The Chatbot Disclosure Dilemma: Desirable and Undesirable Effects of Disclosing the Non-Human Identity of Chatbots

Completed Research Paper

Nika Mozafari
University of Goettingen
Goettingen, Germany
nika.mozafari@wiwi.uni-goettingen.de

Welf H. Weiger
University of Goettingen
Goettingen, Germany
Alfaisal University
Riyadh, KSA
welf.weiger@wiwi.uni-goettingen.de

Maik Hammerschmidt
University of Goettingen
Goettingen, Germany
maik.hammerschmidt@wiwi.uni-goettingen.de

Abstract

Fueled by recent technological advancements, chatbots are more frequently used in the online customer service landscape. As chatbots are more and more capable to pose as humans, the question for firms arises whether they should disclose their chatbots' non-human identity or not. While identity disclosure seems to be the intuitive approach as it promotes transparency, previous research shows that disclosure comes at the cost of lower interaction efficiency, as many consumers today are still skeptical towards chatbots. This research adds to solving this chatbot disclosure dilemma by considering the mediating role of trust in the conversational partner and service-related context factors to understand the repercussions of chatbot disclosure for customer retention. Results of two scenario-based experimental studies show that depending on service context, chatbot disclosure does not only have negative consequences, but can lead to positive outcomes as well.

Keywords: Chatbots, chatbot identity disclosure, chatbot failure, trust, customer retention

Introduction

Chatbots are on the rise. Fueled by recent advancements in artificial intelligence and machine learning, organizations increasingly deploy chatbots in the service frontline. Chatbots are text-based virtual service robots that interact and communicate with users to deliver services (Wirtz et al. 2018). They build on natural language processing to emulate human-to-human communication, providing the potential to replace real-life or computer-mediated encounters with human service providers (Schuetzler et al. 2018). The market for chatbots is estimated to grow from 2.6 billion US $ in 2019 to 9.4 billion US $ in 2024, suggesting an annual growth rate of about 30 % (Markets and Markets 2019). As chatbots are not only capable of handling service requests 24/7 with highly scalable features, they also allow real-time and individualized interactions and therefore can mimic real-life human interactions (Go and Sundar 2019; Shevat 2017).

In particular, it is the additional socio-emotional and relational layer of user interactions with chatbots which contrasts it to interactions with traditional self-service technologies that merely have functional character (Wirtz et al. 2018). Not only may chatbots take on roles traditionally fulfilled by humans, they also allow for interactive conversations based on sophisticated speech recognition tools – thus emitting social cues that hint to a human conversational partner (Nass and Moon 2000; Wilson et al. 2017; Wünderlich and Paluch 2017). The combination of technological advancement and usage of social cues bears the challenge for consumers to correctly differentiate between algorithmic or human conversational partners (Candello et al. 2017). This development is highlighted by an empirical study on Google's chatbot Meena, in which its conversation quality was rated nearly as highly as the quality of real human conversations, leaving previously appraised chatbots such as Cleverbot or Mitsuku far behind (Adiwardana et al. 2020).[1]

Due to consumers not being able to identify their conversational partner when interacting via chats online, firms face the challenge to decide whether to provide chatbot identity information to their customers or not. For companies, both approaches seem to be viable options. While chatbot "Anna" of electric utility company E.ON explicitly introduces herself as a chatbot and thus focuses on transparency, telecommunications provider Vodafone does not reveal the non-human identity of their chatbot "Julia", aiming to appear as little artificial as possible to avoid customers feeling uncomfortable. Research indicates that the latter approach may be more efficient, as despite steadily improving chatbot performance, many consumers are still averse towards algorithms due to lack of trust in their performance (Dietvorst et al. 2015).

Obviously, firms face a chatbot disclosure dilemma in terms of having to trade off transparency (which calls for disclosure) and efficiency (which speaks against disclosure). For addressing this dilemma, a key question is whether disclosing the non-human identity of chatbots to consumers will yield solely negative consumer responses, or if there are situations in which the identity disclosure can also produce favorable business-relevant outcomes. More precisely, there is a need for firms to identify contextual factors under which the disclosure of the chatbot identity may lead to desirable or undesirable outcomes.

Prior research on the consequences of chatbot disclosure is in a nascent stage and points to two deficiencies. First, so far it has been limited to negative consequences of chatbot identity disclosure and, second, it has neglected to consider key aspects of chatbot–mediated firm–customer interactions. In tackling these gaps, this study contributes to research on designing chatbot systems in two important ways. First, we provide empirical insight on how reactions to chatbot disclosure vary for different types of service interactions by considering two service-related context factors: the consumer's service issue (i. e., whether the issue is routine or critical) and the outcome of the service interaction (i. e., whether the chatbot can resolve the issue or fails at doing so). We are the first to show that, for certain service types, positive outcomes of chatbot disclosure prevail and hence firms can achieve both transparency and efficiency. Second, we highlight the role of trust in the conversational partner as a mediator between chatbot identity disclosure and firm-beneficial consumer behavior. Thus, we show the relevance of psychological consumer responses in chatbot–mediated interactions beyond behavioral outcomes. Together, these insights guide firm's design of chatbot systems in terms of whether and under which circumstances to disclose chatbot identity.

We structure the rest of this article as follows: We begin by presenting our research framework to give an overview of our proposal. We then provide a literature review by discussing related research on chatbot identity disclosure and the roles of service-related contingencies and trust in the context of human–chatbot interactions. Further, we introduce attribution theory for grounding the variables contained in the framework and deriving our hypotheses regarding the links between them. Next, we outline two studies based on experiments simulating online interactions with a chatbot to identify whether chatbot disclosure will yield positive or negative effects on customer behavior in different settings. Our first study examines the impact of chatbot disclosure across different service issues (i. e., critical vs. routine). The second study investigates effects of chatbot disclosure for different service outcomes (i. e., chatbot failure vs. no chatbot

[1] We acknowledge that the results on the superiority of a Google chatbot published by Google itself have to be treated with caution.

failure). Finally, in our concluding comments, we summarize our findings, outline theoretical and practical implications and present limitations.

Conceptual Background

This section presents the research framework of our study. To embed this study into existing literature, we also discuss related work on chatbot disclosure. Further, we highlight the need for considering contextual factors and the mediating role of trust in human–chatbot interactions. To provide a theoretical base for hypotheses development, we present attribution theory and apply it to our research context.

Research Framework

Figure 1 illustrates our research framework. To evaluate whether and under what circumstances chatbot disclosure produces favorable outcomes for firms, we consider the effect of chatbot disclosure for different service-related context factors, i. e. service issue and service outcome, on retention behavior through trust. To represent retention behavior, we consider both desirable behavior, i. e. loyalty, and undesirable behavior, i. e. churn. We choose customer retention measures to capture the repercussions of chatbot disclosure as they are central to company profitability (McCollough et al. 2000). Further, we focus on retention instead of purchase behavior, as most chatbots today are deployed in post-purchase customer service settings (Shevat 2017).

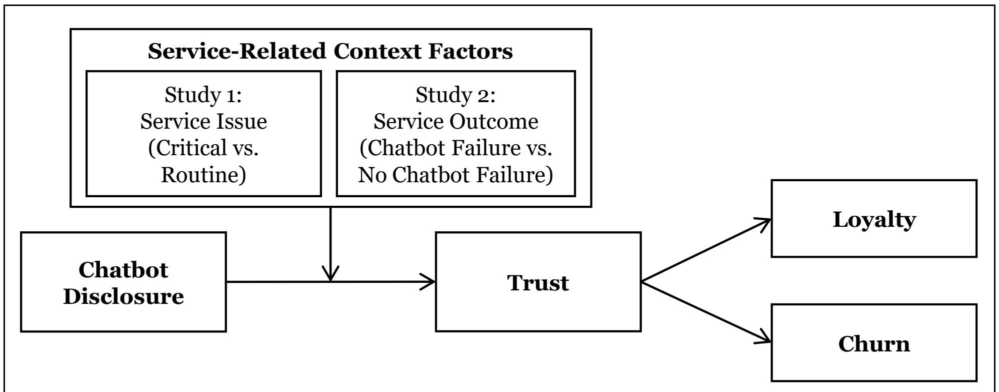

Figure 1. Research Model

Related Work on Chatbot Disclosure

As chatbot technology becomes increasingly sophisticated and chatbots are increasingly able to pose as humans, the more relevant it becomes for firms to understand the repercussions of disclosing or not disclosing chatbot identity (Skjuve et al. 2019), as Google's Meena proves again. The discussion was already sparked in 2018 by Google Duplex. The intelligent phone assistant employed a variety of social cues that were characteristic to human conversations, e. g. the incorporation of speech disfluencies (Leviathan and Matias 2018), creating an uncannily realistic experience. To prevent perceived eeriness, scholars argue that bots should be open about their algorithmic, non-human identity (Mone 2016). This should not only be done from a transparency and ethics point of view, but further to prevent misalignment of consumer expectations and chatbot performance (Luger and Sellen 2016).

However, existing empirical research on the effect of chatbot disclosure reveals a chatbot disclosure dilemma as it has thus far found largely negative reactions to disclosed (vs. undisclosed) chatbots, despite identical performance, suggesting that transparency about identity comes at a high cost. For instance, IS research shows that disclosing the non-human identity of chatbots negatively impacts efficiency of human-machine cooperation (Ishowo-Oloko et al. 2019) and perceived social presence and humanness (Hendriks et al. 2020). Further, studies from related research on human-computer interactions find negative effects of chatbot disclosure on user acceptance (Murgia et al. 2016) and persuasion efficiency (Shi et al. 2020). Notably, one study argues adversatively in saying that undisclosed bots will negatively affect user experience due to feelings of uncertainty, however finds no significant evidence that a chatbot believed to be human is perceived as more pleasant than a chatbot whose identity is disclosed (Skjuve et al. 2019). Finally, the effect of chatbot disclosure has also been highlighted from a marketing perspective, where Luo et al. (2019) show that chatbot disclosure negatively impacts duration of service interactions and purchases. The consistent absence of positive effects of chatbot disclosure in any field of research is startling: Not only has research observed negative biases towards disclosed bots, although performance levels in service delivery were held constant across disclosed and undisclosed bots, but also did some studies provide evidence on superior performance of bots over humans.

These somewhat contradictory implications of prior research on the meaningfulness of disclosing chatbot identity call for considering contextual factors that take into account different service settings and varying performance levels of chatbots. Considering such factors, highlighted in the next section, helps to inform companies when only negative or also positive effects of chatbot disclosure can be expected.

The Relevance of Considering Different Service-Related Contexts

As all the studies discussed above have only examined main effects of disclosure implying universal consequences of disclosure across service situations, we suggest that these insights can be better understood by testing whether the effects of chatbot disclosure vary across different service-related contexts.

First, consumers may have different types of service issues when choosing to interact with a firm via an online chat, ranging from routine FAQ-style questions to more complex, critical issues. Existing studies fail to address that consumers are likely to react differently to chatbot disclosure for different service issues, as research indicates consumers find automation more or less desirable for different types of services (Leung et al. 2018). Traditional service literature advises to consider the moderating role of the criticality of a service issue to evaluate the efficiency of service communication (Webster and Sundaram 2009). Criticality can be defined as how important the service issue is perceived by the customer (Webster and Sundaram 1998). To gain a better understanding of the effect of chatbot disclosure on firm-beneficial outcomes and to enrich insights from existing studies, we therefore consider the type of service issue (i. e., critical or routine issue) as a moderating factor.

Second, in all studies mentioned above, bot performance is at a high level, so that reactions to chatbot disclosure in failure settings remain yet to be investigated. However, research shows that consumers react differently to robot errors than to human errors (Robinette et al. 2017). Particularly, as chatbot design influences error tolerance and trust resilience (De Visser et al. 2016), consumers should react differently to errors from disclosed vs. undisclosed chatbots. Hence, to examine the effect of chatbot disclosure in failure settings, we further include the outcome of the service interaction (i. e., chatbot failure or no chatbot failure) as a moderating variable.

The Relevance of Trust for Human–Chatbot Interactions

Finally, some of the existing studies on bot identity disclosure attempt to provide explanations for negative biases. A common argument is the lack of trust in algorithms, that is suggested in all, but tested in none of the studies (Murgia et al. 2016; Ishowo-Oloko et al. 2019; Luo et al. 2019; Skjuve et al. 2019; Hendriks et al. 2020; Shi et al. 2020). Below, we highlight arguments for considering the role of trust for human–chatbot interactions as an explanatory mechanism for behavioral responses to chatbot disclosure.

To explain the emergence of behavior as a reaction to a stimulus, the mediating psychological response has to be considered (Fishbein and Ajzen 1975). In the context of our study, we identify trust as the relevant variable that explains behavioral responses to chatbot disclosure, as trust is central for environments that produce high levels of uncertainty (Riedl et al. 2011). We define trust as the willingness to rely on an exchange partner; more specifically the willingness to rely on the trustee to be able to fulfill their obligations, to act in the trustor's interest and to tell the truth (Komiak and Benbasat 2004; Moorman et al. 1993).

According to commitment-trust theory, trust in an exchange partner is a key mediating variable between service attributes and subsequent business-relevant consumer behavior (Hart and Johnson 1999; Morgan and Hunt 1994). If trust is established in a relationship, the trustor will commit themselves to that relationship (Hrebiniak 1974). Importantly, the key role of trust is also enforced in agent-mediated interactions (Komiak and Benbasat 2004) where agents are not human. Consumers generalize social concepts like trust to computers, even if they know they are not interacting with a living being (Nass and Moon 2000). Neurological research confirms that trust building processes within human-computer interactions can in fact be compared to that of human-human interactions (Riedl et al. 2011). On a broader level, trust has been long established as a focal variable for the evaluation of human-technology interactions (Hancock et al. 2011; Schaefer et al. 2016).

In a chatbot context, a variety of studies has focused on the examination of trust as a reaction to chatbot design (e. g., Cowell and Stanney 2005; De Visser et al. 2016; Nunamaker et al. 2011; Sameh et al. 2010). However, in the context of chatbot disclosure, the difference in trust between disclosed and undisclosed chatbots has not been assessed yet. Further, we aim to create a comprehensive framework by not only including trust, but also subsequent desirable and undesirable behavioral outcomes.

Attribution Theory

To explain how and why chatbot disclosure affects consumer trust and thus retention differently depending on service issue and service outcome, we draw upon attribution theory.

People are inherently driven to assign causes to other's behavior and events in order to better understand their environment. Attribution theory investigates this formation of causal judgement. Specifically, attributions are made based on situational factors, such as external circumstances, or dispositional factors, such as beliefs about characteristics like ability or motivation of others (Heider 1958). These causal attributions subsequently will affect psychological and behavioral responses (van Vaerenbergh et al. 2014), like trust or retention.

A core tenet of attribution theory suggests that the process of inferring a cause for behavior or events is prone to the attribution bias (Forsyth 1987). The attribution bias describes the tendency of humans to overly rely on dispositions relative to situational influences, i. e. hastily forming judgements based on personal beliefs, overlooking the actual situational behavior of an exchange partner (Ross 1977). That means, when reacting to an event, people tend to ascribe the outcome of a situation to the perceived internal characteristics of involved parties instead of the actual situational environment. This cognitive bias occurs as a result of a spontaneous, premature attribution. However, an attribution becomes less spontaneous and more elaborated if the valence of an outcome is negative (Kanazawa 1992). That is, if a negative outcome occurs, individuals feel the need to comprehend, control and predict their environment in order to effectively cope with the situation (Weiner 2000). In this case, when searching for a cause, individuals invest higher effort to more deeply understand why a negative outcome happened (van Vaerenbergh et al. 2014; Weiner 1985).

Hypotheses Development

Embedding the context of our study into attribution theory, the service-related context factors (i. e., service issue and service outcome) represent situational attributes, while customers' beliefs about chatbots represent their disposition. If service delivery proceeds in a normal course of action (i. e., without failure), when presented with the information of the non-human identity of the conversational partner, the attribution is likely to be made spontaneous and less elaborated and hence likely to be biased by customers' beliefs about the chatbot characteristics.

As stated above, humans show skepticism towards algorithms (Dawes 1979), tend to have less confidence in their performance (Dietvorst et al. 2015) and perceive chatbots as less knowledgeable and empathetic, especially with regard to complex and difficult tasks (Luo et al. 2019). Following this line of reasoning, consumers have the belief that chatbots are not capable of handling critical issues. This should further be enforced by the fact that retrospectively, chatbots are often used for handling simple repetitive routine tasks only (Huang and Rust 2018). Based on the attribution bias, we argue that for services delivered in response to critical service issues consumers are relying on their negative disposition towards chatbots when learning about the chatbot identity of the conversational partner and hence form reduced trust. Therefore:

> *H1: Disclosing (vs. not disclosing) chatbot identity reduces trust in the conversational partner more for critical than for routine service issues.*

Under the notion that trust positively affects loyalty and negatively affects churn, if chatbot disclosure in a critical service request setting reduces trust, loyalty (churn) will indirectly be affected negatively (positively). In this case, trust takes in a mediating role between the service interaction and the behavioral outcomes (Morgan and Hunt 1994). Hence:

H2a: The negative indirect effect of the interaction between chatbot disclosure and critical service issue on loyalty is mediated by trust in the conversational partner.

H2b: The positive indirect effect of the interaction between chatbot disclosure and critical service issue on churn is mediated by trust in the conversational partner.

In a failure situation, however, customers seek to better understand the outcome of an event. In search for a cause of the negative outcome, chatbot disclosure represents a concrete cue that stimulates attributional activity and allows a better understanding of the reasons of the failure (Weiner 1985). If chatbot identity is not disclosed, information on the cause of the outcome remains abstract and the customer is not able to identify a specific entity the failure can be attributed to. As a result of chatbot disclosure, customers should be able to better cope with the situation. Therefore, locating the cause for failure should mitigate the loss in trust that is induced by the chatbot failure.

H3: Disclosing (vs. not disclosing) chatbot identity enhances trust in the conversational partner if the service outcome is a failure.

Under the notion that trust positively affects loyalty and negatively affects churn, if chatbot disclosure (vs. no disclosure) in a failure setting enhances trust, loyalty (churn) will indirectly be affected positively (negatively). Again, trust takes in a mediating role between the service interaction and the behavioral outcomes (Morgan and Hunt 1994). Hence:

H4a: The positive indirect effect of the interaction between chatbot disclosure and service failure on loyalty is mediated by trust in the conversational partner.

H4b: The negative indirect effect of the interaction between chatbot disclosure and service failure on churn is mediated by trust in the conversational partner.

Studies

To test our research model, two main studies have been conducted (complemented by two prestudies), one for each service-related context factor. In study 1, we focus on the effect of chatbot disclosure for different service issues, while holding service outcome constant. In study two, we examine the effect of chatbot disclosure for different service outcomes, while holding service issue constant. Table 1 offers an overview with the goals of each study including descriptive statistics of the samples.

Study	**Purpose of Study**	**Dependent Variables**	**Sample**
Prestudy 1	Pretest of scenarios to rule out that participants had predisposition on the identity of their conversational partner	Perceived Humanness	N = 18
Prestudy 2		Anticipated Identity	N = 26
Study 1	Examination of the effect of chatbot disclosure for different service issues (critical vs. routine)	Trust, Loyalty, Churn	N = 252 68% female M_{age} = 27.5 years
Study 2	Examination of the effect of chatbot disclosure for different service outcomes (chatbot failure vs. no chatbot failure)		N = 270 71% female M_{age} = 27 years
Notes: The prestudies are highlighted in the description of design and sample of study 1.			

Table 1. Overview of studies

Study 1: Chatbot Disclosure for Different Service Issues

In our first main study, we investigate the effect of chatbot disclosure on trust, loyalty and churn under the consideration of different service issues. In the following sections, we present the design, sample, method, results and discussion for study 1.

Design and Sample

The goal of study 1 was to examine how disclosing the non-human chatbot identity impacts consumer trust, loyalty and churn for different service issues (but holding service outcome constant in terms of considering successful service delivery). To examine this, we conducted a 2 (chatbot disclosure vs. no disclosure) × 2 (critical vs. routine service issue) between-subject experiment.

To implement realistic manipulations, real life online chats were evaluated prior to designing the experiment. For our studies, we chose to use scenario-based experiments, to be able to control for confounding influences and ensure high internal validity. This enabled us to easily create a human–chatbot interaction, from which participants could not infer the identity of the conversational partner without disclosure. As the disclosure of the non-human chatbot identity is the central manipulation of our study, two prestudies were conducted to test perceived humanness and whether participants had a predisposition on the identity of their conversational partner. In a first prestudy, we found a significant negative effect of chatbot disclosure on perceived humanness of the conversational partner (N = 18; $M_{\text{disclosed}}$ = 4.96, SD = 1.71; $M_{\text{undisclosed}}$ = 6.07, SD = 0.89; t = 1,73, $p < 0.1$), suggesting when the non-human identity was not revealed, the conversational partner was perceived as more human. Thus, the construction of a scenario where the chatbot is not immediately recognized as such if identity is not disclosed was successful. To further strengthen these results, in a second prestudy we asked participants for their specific guess on the anticipated identity of the conversational partner, if it was not disclosed. The results show that while only 3.8 % of 26 participants guessed they were talking to a bot, 38.5 % of participants were certain the conversational partner was human. The rest of the sample (57.7%) could make no distinct assertion on the identity of the conversational partner in the online chat. Both prestudies were based on the routine service issue scenario described below.

For the main study, we recruited respondents from two sources: First, we collected data through a European online panel provider (i. e., Clickworker) with monetary compensation. Second, using distribution lists and social media we distributed the online survey in the context of a European university and rewarded participation with a raffle of online shopping vouchers. The samples were pooled and included as a control variable in all analyses. The data collection mode yields no significant effect on the dependent variables in any of the analyses ($p > 0.1$).

Participants were instructed to imagine that they were customers of an energy provider and were about to contact the energy provider via their online chat. In the online chat, to initially conceal the identity of the chatbot, the chatbot did not present itself as a bot, but simply introduced himself as “Leon”. Participants were randomly assigned to one of the two service issues, i. e. a routine service issue or a critical service issue. In the routine scenario, participants were instructed to imagine having to enter their meter readings in the online chat. In the critical scenario, participants were instructed to imagine contacting the company because of a wrongfully conducted double debiting of their connected bank account. We chose these service issues for the routine and critical manipulations respectively, as the former merely represents a minor inconvenience, the latter involves monetary costs for the customer. Pieces of the conversation were presented to the participants in sequence. During the conversation, the customer’s issue was resolved in the chat, in that the meter reading was successfully recoded and the wrongfully withdrawn amount was reimbursed. At the end of the conversation, half of the participants were informed that the service agent of the presented chat dialogue was in fact not a human person. Instead, it was revealed to them that the customer had been interacting with a chatbot. This information was delivered as descriptive text in the chat window. The other half of participants did not receive this information, but instead read the descriptive text that they may now close the chat window. Apart from disclosure and service issue manipulations, the course of the chatbot interactions was of identical length and depth. For exemplary screenshots from the disclosure × routine service issue scenario see Figure 2.

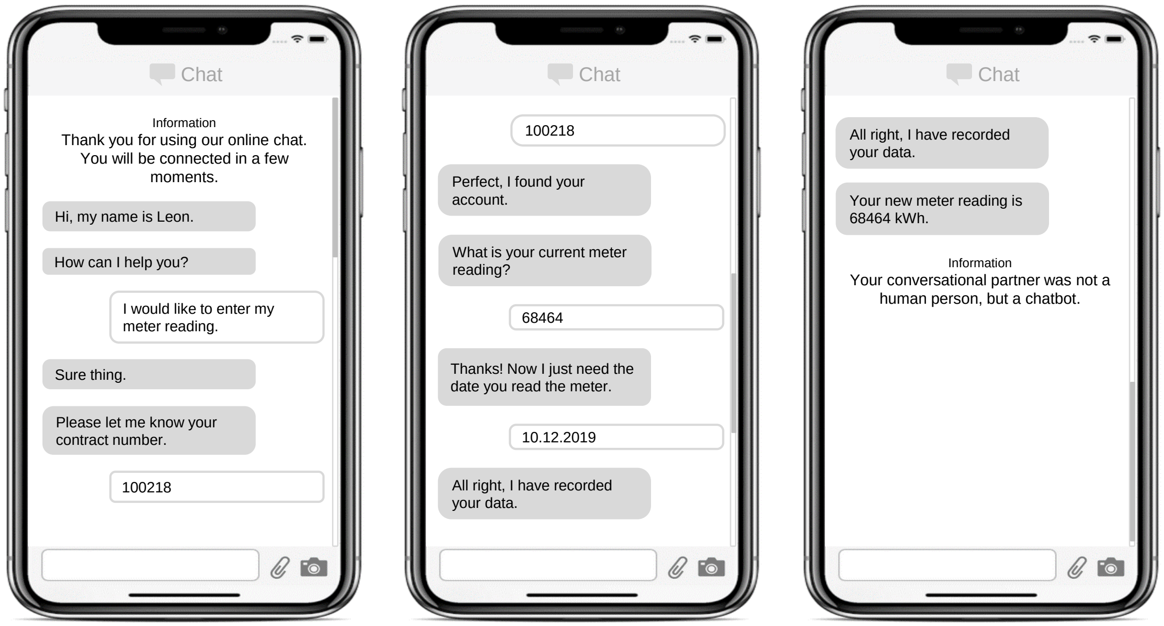

Figure 2. Exemplary Scenario

After going through these scenarios of service encounters, participants reported their trust in the conversational partner. Further, to assess customer retention, measures for loyalty and churn were taken on a firm level (see Table 2 for items). As control variables, in addition to controlling for data collection mode, we further gathered measures on socio-demographics, resistance to information systems (Kim and Kankanhalli 2009) and need for interaction (Dabholkar and Bagozzi 2002). The study was closed with manipulation checks.

Multi-item constructs were measured by taking the mean of participants' statements on 7-point-likert scales, anchored by 1 = strongly disagree and 7 = strongly agree. Finally, after merging the two subsamples, the sample consisted of 338 participants. Those who did not fill out the entire survey, those who failed to answer attention checks correctly and those who did not identify correctly whether the chatbot identity was revealed were discarded from further analyses. The effective sample thus consisted of 252 participants (68% female, M_{age} = 27.5 years). The manipulation check for perceived criticality of service issue is significant at $p < 0.0001$, with respondents in the critical service issue scenario perceiving the service to be significantly more critical than in the routine service issue scenario.

We examined construct reliability and validity of our focal constructs by employing different methods. First, all Cronbach's alpha and composite reliability measures are above the cut-off value of .7, indicating construct-level reliability (see Table 2) (Hulland et al. 2018). Second, we rely on the Fornell and Larcker (1981) approach to obtain convergent validity as the average variance extracted (AVE) for each multiple-item construct exceeds .50. Furthermore, the AVE for each multi-item construct is larger than the shared variance with any possible pairings of the remaining constructs, suggesting initial evidence for discriminant validity (Hulland et al. 2018). Third, as suggested by prior research, we additionally rely on the heterotrait-monotrait (HTMT) method to further demonstrate discriminant validity (Henseler et al. 2015; Krämer et al. 2020). Estimating the HTMT ratios for all multi-item constructs yields values that range from .22 to .72 which are well below the conservative cut-off value of .85. The highest upper limit of the 97.5% bias-corrected confidence intervals for all multi-item constructs is .79, which strengthens our confidence in the discriminant validity exhibited by the focal constructs.

Construct	Measurement	Study 1				Study 2			
		Item loadings	α	AVE	CR	Item loadings	α	AVE	CR
Trust in conversational partner (Bhattacherjee 2002)	The conversational partner has the necessary skills to deliver the service.	$.71^{***}$	.88	.59	.91	$.93^{***}$	.96	.82	.97
	The conversational partner has access to the information needed to handle my service request adequately.	$.65^{***}$				$.89^{***}$			
	The conversational partner is fair in its conduct of my service request.	$.82^{***}$				$.92^{***}$			
	The conversational partner has high integrity.	$.72^{***}$				$.94^{***}$			
	The conversational partner is receptive to my service request.	$.67^{***}$				$.92^{***}$			
	The conversational partner makes efforts to address my service request.	$.78^{***}$				$.90^{***}$			
	Overall, the conversational partner is trustworthy.	$.69^{***}$				$.82^{***}$			
Loyalty (firm level) (Wallenburg 2009)	I would continue being a customer of the energy provider.	$.87^{***}$	.87	.92	.79	$.92^{***}$	.94	.89	.96
	I would extend my existing contract with the energy provider when it expires.	$.90^{***}$				$.96^{***}$			
	If I had to decide, I would again select this energy provider.	$.90^{***}$				$.95^{***}$			
Churn (firm level) (Bhattacherjee et al. 2012)	I would terminate my existing contract with the energy provider.	$.94^{***}$	.92	.87	.95	$.97^{***}$	.96	.93	.98
	I would intend to switch my energy provider.	$.95^{***}$				$.97^{***}$			
	I would plan to abandon the energy provider.	$.90^{***}$				$.96^{***}$			

Table 2. Measures of Focal Constructs, Indicator and Construct Reliability

Method and Results

To first test the effect of the manipulations on trust, we used analysis of covariance (ANCOVA). Chatbot disclosure, service issue and the interaction of disclosure and service issue were used as independent variables, age, academic education, resistance to information systems, need for interaction, data collection mode as covariates and trust in the conversational partner as the dependent variable. For an overview of the results see Table 3.

Source	**Partial *SS***	***df***	***MS***	***F***	***p***
Model	18.46	8	2.31	3.97	0.0002
Chatbot Disclosure (vs. No Disclosure)	1.85	1	1.85	3.19	0.0755
Service Issue (Critical vs. Routine Request)	0.92	1	0.92	1.58	0.2093
Chatbot Disclosure × Service Issue	1.73	1	1.73	2.97	0.086
Age	2.04	1	2.04	3.51	0.0623
Academic Education	1.48	1	1.48	2.55	0.1117
Resistance to Information Systems	8.13	1	8.13	13.99	0.0002
Need for Interaction	2.46	1	2.46	4.23	0.0407
Data Collection Mode	0.08	1	0.08	0.13	0.7144
Residual	141.30	243	0.58		
Total	159.75	251	0.64		
Adjusted R^2	0.09				
Notes: N = 252; SS: Sum of Squares, df: Degrees of Freedom, MS: Mean Square					

Table 3. Study 1: Analysis of Covariance (DV: Trust)

The results show a significant negative main effect of chatbot disclosure on trust ($M_{\text{disclosed}} = 6.08$, $SE = 0.07$; $M_{\text{undisclosed}} = 6.26$, $SE = 0.08$; $F = 3.19$, $p < 0.1$). The main effect of service issue is not significant ($M_{\text{critical}} = 6.23$, $SE = 0.07$; $M_{\text{routine}} = 6.11$, $SE = 0.08$; $F = 1.58$, $p > 0.1$). The interaction effect of chatbot disclosure and service issue is significant ($F = 2.97$, $p < 0.1$). We found a negative effect of chatbot disclosure if the service issue is critical ($\Delta_{\text{Trust}} = -0.34$, $SE = 0.14$, $t = -2.49$, $p < 0.05$). There was no effect of chatbot disclosure on trust if service issue is routine ($\Delta_{\text{Trust}} = -0.003$, $SE = 0.14$, $t = -0.02$, $p > 0.1$). Taken together, these two results provide support for H1, which stated that disclosing chatbot identity reduces trust more for critical than for routine service issues. Figure 3 illustrates the interaction effect of chatbot disclosure and service issue on trust.

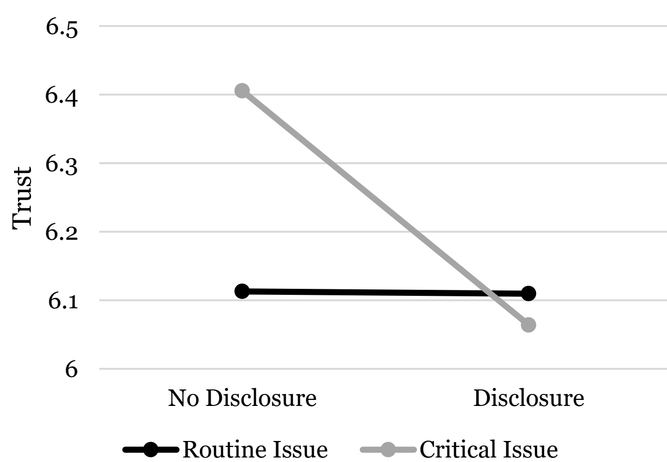

Figure 3. Study 1: Interaction of Chatbot Disclosure × Service Issue on Trust

To test hypotheses 2a and 2b, we conducted a mediation analysis using the products of coefficient method to estimate the indirect effects and bias-corrected bootstrapped confidence intervals (Zhao et al. 2010). Results are shown in Table 4. In line with our expectations, results show that the interaction of chatbot disclosure and service issue has a significant negative indirect effect on loyalty (Chatbot Disclosure × Critical Service Issue → Trust → Loyalty = –0.1944, lower-level confidence interval [LLCI] = –0.4075; upper-level confidence interval [ULCI] = –0.0212) through trust because the 90% confidence intervals do not include zero, supporting H2a. Further, the interaction of chatbot disclosure and service issue has a significant positive indirect effect on churn (Chatbot Disclosure × Critical Service Issue → Trust → Churn = 0.1673, LLCI = 0.0179, ULCI = 0.3525), thus also supporting H2b. We found no significant direct effects of chatbot disclosure or service issue on loyalty or churn, suggesting full (or indirect-only) mediation for both paths (Zhao et al. 2010).

	Coeff.	SE	LLCI	ULCI	
Chatbot Disclosure → Trust → Loyalty	–0.0019	0.0823	–0.1380	0.1326	n. s.
Chatbot Disclosure → Trust → Churn	0.0016	0.0710	–0.1136	0.1192	n. s.
Chatbot Disclosure × Critical Service Issue → Trust → Loyalty	–0.1944	0.1183	–0.4075	–0.0212	H2a ✓
Chatbot Disclosure × Critical Service Issue → Trust → Churn	0.1673	0.1031	0.0179	0.3525	H2b ✓

Notes: N = 252; number of bootstrap samples = 5000; Coeff. = coefficient; SE = standard error; LLCI = 90 % lower level confidence interval; ULCI = 90 % upper level confidence interval.

Table 4. Study 1: Mediation Testing

Discussion

The results of study 1 show that despite identical course of conversation, the disclosure of the non-human chatbot identity leads to lower trust, and thus lower customer retention. As expected, consumers do not trust chatbots to handle critical service issues. This happens despite the fact that the customer's issue was resolved. Interestingly, there is no change in trust for chatbots with or without disclosure in the routine service issue scenario. This implies that consumers may trust chatbots to handle simple, routine tasks, while they may not feel secure confiding in an algorithmic conversational partner for a more complex, critical issue. Notably however, if the algorithmic identity was not revealed, trust levels were significantly higher for the critical service issue than for the routine service issue. It can be assumed that this is the result of a higher emotional involvement with the issue.

Study 2: Chatbot Disclosure for Different Service Outcomes

We assume that the trust-eroding effect of disclosed bots for critical service requests may not prevail for situations when customer's service inquiries cannot be resolved in the online chat due to a more elaborated attribution process triggered in failure situations. If a negative service outcome comes into play, we expect disclosing chatbot identity to enhance trust compared to not disclosing as the disclosure allows to better understand the reason for the negative outcome and to cope with the situation. Therefore, in our second study, we examine the effect of chatbot disclosure on trust, loyalty and churn under the consideration of different service outcomes to test hypotheses H3, H4a and H4b. In the following sections, we present the design, sample, method, results and discussion for study 2.

Design and Sample

The goal of study 2 was to examine the effect of disclosing the non-human chatbot identity on consumer trust, loyalty and churn for different service outcomes. Therefore, another 2 (chatbot disclosure vs. no disclosure) × 2 (chatbot failure vs. no chatbot failure) between-subject experiment was conducted. For chatbot design, we relied on the same materials as in study 1. Thus, there was no need for a further prestudy. Again, respondents were acquired by the two data collection modes mentioned above.

As study 1 offered compelling results for the critical service issue scenario, we selected it as a basis for study 2. In addition, it is more realistic that a chatbot failure could happen for such more complex services. The critical service issue scenario mimicked that of study 1: Participants had to imagine that they were customers of an energy provider and about to use the online chat to contact the energy provider on the issue of the wrongfully conducted double debiting. Pieces of the conversation were presented to the participants in sequence. Participants were randomly assigned to one of the service outcome conditions, either the chatbot failure or the no chatbot failure condition. In the chatbot failure condition, the conversational partner was not able to handle the customer's inquiry and thus could not resolve the customer problem, whereas in the other condition the customer's request was handled successfully. Identically to study 1, at the end of the conversation, half of the participants received the information that the conversational partner was a chatbot.

After this, we collected measures on trust, loyalty and churn, control variables, and manipulation checks. All items and scales used were identical to study 1, except for the manipulation check used for service outcome. The initial sample consisted of 336 participants. Again, those who did not fill out the entire survey, those who failed to attention checks and those who did not identify correctly whether the chatbot identity was revealed were discarded from further analyses. The effective sample consisted of 270 participants (71% female, M_{age} = 27 years). All participants passed the manipulation check for service outcome, i. e. they correctly stated whether the conversational partner was able to resolve their service issue.

Method and Results

We conducted an ANCOVA to test the effect of chatbot disclosure on trust for different service outcomes. Chatbot disclosure, service outcome and the interaction of disclosure and service outcome were used as independent variables, age, academic education, resistance to information systems, need for interaction, date collection mode as covariates and trust in the conversational partner as the dependent variable. For an overview of the results see Table 5.

Source	**Partial *SS***	***df***	***MS***	***F***	***p***
Model	968.25	8	121.03	139.19	0.0000
Chatbot Disclosure (vs. No Disclosure)	0.46	1	0.46	0.53	0.4659
Service Outcome (Chatbot Failure vs. No Chatbot Failure)	821.45	1	821.45	944.69	0.0000
Chatbot Disclosure × Service Outcome	9.06	1	9.06	10.42	0.0014
Age	0.25	1	0.25	0.29	0.5901
Academic Education	0.01	1	0.01	0.01	0.9044
Resistance to Information Systems	1.20	1	1.20	1.38	0.2414
Need for Interaction	1.83	1	1.83	2.11	0.1478
Data Collection Mode	0.57	1	0.57	0.66	0.4184
Residual	226.95	261	0.87		
Total	1195.20	269	4.44		
Adjusted R^2	0.8043				
Notes: N = 270; SS: Sum of Squares, df: Degrees of Freedom, MS: Mean Square					

Table 5. Study 2: Analysis of Covariance (DV: Trust)

We found no significant main effect of chatbot disclosure on trust ($M_{disclosed}$ = 4.42, *SE* = 0.09; $M_{undisclosed}$ = 4.33, *SE* = 0.09; $F = 0.53$, $p > 0.1$). Not surprisingly, the main effect of service outcome on trust is negative ($M_{failure}$ = 2.51, *SE* = 0.09; $M_{nofailure}$ = 6.23, *SE* = 0.09; $F = 944.68$, $p < 0.001$). The interaction of chatbot disclosure and service outcome yields a significant effect on trust ($F = 10.42$, $p < 0.01$). Mirroring the results of study 1, the effect of chatbot disclosure was negative when no chatbot failure occurred ($\Delta_{Trust} = -0.29$, *SE* = 0.17, $t = -1.71$, $p < 0.1$). However and interestingly, the effect of chatbot disclosure on trust is positive in case of a chatbot failure (Δ_{Trust} = 0.46, SE = 0.16, t = 2.86, p < 0.01), supporting H3, which stated that disclosing chatbot identity enhances trust if the service outcome is a failure. For an illustration of the interaction, see Figure 4.

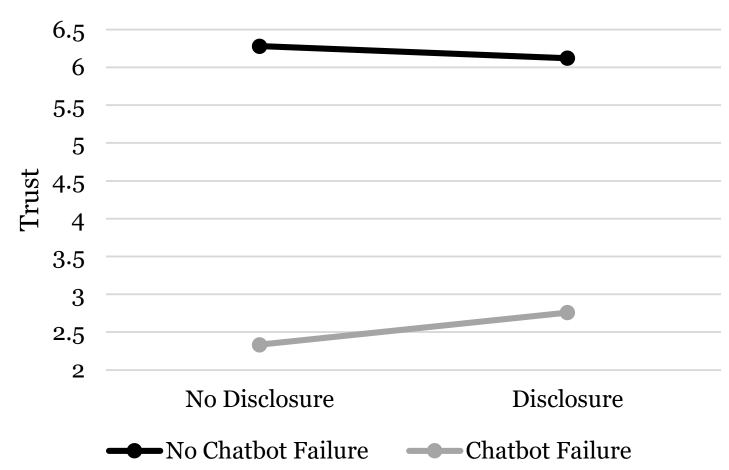

Figure 4. Study 2: Interaction of Chatbot Disclosure × Service Outcome on Trust

To further test hypotheses 4a and 4b, mediation analysis was conducted. Results are shown in Table 6. As expected, results show a significant positive indirect effect of the interaction of chatbot disclosure and service outcome on loyalty (Chatbot Disclosure × Chatbot Failure → Trust → Loyalty = 0.3185, LLCI = 0.1695; ULCI = 0.5258), supporting H4a. Further, the interaction of chatbot disclosure and service outcome has a significant negative indirect effect on churn (Chatbot Disclosure × Chatbot Failure → Trust → Churn = −0.4419, LLCI = −0.7161, ULCI = −0.2406), supporting H4b. We found no significant direct effects of chatbot disclosure or service outcome on loyalty or churn, suggesting full (or indirect-only) mediation for both paths (Zhao et al. 2010).

	Coeff.	SE	LLCI	ULCI	
Chatbot Disclosure → Trust → Loyalty	−0.1228	0.0647	−0.2433	−0.0326	sign.
Chatbot Disclosure → Trust → Churn	0.1703	0.0892	0.0447	0.3339	sign.
Chatbot Disclosure × Chatbot Failure → Trust → Loyalty	0.3185	0.1063	0.1695	0.5258	H4a ✓
Chatbot Disclosure × Chatbot Failure → Trust → Churn	−0.4419	0.1412	−0.7161	−0.2406	H4b ✓

Notes: N = 270; number of bootstrap samples = 5000; Coeff. = coefficient; SE = standard error; LLCI = 90 % lower level confidence interval; ULCI = upper level confidence interval.

Table 6. Study 2: Mediation Testing

Discussion

As expected, the main driver of trust in the conversational partner was whether or not the customer's issue could be solved (De Matos et al. 2007; Kelley et al. 1993). This result is well established in service research and should therefore not be the focus of our discussion. However, chatbot disclosure still plays a significant role: the interaction effect shows that a significant increase in trust can be observed if chatbot identity is disclosed in a chatbot failure setting. This suggests, that when the customer's issue cannot be resolved in the online chat, chatbot disclosure helps mitigate the negative failure effect. We suppose that this happens, as the disclosure offers a type of explanation for the negative outcome. While trust in the failure setting is significantly lower than in the no failure setting, we are able to demonstrate a positive reaction to chatbot disclosure, compared to no disclosure.

General Discussion

The goal of this article was to provide empirical insight on the reaction to chatbot disclosure, specifically the implications for customer retention through trust. Our results contribute to the current literature on human–chatbot interaction. First, we offer insights on how customers respond to chatbot disclosure psychologically and behaviorally. More precisely, we show that trust mediates the relationship between chatbot disclosure and customer retention. Also, we demonstrate that reactions to chatbot disclosure differ depending on the service context. In a similar vein as prior studies, we find that chatbot disclosure will negatively impact psychological and behavioral outcomes if the conversation proceeds flawlessly, i. e. if the chatbot delivers the expected service. More specifically, if the chatbot is able to solve the customer's issue, it will either negatively impact customer trust and thus hampers retention for critical service issues or not impact trust at all for routine service issues, suggesting the existence of an attribution bias in a chatbot context and that transparency may come at cost of efficiency. However, our findings reveal a "disclosure paradox" in terms that disclosing a chatbot's non-human identity enhances trust and thus retention (instead of mitigating it) in cases where the chatbot fails to deliver the expected service, as the chatbot represents an object to which the outcome can be attributed to. In other words, "merely" disclosing chatbot identity serves as an effective, yet very inexpensive and easy-to-implement means for failure recovery. Thus, there do seem to be cases in which chatbot disclosure does have positive effects on business-relevant consumer behavior. While we acknowledge that this enhancement of trust may seem small compared to the loss of trust induced by the chatbot failure, it still represents a significant effect and thus disclosure is a viable lever for damage control in case of chatbot failure. Of course, service providers should continue striving for error-free service delivery with chatbots. Finally, the results of our studies confirm the theoretical mechanisms of attribution theory in the context of chatbot disclosure (Davison and Martinsons 2016). That is, the results demonstrate a spontaneous and biased attribution for critical service issues despite being addressed successfully as well as an elaborated attribution following a negative service outcome.

Outlook

This study was conducted to test whether and under what circumstances chatbot identity should be disclosed. In case of adequate chatbot performance, disclosure will result in negative consequences, confirming a general aversive attitude towards bots, disregarding actual performance. This implies that companies should prevent disclosing the algorithmic identity to their customers in online chats. However, it is questionable if withholding chatbot identity is tenable ethically and legally in the long term. The state of California has already passed a bill to prevent companies from doing so for political and commercial bots (California Legislative Information 2018). If this development gains traction worldwide and disclosure becomes legally inevitable, based on the largely negative effects shown in prior studies, firms would have to scrutinize whether they deploy chatbots at all. While we find negative reactions to chatbot disclosure too, our results also prove that disclosure can in fact produce positive reactions. The way forward should thus not be to question deployment of chatbots, but to develop a disclosure strategy that consistently produces positive outcomes, is ethically tenable and hence eliminates the chatbot disclosure dilemma.

This study has so far only focused on disclosure of chatbot identity at the end of the conversation. For future research, we plan to examine different timing strategies for disclosure. In addition, we aim to examine different framing strategies of disclosure. The goal is to create a disclosure strategy that minimizes negative reactions and may even produce positive reactions beyond those that we found in our results. Furthermore, it is reasonable to assume the effect identified in this study is strongest for the initial customer–chatbot interaction. Eventually, the initial service encounter is of high relevance and will shape following encounters. Further analyses are necessary to examine how repeated chatbot interactions turn out. Since chatbot technology is developing very quickly, customers can often not apply prior knowledge about certain chatbots when evaluating every new interaction and so learning effects across interactions which could impede the effectiveness of disclosure strategies are likely to be low.

References

Adiwardana, D., Luong, M.-T., So, D. R., Hall, J., Fiedel, N., Thoppilan, R., Yang, Z., Kulshreshtha, A., Nemade, G., Lu, Y., and Le V, Q. 2020. "Towards a Human-like Open-Domain Chatbot," Google Research.

Bhattacherjee, A. 2002. "Individual Trust in Online Firms: Scale Development and Initial Test," *Journal of Management Information Systems* (19:1), pp. 211-241 (doi: 10.1080/07421222.2002.11045715).

Bhattacherjee, A., Limayem, M., and Cheung, C. M.K. 2012. "User switching of information technology: A theoretical synthesis and empirical test," *Information & Management* (49:7-8), pp. 327-333 (doi: 10.1016/j.im.2012.06.002).

California Legislative Information 2018. *SB-1001 Bots: Disclosure*. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001. Accessed 21 April 2020.

Candello, H., Pinhanez, C., and Figueiredo, F. 2017. "Typefaces and the Perception of Humanness in Natural Language Chatbots," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 3476-3487.

Cowell, A. J., and Stanney, K. M. 2005. "Manipulation of Non-Verbal Interaction Style and Demographic Embodiment to Increase Anthropomorphic Computer Character Credibility," *International Journal of Human-Computer Studies* (62:2), pp. 281-306 (doi: 10.1016/j.ijhcs.2004.11.008).

Dabholkar, P. A., and Bagozzi, R. P. 2002. "An Attitudinal Model of Technology-Based Self-Service: Moderating Effects of Consumer Traits and Situational Factors," *Journal of the Academy of Marketing Science* (30:3), pp. 184-201 (doi: 10.1177/00970302030003001).

Davison, R. M., and Martinsons, M. G. 2016. "Context is king! Considering particularism in research design and reporting," *Journal of Information Technology* (31:3), pp. 241-249 (doi: 10.1057/jit.2015.19).

Dawes, R. M. 1979. "The robust beauty of improper linear models in decision making," *American Psychologist* (34:7), pp. 571-582 (doi: 10.1037/0003-066x.34.7.571).

De Matos, C. A., Henrique, J. L., and Alberto Vargas Rossi, C. 2007. "Service Recovery Paradox: A Meta-Analysis," *Journal of Service Research* (10:1), pp. 60-77 (doi: 10.1177/1094670507303012).

De Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., and Parasuraman, R. 2016. "Almost Human: Anthropomorphism Increases Trust Resilience in Cognitive Agents," *Journal of Experimental Psychology: Applied* (22:3), pp. 331-349 (doi: 10.1037/xap0000092).

Dietvorst, B. J., Simmons, J. P., and Massey, C. 2015. "Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err," *Journal of experimental psychology. General* (144:1), pp. 114-126 (doi: 10.1037/xge0000033).

Fishbein, M., and Ajzen, I. 1975. *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*, Massachussetts: Addison-Wesley.

Fornell, C., and Larcker, D. F. 1981. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research* (18:1), pp. 39-50 (doi: 10.1177/002224378101800104).

Forsyth, D. R. 1987. *Social psychology*, Pacific Grove, California: Brooks/Cole Publ.

Go, E., and Sundar, S. S. 2019. "Humanizing Chatbots: The Effects of Visual, Identity and Conversational Cues on Humanness Perceptions," *Computers in Human Behavior* (doi: 10.1016/j.chb.2019.01.020).

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., De Visser, E. J., and Parasuraman, R. 2011. "A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction," *Human factors* (53:5), pp. 517-527 (doi: 10.1177/0018720811417254).

Hart, C. W. L., and Johnson, M. D. 1999. "Growing the Trust Relationship," *Marketing Management* (8:1), pp. 9-19.

Heider, F. 1958. *The Psychology of Interpersonal Relations*, Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Hendriks, F., Ou, C., Amiri, A. K., and Bockting, S. 2020. "The power of computer-mediated communication theories in explaining the effect of chatbot introduction on user experience," *Proceedings of the 53 Hawaii International Conference on System Sciences*.

Henseler, J., Ringle, C. M., and Sarstedt, M. 2015. "A new criterion for assessing discriminant validity in variance-based structural equation modeling," *Journal of the Academy of Marketing Science* (43:1), pp. 115-135 (doi: 10.1007/s11747-014-0403-8).
Hrebiniak, L. G. 1974. "Effects of Job Level and Participation on Employee Attitudes and Perceptions of Influence," *Academy of Management Journal* (17:4), pp. 649-662 (doi: 10.5465/255644).
Huang, M.-H., and Rust, R. T. 2018. "Artificial Intelligence in Service," *Journal of Service Research* (21:2), pp. 155-172 (doi: 10.1177/1094670517752459).
Hulland, J., Baumgartner, H., and Smith, K. M. 2018. "Marketing Survey Research Best Practices: Evidence and Recommendations from a Review of JAMS articles," *Journal of the Academy of Marketing Science* (46:1), pp. 92-108 (doi: 10.1007/s11747-017-0532-y).
Ishowo-Oloko, F., Bonnefon, J.-F., Soroye, Z., Crandall, J., Rahwan, I., and Rahwan, T. 2019. "Behavioural evidence for a transparency–efficiency tradeoff in human–machine cooperation," *Nature Machine Intelligence* (1:11), pp. 517-521 (doi: 10.1038/s42256-019-0113-5).
Kanazawa, S. 1992. "Outcome or Expectancy? Antecedent of Spontaneous Causal Attribution," *Personality and Social Psychology Bulletin* (18:6), pp. 659-668 (doi: 10.1177/0146167292186001).
Kelley, S. W., Hoffman, K. D., and Davis, M. A. 1993. "A Typology of Retail Failures and Recoveries," *Journal of Retailing* (69:4), pp. 429-452.
Kim, and Kankanhalli 2009. "Investigating User Resistance to Information Systems Implementation: A Status Quo Bias Perspective," *MIS Quarterly* (33:3), p. 567 (doi: 10.2307/20650309).
Komiak, S. X., and Benbasat, I. 2004. "Understanding Customer Trust in Agent-Mediated Electronic Commerce, Web-Mediated Electronic Commerce, and Traditional Commerce," *Information Technology and Management* (5:1/2), pp. 181-207 (doi: 10.1023/B:ITEM.0000008081.55563.d4).
Krämer, T., Weiger, W. H., Gouthier, M. H. J., and Hammerschmidt, M. 2020. "Toward a theory of spirals: the dynamic relationship between organizational pride and customer-oriented behavior," *Journal of the Academy of Marketing Science* (48:6), p. 1-21 (doi: 10.1007/s11747-019-00715-0).
Leung, E., Paolacci, G., and Puntoni, S. 2018. "Man Versus Machine: Resisting Automation in Identity-Based Consumer Behavior," *Journal of Marketing Research* (55:6), pp. 818-831 (doi: 10.1177/0022243718818423).
Leviathan, Y., and Matias, Y. 2018. *Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone*. https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html. Accessed 22 April 2020.
Luger, E., and Sellen, A. 2016. ""Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents," in *Proceedings of the 34th Annual CHI Conference on Human Factors in Computing Systems*, pp. 5286-5297.
Luo, X., Tong, S., Fang, Z., and Qu, Z. 2019. "Machines versus Humans: The Impact of AI Chatbot Disclosure on Customer Purchases," *Marketing Science,* Forthcoming.
Markets and Markets 2019. *Chatbot Market*. https://www.marketsandmarkets.com/Market-Reports/smart-advisor-market-72302363.html.
McCollough, M. A., Berry, L. L., and Yadav, M. S. 2000. "An Empirical Investigation of Customer Satisfaction after Service Failure and Recovery," *Journal of Service Research* (3:2), pp. 121-137 (doi: 10.1177/109467050032002).
Mone, G. 2016. "The Edge of the Uncanny," *Communications of the ACM* (59:9), pp. 17-19 (doi: 10.1145/2967977).
Moorman, C., Deshpandé, R., and Zaltman, G. 1993. "Factors Affecting Trust in Market Research Relationships," *Journal of Marketing* (57:1), pp. 81-101 (doi: 10.1177/002224299305700106).
Morgan, R. M., and Hunt, S. D. 1994. "The Commitment-Trust Theory of Relationship Marketing," *Journal of Marketing* (58:3), pp. 20-38 (doi: 10.1177/002224299405800302).
Murgia, A., Janssens, D., Demeyer, S., and Vasilescu, B. 2016. "Among the Machines: Human-Bot Interaction on Social Q&A Websites," in *#chi4good: CHI 2016: San Jose, CA, USA, May 7-12*, pp. 1272-1279.
Nass, C., and Moon, Y. 2000. "Machines and Mindlessness: Social Responses to Computers," *Journal of Social Issues* (56:1), pp. 81-103 (doi: 10.1111/0022-4537.00153).
Nunamaker, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., and Patton, M. W. 2011. "Embodied Conversational Agent-Based Kiosk for Automated Interviewing," *Journal of Management Information Systems* (28:1), pp. 17-48 (doi: 10.2753/MIS0742-1222280102).

Riedl, R., Mohr, P., Kenning, P., Davis, F., and Heekeren, H. 2011. "Trusting Humans and Avatars: Behavioral and Neutral Evidence," in *Proceedings of the International Conference on Information Systems,* Shanghai, China.

Robinette, P., Howard, A., and Wagner, A. R. 2017. "Conceptualizing Overtrust in Robots: Why Do People Trust a Robot That Previously Failed?" in *Autonomy and Artificial Intelligence: A Threat or Savior?* W. F. Lawless, R. Mittu, D. Sofge and S. Russell (eds.), Cham: Springer International Publishing, pp. 129-155.

Ross, L. 1977. "The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process," in *Advances in experimental social psychology*, L. Berkowitz (ed.), New York: Academic Press, pp. 173-220.

Sameh, A.-N., Benbasat, I., and Cenfetelli, R. 2010. "Trustworthy Virtual Advisors and Enjoyable Interactions: Designing for Expressiveness and Transparency," in *Proceedings of the 18th European Conference on Information Systems,* Regensburg, Germany.

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., and Hancock, P. A. 2016. "A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems," *Human factors* (58:3), pp. 377-400 (doi: 10.1177/0018720816634228).

Schuetzler, R. M., Giboney, J. S., Grimes, G. M., and Nunamaker, J. F., JR. 2018. "The Influence of Conversational Agents on Socially Desirable Responding," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, pp. 283-292.

Shevat, A. 2017. *Designing Bots*: *Creating Conversational Experiences*, Beijing: O'Reilly Media.

Shi, W., Wang, X., Oh, Y. J., Zhang, J., Sahay, S., and Yu, Z. 2020. "Effects of Persuasive Dialogues: Testing Bot Identities and Inquiry Strategies," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems - CHI '20,* Hawaii, USA.

Skjuve, M., Haugstveit, I. M., Følstad, A., and Brandtzaeg, P. B. 2019. "Help! Is My Chatbot Falling into the Uncanny Valley? An Empirical Study of User Experience in Human-Chatbot Interaction," *Human Technology* (15:1), pp. 30-54 (doi: 10.17011/ht/urn.201902201607).

van Vaerenbergh, Y., Orsingher, C., Vermeir, I., and Larivière, B. 2014. "A Meta-Analysis of Relationships Linking Service Failure Attributions to Customer Outcomes," *Journal of Service Research* (17:4), pp. 381-398 (doi: 10.1177/1094670514538321).

Wallenburg, C. M. 2009. "Innovation in Logistics Outsourcing Relationships: Proactive Improvement by Logistics Service Providers as a Driver of Customer Loyality," *Journal of Supply Chain Management* (45:2), pp. 75-93.

Webster, C., and Sundaram, D. S. 1998. "Service consumption criticality in failure recovery," *Journal of Business Research* (41:2), pp. 153-159 (doi: 10.1016/S0148-2963(97)00004-0).

Webster, C., and Sundaram, D. S. 2009. "Effect of service provider's communication style on customer satisfaction in professional services setting: the moderating role of criticality and service nature," *Journal of Services Marketing* (23:2), pp. 104-114.

Weiner, B. 1985. ""Spontaneous" Causal Thinking," *Psychological Bulletin* (97:1), pp. 74-84 (doi: 10.1037/0033-2909.97.1.74).

Weiner, B. 2000. "Attributional Thoughts about Consumer Behavior," *Journal of Consumer Research* (27:3), pp. 382-387 (doi: 10.1086/317592).

Wilson, H. J., Daugherty, P. R., and Morini-Bianzino, N. 2017. "The Jobs That Artificial Intelligence Will Create," *MIT Sloan Management Review* (58:4), pp. 13-17.

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., and Martins, A. 2018. "Brave New world: Service Robots in the Frontline," *Journal of Service Management* (29:5), pp. 907-931 (doi: 10.1108/JOSM-04-2018-0119).

Wünderlich, N. V., and Paluch, S. 2017. "A Nice and Friendly Chat with a Bot: User Perceptions of AI-Based Service Agents," in *38th International Conference on Information Systems,* Seoul, South Korea, pp. 1-17.

Zhao, X., Lynch, J. G., and Chen, Q. 2010. "Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis," *Journal of Consumer Research* (37:2), pp. 197-206.