experience of anger can be attributed to the combined presence of four factors: frustration, the perception of others being present, a sense of accountability for the situation, and the perception of hostile intentions from others (Van Mechelen & Hennes, 2009). In the context of CAs, there are some indications of similar effects, but no empirical study has been conducted to explore these. For instance, Riquel, Brendel, Hildebrandt, Greve, and Dennis (2021) reported that perceived humanness leads to greater frustration when an error occurs, and attributed this effect to the potential perception of intentions. In view of the reported effects of perceived humanness on frustration when errors occur, we put forward the following hypothesis:

H1: Greater perceived humanness increases the effect of error on anger and frustration.

Confirmation of Expectations

Humans are constantly making assumptions about what is likely to happen or what characteristics an entity such as a product or service will possess (Zeithaml et al., 1993). They then assess the extent to which these expectations are met by future events (Coye, 2004; Zeithaml et al., 1993). In general, people prefer their expectations to be confirmed, as this validates their predictions (Oliver, 1981). The evaluation of confirmation or disconfirmation is a subjective process that is influenced by the information that is available, and by emotions and biases (Boulding et al., 1993; Coye, 2004). All forms of thinking, including perception and information processing, are subjective in nature (Lerner et al., 2015; Levinson, 1995).

Social response theory (Nass & Moon, 2000) suggests that the perception of humanness leads users to apply social scripts and behavior to an interaction with a computer, including a CA (Lang et al., 2013; Nass & Moon, 2000). Thus, their expectations are (to some degree) similar to those associated with a human-to-human interaction (Diederich et al., 2022; Seeger & Heinzl, 2021). In general, humans are aware that other humans make errors, and hence expect some errors to occur (Mirnig et al., 2017; Renier et al., 2021). This general understanding of human nature ("to err is human") may also translate to an interaction with a CA that is perceived to be humanlike. Some support for this view can be found in the literature. For instance, Riquel, Brendel, Hildebrandt, Greve, and Kolbe (2021) found that despite the occurrence of an error, perceived humanness increased service satisfaction. Service satisfaction is very closely related to the confirmation of expectations (Oliver, 1981). It can therefore be deduced that the perception of humanness may increase confirmation of expectations, although empirical evidence for this has yet to be produced. We propose the following hypothesis:

H2a: Greater perceived humanness reduces the effect of error on confirmation of expectations.

The tendency of humans to attribute humanlike characteristics to non-human entities such as objects, commonly known as anthropomorphism (Epley et al., 2007), is relevant in this context. The perception of humanness can potentially impact the user's thought process and their assessment of expectation confirmation for two reasons. Firstly, humans generally enjoy social interaction (Levinson, 1995), which is likely to have a positive influence on the evaluation of confirmation of expectations, since in general, a favorable state of mind leads to more positive perceptions and evaluations (Blanchette & Richards, 2010). For instance, Babel et al. (2021) showed that users tend to trust a robot more when it is designed to be humanlike. Hence, a humanlike design influences the user's evaluation of the trustworthiness of the CA, despite no logical relationship between humanlike design and trust. Another example was reported by Pak et al. (2012), who found an increase in users' perceived performance if a CA was designed to be humanlike. However, to the best of our knowledge, no such evidence exists for the specific relationship between perceived humanness and confirmation of expectations.

Secondly, CAs are now used by many companies, meaning that people have become familiar with them (Araujo, 2018; Følstad & Brandtzaeg, 2017). Social cues are frequently added to the design of commercial CAs (Araujo, 2018; Seeger et al., 2018), which can lead to a user expecting them to be present when interacting with a CA; for instance, everyday CAs such as Amazon's Alexa are equipped with gendered voices. A lack of social cues would therefore violate the user's expectations. However, we were unable to identify any empirical studies on this matter.

In summary, there are two arguments that can be derived based on theory and research on related phenomena: (i) people enjoy social interaction, which positively influences the cognitive evaluation of expectation confirmation; and (ii) users have become accustomed to the use of social cues in CA interactions, leading them to expect these when interacting with a CA. However, studies have yet to be conducted on these topics, meaning that there is a lack of concrete evidence. In view of this, we hypothesize as follows:

H2b: Perceived humanness increases confirmation of expectations.

Affect infusion theory states that the emotional state of an individual influences their cognition (Forgas, 1995). Positive emotions (e.g., happiness) lead to positive thinking, whereas negative emotions (e.g., anger or sadness) lead to negative thinking (Forgas, 1995). In line with these arguments, a feeling of anger will negatively influence a customer's cognitive evaluation of a service (Liljander & Strandvik, 1997; Stauss et al., 2005). Anger and frustration are therefore expected to lead to a negative evaluation of the users' confirmation of expectations.

For instance, Stauss et al. (2005) showed that frustrating events in a service encounter (e.g., inaccessibility) lead to feelings of frustration in users, resulting in reduced service satisfaction (i.e., less confirmation of expectations). Furthermore, Crolic et al. (2021) found that users who took part in a service encounter while in an existing negative emotional state (e.g., anger or frustration) reported lower satisfaction. Similarly, Bougie et al. (2003) showed that customers experiencing failed service encounters felt high levels of anger and agitation, leading to lower levels of service satisfaction. In the light of these results, we formulate the following hypothesis:

H2c: Anger and frustration reduce confirmation of expectations.

Negative Word of Mouth

In general, the term WoM is used to describe a situation in which people share information with each other (Katz & Lazarsfeld, 1955; Litvin et al., 2008; Wetzer et al., 2007). In a business-to-customer setting, WoM refers to people sharing their experience of a product or service directly with others, for instance, talking about the product or service with their friends (Bougie et al., 2003; Huete-Alcocer, 2017; Lee & Song, 2010), and thus influencing the product and/or service choices, purchase intentions and attitudes of others (Balaji et al., 2016; Grégoire et al., 2015). Negative WoM, which is the more common form (Akhtar et al., 2019), is understood as customers speaking negatively about a service, product or company (Anderson, 1998). The literature shows that product malfunctions can lead to negative WoM (Anderson, 1998; Sundaram et al., 1998); this has parallels in faulty service interactions with a CA, if these prevent users from achieving their goals. When an individual's expectations are not met, negative WoM is likely to occur, since violation of expectations leads to a negative judgment of a service, product, or company (Litvin et al., 2008; Oliver, 1981). Thus, we propose the following hypothesis:

> *H3a: Confirmation of expectation reduces intention towards negative WoM.*

Following social response theory (Nass & Moon, 2000), perceived humanness of a CA triggers behavior associated with human-to-human interactions. In the context of negative WoM, users can be expected to be reluctant to talk negatively about the chatbot, because speaking badly behind someone's back is normally a type of behavior that has negative connotations (Hartung et al., 2019). More specifically, online disinhibition theory (i.e., humans tend to behave negatively when they are less aware of the humanness of others (Lapidot-Lefler & Barak, 2012; Suler, 2004)) suggests that users are likely to be inhibited because of the perceived humanness.

Support for this view can be found in the literature. For instance, based on a study of cyber trolling, Wu et al. (2023) reported that for Chinese college students, high online disinhibition increases the likeliness of trolling behavior. Although no research on this effect exists in the context of CAs, there is some evidence to support a relationship between perceived humanness and negative WoM. Brendel et al. (2023) found that an increase in perceived humanness reduces the likelihood of severe aggression (e.g., using swearwords) towards CAs. In view of this, we theorize as follows:

> *H3b: Greater perceived humanness increases the effect of confirmation of expectations on intentions for negative WoM.*

Given that anger and frustration are sources of WoM by customers (Benbasat & Wang, 2005), angry customers are more likely to engage in negative WoM (Anderson, 1998; Richins, 1983). For instance, Wetzer et al. (2007) reported that angry customers engaged in negative WoM both to take revenge and to vent their negative emotions. In the context of tourism, Sánchez-García & Currás-Pérez (2011) showed that anger leads to increased negative WoM. We therefore formulate the following hypothesis:

> *H4a: Anger and frustration increases intention towards negative WoM.*

When they are angry, users behave more instinctively (e.g., anger and frustration lead to hostile aggression (Anderson & Bushman, 2002)), and their behavior is strongly influenced by subconscious scripts. In this context, venting anger and frustration in form of negative WoM is a type of aggression, which is related to revenge (Sánchez-García & Currás-Pérez, 2011). Negative WoM may have the intention of damaging the company or the employees involved in the service encounter (e.g., bad mouthing employees to get them fired) (Wetzer et al., 2007), which is known as vindictive negative WoM (He & Harris, 2014). Following online disinhibition theory (Suler, 2004), we note that a CA that is perceived to have a high level of humanness should lead to inhibition; users are less likely to engage in negative WoM because they subconsciously do not want to hurt the CA (Hydock et al., 2020; Lapidot-Lefler & Barak, 2012; Suler, 2004). However, no research in the context of CA has addressed this relation. Against this background, based on evidence from the human-to-human context, we hypothesize:

> *H4b: Greater perceived humanness reduces the effect of anger and frustration on intention towards negative WoM.*

Method

Participants

We recruited 179 participants for our experiment via the SurveyCircle platform. Participants received no incentive from us for their voluntary participation. The user interaction involved a responsive web interface that could be accessed from any device (e.g., desktop PC, tablet or smartphone). Hence, participants could participate in our experiment at anytime, anywhere, and with any device. A total of five participants failed the attention checks that were conducted throughout the survey, and when these had been removed, our final sample consisted of 174 valid responses. The mean age of our participants was 27 (SD: 5.85), and 70% of them identified as women. In addition, 68% held a higher education degree (e.g., a Bachelor's or Master's degree).

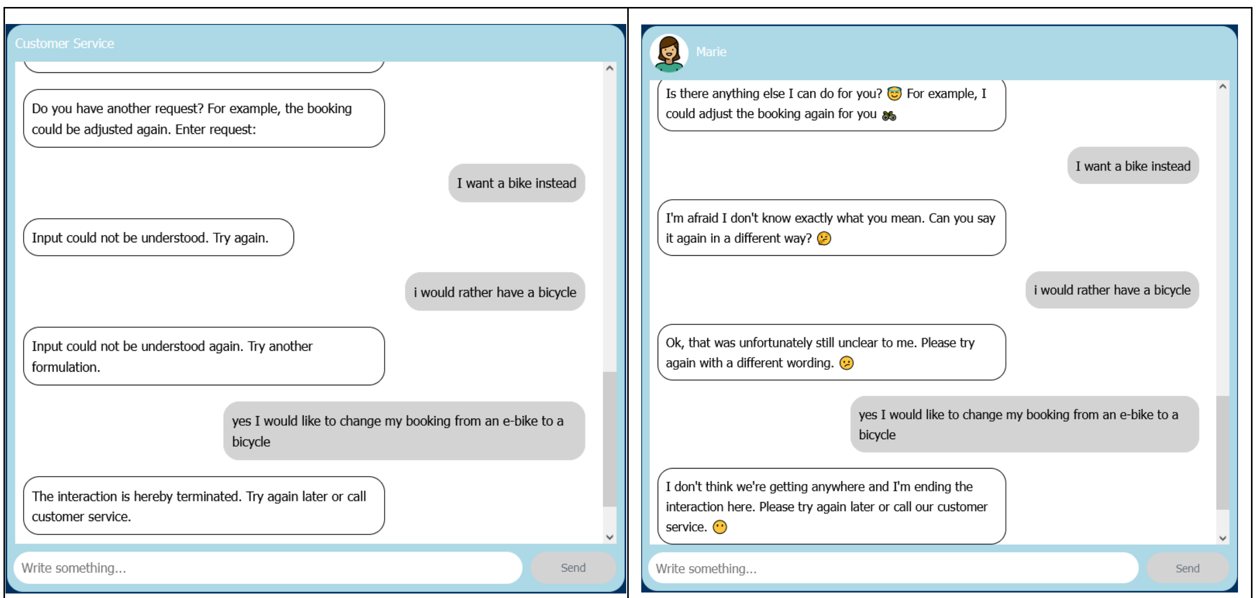
Task and Procedure

We developed our experimental task and procedure based on examples from other studies on the humanlike design of CAs (e.g., Brendel et al. (2020) and Gnewuch, Adam, et al. (2018)). Following these examples, our experiment involved participants completing a specific task with a chatbot, and the dialogue had a clear, task-oriented structure. This meant that each participant had a very similar interaction with the chatbot, as scenarios that would lead to vastly different experiences were omitted (for instance, tasking participants to undertake five minutes of casual conversation would lead to vastly different interactions for each of them).

All participants were provided with the same briefing information that explained their task: to use a chatbot to rent an e-bike, and then to revise the booking directly after completing it. It was explicitly explained in the briefing that the participants would interact with a chatbot (i.e., a computer program) rather than a human. Overall, the task consisted of nine steps: (1) start the booking process, (2) provide a date, (3) select a city, (4) state the reason for the rental, (5) chose one of the types of e-bikes offered, (6 & 7) input a first and last name, (8) provide an e-mail address, and (9) change the type of bike (from an e-bike to a regular bike). We explained to all participants that although the input should be realistic, they should not give out any personal information (e.g., their own e-mail address). Participants completed the experiment (including the briefing and survey at the end) within about 10 minutes.

Treatments

For this experiment, we decided to use a between-subject design to prevent carryover effects (Boudreau et al., 2001). Each participant was randomly assigned to one of the four chatbot treatments: (1) no-humanlike design and no error, (2) no-humanlike design and error, (3) humanlike design and no error, and (4) humanlike design and error.



Note: Dialogue translated from German to English

Figure 2. Examples of Treatments

For the humanlike chatbots (see Figure 2), we based our design on examples from other studies (e.g., Araujo, 2018; Diederich, Brendel, et al., 2020; Gnewuch, Morana, et al., 2018; Seeger et al., 2017) of the perceptions of humanness of CAs. The specific design was guided by an article by Seeger et al. (2018) and their three dimensions of humanlike design: human identity, verbal cues, and non-verbal cues. For each dimension, we selected social cues that resembled those used in prior studies. For the first dimension (human identity), we assigned a human name (Marie), and provided a stereotypical female-gendered avatar. The second dimension (verbal cues) was implemented in the form of a greeting ("Hello, my name is Marie"), self-referencing ("... can I do..."), and politeness ("Can you please..."). Lastly, non-verbal cues (third dimension) were implemented through the use of emojis and dynamic response delays with associated blinking dots.

In the error treatments (see Figure 2), an error was introduced in step nine: the chatbots did not understand the user's request to change the type of bike. The user in each case had to restate this request twice. After the second time, the chatbot terminated the interaction and forwarded the user to the survey, meaning that the user was unable to change the rental as expected. We introduced this type of error because a total breakdown of the interaction and service would be unexpected and unacceptable. Furthermore, this type of error could not be overlooked by participants (e.g., a typing error may have gone unnoticed), and similar errors may arise in practice (Brandtzæg & Følstad, 2018; Zemčík, 2021). In this context, our participants should view the error as realistic and as violating their expectations.

Except for the humanlike design of the chatbots and the occurrence of errors, all four chatbots were implemented identically, using the same interface and development platform (Google Dialogflow). They were also trained on the same data, enabling them to process input with different phrasings. They could extract, validate, and repeat parameters from the user input (e.g., repeating the date of the booking).

Measures

We used various constructs in our survey to evaluate the research model and hypotheses, including questions about perceived humanness (Gefen & Straub, 1997), confirmation of expectations (Bhattacherjee, 2001), anger and frustration (based on Rajaobelina et al., 2022; Ribeiro & Prayag, 2019), and negative WoM (based on Hamilton et al., 2014; Reichheld, 2003). All related items were measured using a seven-point Likert-scale.

Latent variable	Mean	SD	Loading
Perceived humanness (Cronbach's α = .939, CR = .953, AVE = .804)			
I felt a sense of human contact with the chatbot.	2.575	1.730	.903
I felt a sense of personalness with the chatbot.	1.989	1.402	.911
I felt a sense of sociability with the chatbot.	2.793	1.814	.863
I felt a sense of human warmth with the chatbot.	2.040	1.491	.922
I felt a sense of human sensitivity with the chatbot.	1.948	1.319	.881
Confirmation of expectations (Cronbach's α = .896, CR = .927, AVE = .761)			
My experience with the chatbot was better than what I had expected.	3.638	1.974	.936
The service provided by the chatbot was better than what I had expected.	3.569	1.960	.911
Overall, most of my expectations from using the chatbot were confirmed.	4.282	1.727	.867
The expectations I had from the chatbot were correct.	4.293	1.579	.766
Anger and frustration (Cronbach's α = .882, CR = .916, AVE = .732)			
During the chatbot interaction, I felt tense.	2.523	1.642	.832
During the chatbot interaction, I felt frustrated.	3.339	2.121	.866
During the chatbot interaction, I felt stressed.	2.552	1.642	.822
During the chatbot interaction, I felt angry.	2.454	1.780	.901
Negative WoM (Cronbach's α = .793, CR = .880, AVE = .713)			
I would complain to friends or colleagues about the chatbot.	2.937	1.912	.917
I would say (digital/in-person) negative things about this chatbot to others.	2.839	1.875	.903
I was dissatisfied with the overall interaction with the chatbot.	3.672	1.986	.695

Table 1. Validation of Measures

The survey also included questions about demographics (age, gender, and education), three attention checks, two manipulation checks to determine whether participants perceived the humanlike design ("extremely inhuman-like" to "extremely humanlike", using a nine-point semantic differential scale) and error (“The chatbot has made some errors”, using a seven-point Likert scale) as intended, and a realism check (“The scenario of the experiment was realistic”, using a seven-point Likert scale). Table 1 displays the measured constructs, their factor loadings, Cronbach's α, composite reliability (CR), mean, and standard deviation (SD).

We used a threshold value of .60 for factor loadings (Gefen & Straub, 2005) and thus did not need to exclude any of the items from our analysis. Furthermore, all measures were reliable, with a CR greater than .70 (Nunally, 1970) and a Cronbach's α greater than .70 (Cortina, 1993). Our convergent and discriminant validity analyses were also sufficient, as indicated by an average variance extracted (AVE) of at least .50 (Hair et al., 2010), and the Fornell-Larcker criterion was satisfied (see Table 2) (Fornell & Larcker, 1981).

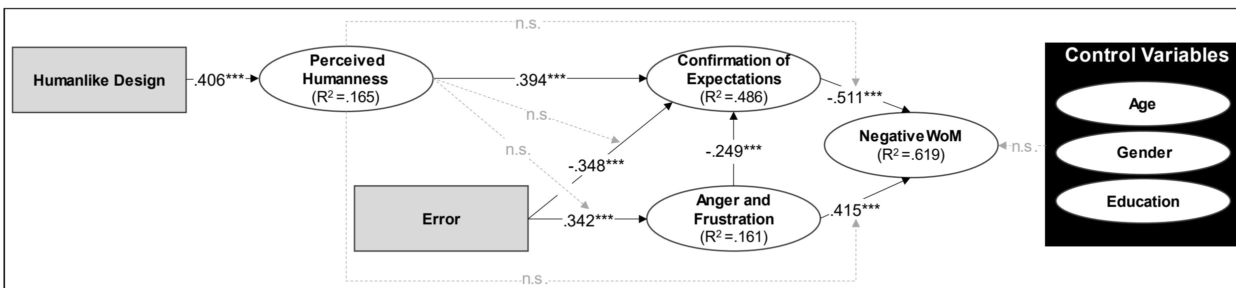
	1	2	3	4	5	6
1. Anger and frustration	**.856**					
2. Error	.367	**n.a.**				
3. Confirmation of expectations	-.461	-.500	**.872**			
4. Humanlike design	-.001	.012	.233	**n.a.**		
5. Perceived humanness	-.217	-.155	.496	.406	**.896**	
6. Negative WoM	.634	.469	-.700	-.076	-.326	**.844**
n.a. = not applicable						

Table 2. Discriminant validity

Results

Before we tested our hypotheses, we carried out some validation checks using a two-way analysis of variance (ANOVA). The results of the manipulation check for humanlike design ($F(1, 170) = 77.049$, $p <.001$) and error ($F(1, 170) = 121.967$, $p <.001$) were significant. The results of one-sample t-tests for the realism check were significantly higher than four (middle of the scale) for all treatments (humanlike with error: $t(46, 47) = 9.876$, $p < .001$; humanlike without error: $t(39, 40) = 6.230$, $p < .001$); non-humanlike with error $t(45, 46) = 7.114$, $p < .001$; non-humanlike without error: $t(40, 41) = 4.569$, $p < .001$)). In summary, our treatments were perceived as intended, and the scenario of the experiment was realistic.

To test our hypotheses on the relationship between the humanlike design of a CA, its errors, perceived humanness, confirmation of expectations, anger and frustration, and negative WoM, we applied a partial least squares (PLS) variance analysis. This analysis was performed using Smart PLS 3.3.7, and we used a bootstrapping resampling method with 5,000 samples to determine the significance of the path coefficients, as suggested by Chinn (1998). We chose a structural equation model (SEM) with latent variables for our research design, as it accounts for measurement errors and the multidimensional nature of theoretical constructs (Bagozzi & Yi, 1988). The PLS estimator is advantageous in terms of restrictive assumptions, and is commonly used in experimental research (Fombelle et al., 2016). Figure 3 illustrates the results, including the path coefficients, R^2 values, and significance levels.



** p < 0.05, ** p < 0.01, *** p < 0.001*
n. s. = not significant

Figure 3. Structural equation model (n = 174)

The use of a humanlike design for the CA was found to have a significant impact on perceived humanness ($\beta = .406$, $p < .001$), and the impact of implementing a CA with an error was also significant on anger and frustration ($\beta = .342$, $p < .001$) and confirmation of expectations ($\beta = -.348$, $p < .001$). Hypothesis **H1** involved the moderating effect of perceived humanness on the relation between error and anger and frustration, which was not found to be significant ($\beta = -.010$, $p = .870$), and **H1** was therefore not supported. Similarly, the moderating effect of perceived humanness on the relation of error and confirmation of expectations also showed no significance ($\beta = .064$, $p = .217$), meaning that **H2a** was not supported. However, perceived humanness had a significant direct effect on confirmation of expectations ($\beta = .394$, $p < .001$), thus supporting **H2b**. Anger and frustration were found to have a significant impact on confirmation of expectations ($\beta = -.249$, $p < .001$), providing support for **H2c**. The next hypotheses were related to the influence of confirmation of expectations and anger and frustration on the outcome variable, negative WoM. Confirmation of expectation was found to significantly decrease negative WoM ($\beta = -.511$, $p < .001$), thereby supporting **H3a**. However, this relationship was not significantly moderated by perceived humanness ($\beta = .055$, $p = .450$), meaning that **H3b** was not supported. Furthermore, we found evidence for **H4a** based on a significant increase in negative WoM due to anger and frustration ($\beta = .415$, $p < .001$). However, this relationship was not significantly moderated by perceived humanness ($\beta = .069$, $p = .336$), and hence **H4b** was not supported. Following Cohen (1988), the value of R^2 indicated a medium explanatory power for perceived humanness ($R^2 = .165$) and anger and frustration ($R^2 = .161$), and a large power for confirmation of expectations ($R^2 = .486$) and negative WoM ($R^2 = .619$).

We then examined the impact of the control variables (age, gender (1: female, 0: other), and higher education (1: university degree, 0: no university degree)) on the outcome variable, negative WoM. All control variables were found to have no significant impact on negative WoM (age: $\beta = .011$, p = .828, gender: $\beta = -.049$, p = .333, education: $\beta = .023$, p = .618). Also, we examined the specific indirect paths from the two treatments on negative WoM. A significant specific indirect path was found for error via the effect of anger and frustration on negative WoM ($\beta = .142$, $p < .001$). The use of a humanlike design was found to have a significant impact via humanness and anger and frustration ($\beta = -.028$, $p = .020$) and humanness and confirmation of expectations ($\beta = -.082$, $p = .001$). Regarding the mediation of perceived humanness on negative WoM, the total effect was significant ($\beta = -.279$, $p < .001$) but the direct effect was not ($\beta = .011$, p = .879). The indirect effects via anger and frustration ($\beta = -.068$, $p = .010$) and confirmation of expectations ($\beta = -.201$, $p < .001$) were significant, meaning that the relationship was fully mediated. Finally, the effect of error on negative WoM was fully mediated by anger and frustration due to a significant total effect ($\beta = .423$, $p < .001$), a non-significant direct effect ($\beta = .088$, p = .182), and a significant indirect effect ($\beta = .130$, $p < .001$).

Discussion

Although companies aim to develop CAs that provide great service, they are unlikely to ever be perfect, as their performance is limited by the technology available and the skills of the developers (Ben Mimoun et al., 2012; Brandtzæg & Følstad, 2018). Furthermore, the everchanging nature of human interaction, with the introduction of new words and phrases (Christiansen & Kirby, 2003), poses a challenge in terms of the ability of CAs to process input (Seeger et al., 2021). Consequently, CAs are prone to producing errors, such as a failure to understand user inputs and respond accordingly (Diederich et al., 2021). In this context, it is paramount to understand the consequences of errors (Benner et al., 2021), as people are likely to speak to others about their experiences. This WoM is of great importance for companies, because it influences their customer base and related revenue (Buttle, 1998; Verhagen et al., 2013).

In view of this, our study addresses how the use of a humanlike design and the resulting perception of humanness influence users' intention towards negative WoM. We derive two theoretical pathways for how perceived humanness influences negative WoM: (i) via a cognitive pathway, and (ii) via an affective pathway. Our results reveal that perceived humanness acts through the cognitive pathway, i.e., via a direct effect on confirmation of expectations. Perceived humanness does not influence the affective pathway. Our results provide new insights into the effect of perceived humanness on users when a CA produces errors and constitute evidence that perceived humanness influences cognitive processes related to expectation confirmation, while not acting on the user's anger and frustration.

Theoretical Implications and Future Research

First and foremost, our results do not point to a moderating effect of perceived humanness on the effects of the occurrence of an error. This goes against our theorizing based on social response theory and the evidence from the literature. Specifically, our results indicate that a perception of humanness does not increase the effects of error on anger and frustration, which we theorized in view of the tendency in human-to-human interactions whereby a perception of intention leads to greater anger and frustration (González-Gómez et al., 2021). Similarly, the perception of humanness does not reduce the negative effect of errors on confirmation of expectations, which contradicts our assumption that the user subconsciously views the CA as a human who may make errors (i.e., to err is human and, therefore, to some degree, expected). Overall, our results provide new insights into the extent to which humans follow social thinking patterns and behavior when perceiving a CA to be humanlike. In view of this, we see future research regarding the perception of intentions and expectation of errors as promising.

Secondly, the literature suggests that users may perceive some degree of intention behind an error made by a CA (Riquel, Brendel, Hildebrandt, Greve, & Dennis, 2021). In this context, we argue that our results indicate that understanding the attribution process of users is essential. Attribution theory states that errors can be attributed to either internal (to oneself, i.e., the user) or external causes (the CA or its developers). We propose that the participants in our study attributed the error either to themselves or to us, as the developers. Thus, perception of intention by the CA was unlikely, because the CA was not perceived as the cause of the error. Future research should therefore investigate when and how a CA's humanlike design and the related perception of humanness lead to attribution of the error to the CA. One promising approach could be to investigate social cues that express accountability (e.g., "I am sorry for the error *I caused*") or express some other constellation of responsibility (e.g., "I think we have to resolve *our issue* by…").

Thirdly, we expected the perception of humanness to lead users to expect some errors from the CA, as errors are part of human nature. However, this mindset was not triggered in our experiment. We believe that the different aspects of mind (Malle, 2019) may offer an explanation. Astington and Jenkins (1995) found that humans ascribe mind to other humans as well as to entities they perceive to be humanlike (e.g., animals or robots). In this context, mind has multiple components, such as agency, experience, moral and metal regulation, affect and reality interaction (Gray et al., 2007; Malle, 2019). For instance, agency involves planning, memory, and self-control, and we can expect this to play a role in users' formation of expectations (i.e., expecting errors). Another critical aspect of perceived agency is perceived intelligence (Seeger & Heinzl, 2021), for which it has already been shown in the context of CA that it influences CA adoption intention (Pillai & Sivathanu, 2020) and brand engagement (McLean et al., 2021). Thus, future research could investigate the effect of humanlike CA design and errors on perceived intelligence and how this further acts as a mediator. Similarly, the type of error might also play a role. The error we implemented caused a total breakdown of the interaction, which may have led users to think of the CA as a broken machine. In contrast, an error that does not disrupt the interaction (e.g., failing to understand an input once and then continuing the interaction) might fit the "to err is human" schema. We would welcome future research to challenge our two potential explanations: either support or contradiction would significantly increase our current understanding of the perception of humanness and social responses by CA users in the context of errors.

Furthermore, our results indicate that perceived humanness increases users' confirmation of expectations. Our results are amongst the first to support this relationship. We theorized that users would report greater confirmation of their expectations as they simply enjoy social interaction (Levinson, 1995; Nass & Moon, 2000) and/or because they have become accustomed to the presence of social cues (Seeger et al., 2018), and expect them when interacting with a new CA. Based on our results, future research could test whether one or both of these propositions are true, or whether a totally different explanation needs to be developed. Future research could investigate the user's state of mind and expectations before an interaction with a CA, for instance by asking them if they expect the CA to have a human name and to greet them. Similarly, an exploration of the influence of the user's experience with CAs could be of value. For example, Diederich et al. (2021) found that experience with CAs reduced the perception of humanness when an error occurred. This suggests that experience could drastically change the users' perceptions and subsequent reaction to a CA. For instance, experienced CA users may be aware that CAs make errors (e.g., Amazon's Alexa frequently failing to understand commands), leading to lowered expectations compared to users that are interacting with a CA for the first time. This proposition should be explored in future work to deepen our understanding of how users process errors made by CAs.

Lastly, we would like to highlight that not all errors made by CAs are the same: some are minor and might not be noticed by users (e.g., typing errors), while others are major (e.g., endangering the time, life, or money of users). For instance, Riquel, Brendel, Hildebrandt, Greve, and Dennis (2021) studied an error that endangered the incentive of participants (i.e., their chatbot behaved as if it did not understand that users wanted to continue to the survey and to receive their money afterwards). Taking a totally different approach, Brendel et al. (2020) implemented an error that resulted in a fictitious appointment for a tire change on Sunday, one hour before midnight; this did not represent a total service breakdown, but a service leading to an unintended outcome. In view of this, we suggest future research to gather and classify the different types of error that a CA can produce. In this way, research on the effects of errors can be clustered, and may reveal how the type of error changes the way in which users react to the CA.

Practical Implications

Our results indicate that a perception of humanness can mitigate the negative effects of errors. The occurrence of an error reduces users' confirmation of expectations, whereas the perception of humanness increases it. We would therefore advise CA designers to add social cues to their design. However, in light of the results of other studies (e.g., Riquel, Brendel, Hildebrandt, Greve, & Dennis, 2021) showing that the perception of humanness can increase frustration caused by errors), we would also recommend a cautious approach when adding social cues. Depending on the context and the errors produced by a CA, the addition of social cues to increase the perception of humanness may be a double-edged sword. On the one hand, it can counteract some of the negative effects of specific errors (e.g., increasing confirmation of expectations), while on the other, it can lead to counterproductive outcomes for other errors (e.g., increasing frustration). Hence, our results must be interpreted with caution, despite their overall positive implications.

Limitations

This work is not free of limitations. Firstly, our study suffers from the limitations and drawbacks typical of experimental work, as our experiment traded realism for controllability. Although we did our best to provide a CA interaction that was similar to real-world examples, it had no real-world consequences for our participants, as interacting with the chatbot did not lead to an actual rental of an e-bike. Thereby, a chosen scenario is always a limitation and in future research the results should be replicated in a different context and setting (e.g., order cancellation or rebooking). Indeed, the scenario chosen could impact not finding support for some hypotheses. Furthermore, the use of a humanlike design constitutes a limitation, in the same way as for all studies of this nature, as there is a nearly endless supply of social cues that can be combined to design a CA. Despite our best efforts to mimic the designs used in other established studies, future research is needed to challenge our results by implementing different designs and/or deliberately changing the social cues available. Thirdly, as outlined in our discussion, there are many different errors that a CA can produce, and our results are limited by the error selected for the experiment. We implemented an error that was realistic, but other errors might lead to different results (e.g., a typing error (Bührke et al., 2021)). Hence, replication of our results using different errors is required. Fourth, we combined frustration and anger in our measurements, as these are the main emotions we feel when goals are not achieved (Riquel, Brendel, Hildebrandt, Greve, & Dennis, 2021). Nevertheless, different emotions are at play and should subject to future research. Lastly, our sample of participants represents a limitation. The experiment was conducted in German, and we recruited our participants from a German-speaking population. Although we see few or no theoretical reasons why other samples or populations should give different results, testing our results with other populations (e.g., from other countries) could provide deeper insights into the role of user characteristics. In summary, our study provides evidence for the relationship between perceived humanness and users' intention to complain about CA errors, but these results need further replication.

Conclusion

The perception of humanness of a CA has a strong influence on its users. However, when a CA produces errors, it is unclear how the perception of humanness interacts with the occurrence of an error in terms of users' intention towards negative WoM. We theorized two main pathways by which perceived humanness could influence intentions towards negative WoM: a cognitive pathway, and an affective pathway. We conducted an 2x2 online experiment in which we varied the use of a humanlike design and the occurrence of errors. The results revealed that perceived humanness influenced users' intentions towards WoM via the cognitive pathway: perceived humanness increased users' confirmation of expectations, which reduced their intentions towards negative WoM. Our study makes three main contributions. Firstly, we are amongst the first to investigate the relationship between perceived humanness and confirmation of expectations in the context of CA interactions. Secondly, we show that perceived humanness influences users' cognition, and not their affect, when errors occur. Lastly, we show that implementing a humanlike design can be a double-edged sword in practice: although it can increase the confirmation of expectations, other studies of different types suggest that the perception of humanness may lead to negative outcomes (such as greater frustration). Thus, our results have significant implications for future research, and we suggest that practitioners should be very cautious when adding social cues to their CA designs.

References

Adam, M., Roethke, K., & Benlian, A. (2022). Human Versus Automated Sales Agents: How and Why Customer Responses Shift Across Sales Stages. *Information Systems Research*, *Articles in Advance*, 1–21.

Akhtar, M., Neidhardt, J., & Werthner, H. (2019). The potential of chatbots: Analysis of chatbot conversations. *Proceedings - 21st IEEE Conference on Business Informatics, CBI 2019*.

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology*, *53*(1), 27–51.

Anderson, E. W. (1998). Customer satisfaction and word of mouth. *Journal of Service Research*, *1*, 5–17.

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.

Arndt, J. (1967). Role of Product-Related Conversations in the Diffusion of a New Product. *Journal of Marketing Research*, *4*(3), 291–295.

Ashktorab, Z., Jain, M., Liao, Q. V., & Weisz, J. D. (2019). Resilient Chatbots: Repair Strategy Preferences for Conversational Breakdowns. *Proceedings of the 2019 Conference on Human Factors in Computing Systems*.

Astington, J. W., & Jenkins, J. M. (1995). Theory of Mind Development and Social Understanding. *Cognition and Emotion*, *9*(2–3), 151–165.

Babel, F., Kraus, J., Hock, P., Asenbauer, H., & Baumann, M. (2021). Investigating the validity of online robot evaluations: Comparison of findings from an one-sample online and laboratory study. *ACM/IEEE International Conference on Human-Robot Interaction*, 116–120.

Back, C., Morana, S., & Spann, M. (2023). When do robo-advisors make us better investors? The impact of social design elements on investor behavior. *Journal of Behavioral and Experimental Economics*, *103*(3), 101984.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*, 74–94.

Balaji, M. S., Khong, K. W., & Chong, A. Y. L. (2016). Determinants of negative word-of-mouth communication using social networking sites. *Information and Management*, *53*(4), 528–540.

Ben Mimoun, M. S., Poncin, I., & Garnier, M. (2012). Case study-Embodied virtual agents: An analysis on reasons for failure. *Journal of Retailing and Consumer Services*, *19*(6), 605–612.

Benbasat, I., & Wang, W. (2005). Trust In and Adoption of Online Recommendation Agents. *Journal of the Association for Information Systems*, *6*(3), 72–101.

Benner, D., Elshan, E., Schöbel, S., & Janson, A. (2021). What do you mean? A Review on Recovery Strategies to Overcome Conversational Breakdowns of Conversational Agents. *Proceedings of the 42nd International Conference on Information Systems (ICIS)*.

Berkowitz, L. (1964). Aggressive cues in aggressive behavior and hostility catharsis. *Psychological Review*, *71*(2), 104–122.

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly*, *25*(3), 351–370.

Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cognition and Emotion*, *24*(4), 561–595.

Bodenhausen, G. V., Sheppard, L. A., & Kramer, G. P. (1994). Negative affect and social judgment: The differential impact of anger and sadness. *European Journal of Social Psychology*, *24*, 45–62.
Boudreau, M. C., Gefen, D., & Straub, D. W. (2001). Validation in information systems research: A state-of-the-art assessment. *MIS Quarterly*, *25*(1), 1–16.
Bougie, R., Pieters, R., & Zeelenberg, M. (2003). Angry Customers don't Come Back, They Get Back: The Experience and Behavioral Implications of Anger and Dissatisfaction in Services. *Journal of the Academy of Marketing Science*, *31*, 377–393.
Boulding, W., Kalra, A., Staelin, R., & Zeithaml, V. A. (1993). A Dynamic Process Model of Service Quality: From Expectations to Behavioral Intentions. *Journal of Marketing Research*, *30*, 7–27.
Brandtzæg, P. B., & Følstad, A. (2018). Chatbots: Changing User Needs and Motivations. *Interactions*, *25*(5), 38–43.
Brendel, A. B., Greve, M., Diederich, S., Bührke, J., & Kolbe, L. M. (2020). "You are an idiot!" - How conversational agent communication patterns influence frustration and harassment. *Proceedings of the 26th Americas Conference on Information Systems (AMCIS)*.
Brendel, A. B., Hildebrandt, F., Dennis, A. R., & Riquel, J. (2023). The Paradoxical Role of Humanness in Aggression Towards Conversational Agents. *Journal of Management Information Systems*, *(Forthcomi*.
Briggs, R. O., Reinig, B. A., & de Vreede, G. J. (2008). The yield shift theory of satisfaction and its application to the IS/IT domain. *Journal of the Association for Information Systems*, *9*(5), 267–293.
Bührke, J., Brendel, A. B., Lichtenberg, S., Greve, M., & Mirbabaie, M. (2021). Is Making Mistakes Human? On the Perception of Typing Errors in Chatbot Communication. *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS)*.
Buttle, F. A. (1998). Word of mouth: Understanding and managing referral marketing. *Journal of Strategic Marketing*, *6*(3), 241–254.
Chin, W. W. (1998). The Partial Least Squares Approach to Structural Equation Modelling. In G. A. Marcoulides (Ed.), *Modern Methods for Business Research* (pp. 295–336). Lawrence Erlbaum Associates Publishers.
Christiansen, M. H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in Cognitive Sciences*, *7*(7), 300–307.
Clore, G. L., & Centerbar, D. B. (2004). Analyzing anger: How to make people mad. *Emotion*, *4*(2), 139–144.
Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences. In *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge.
Cortina, J. M. (1993). What Is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, *78*(1), 98–104.
Coye, R. W. (2004). Managing customer expectations in the service encounter. *International Journal of Service Industry Management*, *15*(1), 54–71.
Crolic, C., Thomaz, F., Hadi, R., & Stephen, A. T. (2021). Blame the Bot: Anthropomorphism and Anger in Customer–Chatbot Interactions. *Journal of Marketing*, *86*(1), 132–148.
Dacey, M. (2017). Anthropomorphism as cognitive bias. *Philosophy of Science*, *84*(5), 1152–1164.
Diederich, S., Brendel, A. B., & Kolbe, L. (2019). Towards a Taxonomy of Platforms for Conversational Agent Design. *Proceedings of the 14th International Conference on Wirtschaftsinformatik (WI)*, 1100–1114.
Diederich, S., Brendel, A. B., & Kolbe, L. M. (2020). Designing Anthropomorphic Enterprise Conversational Agents. *Business and Information Systems Engineering*, *62*, 193–209.
Diederich, S., Brendel, A. B., Morana, S., & Kolbe, L. (2022). On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research. *Journal of the Association for Information Systems*, *23*(1), 96–138.
Diederich, S., Lembcke, T.-B., Brendel, A. B., & Kolbe, L. M. (2021). Understanding the Impact that Response Failure has on How Users Perceive Anthropomorphic Conversational Service Agents: Insights from an Online Experiment. *AIS Transactions on Human-Computer Interaction*, *13*(1), 82–103.
Dollard, J., Miller, N. E., Doob, L. W., Mowrer, O. H., & Sears, R. R. (1939). *Frustration and aggression*. Yale University Press.
Duhan, D. F., Johnson, S. D., Wilcox, J. B., & Harrell, G. D. (1997). Influences on consumer use of word-of-mouth recommendation sources. *Journal of the Academy of Marketing Science*, *25*, 283–295.
East, R., Hammond, K., & Lomax, W. (2008). Measuring the impact of positive and negative word of mouth on brand purchase probability. *International Journal of Research in Marketing*, *25*, 215–224.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On Seeing Human: A Three-Factor Theory of Anthropomorphism. *Psychological Review*, *114*(4), 864–886.
Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human Computer Studies*, *132*(12), 138–161.
Feine, J., Morana, S., & Gnewuch, U. (2019). Measuring Service Encounter Satisfaction with Customer Service Chatbots using Sentiment Analysis. *Proceedings of the 14th International Conference on Wirtschaftsinformatik (WI)*, 1–11.
Følstad, A., & Brandtzaeg, P. B. (2017). Chatbots and the New World of HCI. *Interactions*, *24*(4), 38–42.
Fombelle, P. W., Bone, S. A., & Lemon, K. N. (2016). Responding to the 98%: Face-enhancing strategies for dealing with rejected customer ideas. *Journal of the Academy of Marketing Science*, *44*(6), 685–706.
Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin*, *117*(1), 39–66.
Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39–50.
Gefen, D., & Straub, D. (2005). A Practical Guide To Factorial Validity Using PLS-Graph: Tutorial And Annotated Example. *Communications of the Association for Information Systems*, *16*(1), 91–109.
Gefen, D., & Straub, D. W. (1997). Gender differences in the perception and use of e-mail: An extension to the technology acceptance model. *MIS Quarterly*, *21*(4), 389–400.
Gnewuch, U., Adam, M. T. P., Morana, S., & Maedche, A. (2018). “The Chatbot is typing …” - The Role of Typing Indicators in Human-Chatbot Interaction. *Proceedings of the 17th Annual Pre-ICIS Workshop on HCI Research in MIS*, 1–5.
Gnewuch, U., Morana, S., Adam, M. T. P., & Maedche, A. (2018). Faster Is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction. *Proceedings of the 26th European Conference on Information Systems (ECIS)*, 1–17.
Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards Designing Cooperative and Social Conversational Agents for Customer Service. *Proceedings of the 38th International Conference on Information Systems (ICIS)*.
Goetsu, S., & Sakai, T. (2020). Different Types of Voice User Interface Failures May Cause Different Degrees of Frustration. *ArXiv Preprint ArXiv:2002.03582*.
González-Gómez, H. V., Hudson, S., & Rychalski, A. (2021). The psychology of frustration: Appraisals, outcomes, and service recovery. *Psychology and Marketing*, *38*(9), 1550–1575.
Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619.
Grégoire, Y., Salle, A., & Tripp, T. M. (2015). Managing social media crises with your customers: The good, the bad, and the ugly. *Business Horizons*, *58*(2), 173–182.
Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th ed.). Pearson Education.
Hamilton, D. F., Lane, J. V., Gaston, P., Patton, J. T., MacDonald, D. J., Simpson, A. H. R. W., & Howie, C. R. (2014). Assessing treatment outcomes using a single question: The Net Promoter Score. *Bone and Joint Journal*, *96-B*(5), 622–628.
Han, E., Yin, D., & Zhang, H. (2021). Interruptions during a service encounter: Dealing with imperfect chatbots. *Proceedings of the 42nd International Conference on Information Systems (ICIS)*.
Harrison-Walker, L. J. (2019). The effect of consumer emotions on outcome behaviors following service failure. *Journal of Services Marketing*, *33*, 285–302.
Hartung, F. M., Krohn, C., & Pirschtat, M. (2019). Better than its reputation? Gossip and the reasons why we and individuals with “dark” personalities talk about others. *Frontiers in Psychology*, *10*, 1162.
He, H., & Harris, L. (2014). Moral disengagement of hotel guest negative WOM: Moral identity centrality, moral awareness, and anger. *Annals of Tourism Research*, *45*, 132–151.
Huete-Alcocer, N. (2017). A literature review of word of mouth and electronic word of mouth: Implications for consumer behavior. *Frontiers in Physiology*, *8*, 1256.
Hughes, L., Gauld, R., Grover, V., Hu, M., Edwards, J. S., Flavi, C., Janssen, M., Jones, P., Junglas, I., Khorana, S., & Kraus, S. (2023). “So What if ChatGPT Wrote It ?” Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy. *International Journal of Information Management*, *71*, 102642.
Hydock, C., Chen, Z., & Carlson, K. (2020). Why Unhappy Customers Are Unlikely to Share Their Opinions with Brands. *Journal of Marketing*, *84*, 95–112.
Janssen, A., Grützner, L., & Breitner, M. H. (2021). Why do Chatbots fail? A Critical Success Factors Analysis. *Proceedings of the 42nd International Conference on Information Systems (ICIS)*.

Jiang, J. J., Klein, G., & Carr, C. L. (2002). Measuring Information System Service Quality: SERVQUAL from the Other Side. *MIS Quarterly*, *26*(2), 145–166.

Katz, E., & Lazarsfeld, P. F. (1955). Personal Influence: The Part Played by People in the Flow of Mass Communications. *Free Press*.

Lambert de Diesbach, P. B., & Bagozzi, R. P. (2022). Do Embodied Virtual Agents Influence Word-of-Mouth and Loyalty Intentions?The Role of Functional, Hedonic and Brand Attitudes. *SSRN Electronic Journal*.

Lang, H., Seufert, T., Klepsch, M., Minker, W., & Nothdurft, F. (2013). Are Computers Still Social Actors? *Conference on Human Factors in Computing Systems - Proceedings*, 859–864.

Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior*, *28*(2), 434–443.

Lau, G. T., & Ng, S. (2001). Individual and situational factors influencing negative word-of-mouth behaviour. *Canadian Journal of Administrative Sciences*, *18*, 163–178.

Lee, M. K., Kielser, S., Forlizzi, J., Srinivasa, S., & Rybski, P. (2010). Gracefully mitigating breakdowns in robotic services. *5th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2010*.

Lee, Y. L., & Song, S. (2010). An empirical investigation of electronic word-of-mouth: Informational motive and corporate response strategy. *Computers in Human Behavior*, *26*(5), 1073–1080.

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and Decision Making. *Annual Review of Psychology*, *66*(1), 799–823.

Levinson, S. C. (1995). Interactional biases in human thinking. In E. N. Goody (Ed.), *Social Intelligence and Interaction* (pp. 221–260). Cambridge University Press.

Liljander, V., & Strandvik, T. (1997). Emotions in service satisfaction. *International Journal of Service Industry Management*, *8*(2), 148–169.

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, *29*, 458–468.

Malle, B. F. (2019). How Many Dimensions of Mind Perception Really Are There? *Proceedings of the 41st Annual Meeting of the Cognitive Science Society, CogSci 2019*, 2268–2274.

Martin, R., Watson, D., & Wan, C. K. (2000). A Three-Factor Model of Trait Anger: Dimensions of Affect, Behavior, and Cognition. *Journal of Personality*, *68*, 869–897.

McLean, G., Osei-Frimpong, K., & Barhorst, J. (2021). Alexa, do voice assistants influence consumer brand engagement? – Examining the role of AI powered voice assistants in influencing consumer brand engagement. *Journal of Business Research*, *124*, 312–328.

McTear, M. F. (2017). The rise of the conversational interface: A new kid on the block? *International Workshop on Future and Emerging Trends in Language Technology*, 38–49.

McTear, M. F., Callejas, Z., & Griol, D. (2016). Conversational Interfaces: Past and Present. In *The Conversational Interface* (pp. 51–72). Springer.

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 Robotics AI*, *4*(21), 1–23.

Moussawi, S., Koufaris, M., & Benbunan-Fich, R. (2022). The role of user perceptions of intelligence, anthropomorphism, and self-extension on continuance of use of personal intelligent agents. *European Journal of Information Systems*, *Online*, 1–18.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues: A Journal of the Society for the Psychological Studies of Social Issues*, *56*(1), 81–103.

Nicolescu, L., & Tudorache, M. T. (2022). Human-Computer Interaction in Customer Service: The Experience with AI Chatbots - A Systematic Literature Review. *Electronics*, *11*(10), 1579.

Nunally, J. C. (1970). *Introduction to psychological measurement*. McGraw-Hill.

Oliver, R. L. (1981). Measurement and Evaluation of Satisfaction Processes in Retail Settings. *Journal of Retailing*, *57*(3), 25–48.

Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics*, *55*(9), 1059–1072.

Pillai, R., & Sivathanu, B. (2020). Adoption of AI-based chatbots for hospitality and tourism. *International Journal of Contemporary Hospitality Management*, *32*(10), 3199–3226.

Rajaobelina, L., Brun, I., Kilani, N., & Ricard, L. (2022). Examining emotions linked to live chat services: The role of e-service quality and impact on word of mouth. *Journal of Financial Services Marketing*, *27*, 232–249.

Reichheld, F. F. (2003). The One Number You Need to Grow. *Harvard Business Review*, *81*, 46–55.

Renier, L. A., Schmid Mast, M., & Bekbergenova, A. (2021). To err is human, not algorithmic – Robust reactions to erring algorithms. *Computers in Human Behavior*, *124*, 106879.

Ribeiro, M. A., & Prayag, G. (2019). Perceived quality and service experience: mediating effects of positive and negative emotions. *Journal of Hospitality Marketing and Management*, *28*, 285–308.

Richins, M. L. (1983). Negative Word-of-Mouth by Dissatisfied Consumers: A Pilot Study. *Journal of Marketing*, *47*, 68–78.

Riquel, J., Brendel, A. B., Hildebrandt, F., Greve, M., & Dennis, A. R. (2021). "F*** You !" – An Investigation of Humanness, Frustration, and Aggression in Conversational Agent Communication. *Proceedings of the 42nd International Conference on Information Systems (ICIS)*, 1–16.

Riquel, J., Brendel, A. B., Hildebrandt, F., Greve, M., & Kolbe, L. M. (2021). "Even the Wisest Machine Makes Errors" – An Experimental Investigation of Human-like Designed and Flawed Conversational Agents. *Proceedings of the 42nd International Conference on Information Systems (ICIS)*, 1–16.

Sánchez-García, I., & Currás-Pérez, R. (2011). Effects of dissatisfaction in tourist services: The role of anger and regret. *Tourism Management*, *32*, 1397–1406.

Seeger, A.-M., & Heinzl, A. (2021). Chatbots often Fail! Can Anthropomorphic Design Mitigate Trust Loss in Conversational Agents for Customer Service? *Proceedings of the 29th European Conference on Information Systems (ECIS)*.

Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2021). Texting with Human-like Conversational Agents: Designing for Anthropomorphism. *Journal of the Association for Information Systems*, *22*(4), 1–58.

Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2018). Designing Anthropomorphic Conversational Agents: Development and Empirical Evaluation of a Design Framework. *Proceedings of the 39th International Conference on Information Systems (ICIS)*, 1–17.

Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2017). When Do We Need a Human? Anthropomorphic Design and Trustworthiness of Conversational Agents. *Special Interest Group on Human-Computer Interaction Proceedings*, *15*, 1–6.

Spatola, N., & Wudarczyk, O. A. (2021). Ascribing emotions to robots: Explicit and implicit attribution of emotions and perceived robot anthropomorphism. *Computers in Human Behavior*, *124*, 106934.

Spexard, T. P., Hanheide, M., Li, S., Wrede, B., & others. (2008). Oops , Something Is Wrong Error Detection and Recovery for Advanced Human-Robot-Interaction. *Proceedings of the ICRA Workshop on Social Interaction with Intelligent Indoor Robots*.

Stauss, B., Schmidt, M., & Schoeler, A. (2005). Customer frustration in loyalty programs. *International Journal of Service Industry Management*, *16*(3), 229–252.

Suler, J. (2004). The online disinhibition effect. In *Cyberpsychology and Behavior* (pp. 321–326).

Sundaram, D. S., Mitra, K., & Webster, C. (1998). Word-of-mouth communications: A motivational analysis. *Advances in Consumer Research, Vol. XXV*, *25*, 527–531.

Van Mechelen, I., & Hennes, K. (2009). The appraisal basis of anger occurrence and intensity revisited. *Cognition & Emotion*, *23*, 1373–1388.

van Pinxteren, M. M. E., Pluymaekers, M., & Lemmink, J. G. A. M. (2020). Human-like communication in conversational agents: a literature review and research agenda. *Journal of Service Management*, *31*(2), 203–225.

Verhagen, T., Nauta, A., & Felberg, F. (2013). Negative online word-of-mouth: Behavioral indicator or emotional release? *Computers in Human Behavior*, *29*(4), 1430–1440.

Verhagen, T., van Nes, J., Feldberg, F., & van Dolen, W. (2014). Virtual customer service agents: Using social presence and personalization to shape online service encounters. *Journal of Computer-Mediated Communication*, *19*(3), 529–545.

Wetzer, I. M., Zeelenberg, M., & Pieters, R. (2007). "Never eat in that restaurant, I did!": Exploring why people engage in negative word-of-mouth communication. *Psychology and Marketing*, *24*(8), 661–680.

Williams, M., & Buttle, F. (2014). Managing negative word-of-mouth: an exploratory study. *Journal of Marketing Management*, *30*, 1423–1447.

Wu, B., Xiao, Y., Zhou, L., Li, F., & Liu, M. (2023). Why Individuals with Psychopathy and Moral Disengagement Are More Likely to Engage in Online Trolling ? The Online Disinhibition Effect. *Journal of Psychopathology and Behavioral Assessment*, *Online fir*.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1993). The nature and determinants of customer expectations of service. *Journal of the Academy of Marketing Science*, *21*(1), 1–12.

Zemčík, T. (2021). Failure of chatbot Tay was evil, ugliness and uselessness in its nature or do we judge it through cognitive shortcuts and biases? *AI and Society*, *36*, 361–367.