Faster Is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction

4 authors:

Ulrich Gnewuch
Karlsruhe Institute of Technology
13 PUBLICATIONS 189 CITATIONS

Stefan Morana
Universität des Saarlandes
67 PUBLICATIONS 491 CITATIONS

Marc T. P. Adam
University of Newcastle
97 PUBLICATIONS 971 CITATIONS

Alexander Maedche
Karlsruhe Institute of Technology
349 PUBLICATIONS 12,290 CITATIONS

FASTER IS NOT ALWAYS BETTER: UNDERSTANDING THE EFFECT OF DYNAMIC RESPONSE DELAYS IN HUMAN-CHATBOT INTERACTION

Research paper

Ulrich Gnewuch, Karlsruhe Institute of Technology (KIT), Institute of Information Systems and Marketing (IISM), Karlsruhe, Germany, ulrich.gnewuch@kit.edu

Stefan Morana, Karlsruhe Institute of Technology (KIT), Institute of Information Systems and Marketing (IISM), Karlsruhe, Germany, stefan.morana@kit.edu

Marc T. P. Adam, The University of Newcastle, School of Electrical Engineering and Computing, Callaghan, New South Wales, Australia, marc.adam@newcastle.edu.au

Alexander Maedche, Karlsruhe Institute of Technology (KIT), Institute of Information Systems and Marketing (IISM), Karlsruhe, Germany, alexander.maedche@kit.edu

Abstract

A key challenge in designing conversational user interfaces is to make the conversation between the user and the system feel natural and human-like. In order to increase perceived humanness, many systems with conversational user interfaces (e.g., chatbots) use response delays to simulate the time it would take humans to respond to a message. However, delayed responses may also negatively impact user satisfaction, particularly in situations where fast response times are expected, such as in customer service. This paper reports the findings of an online experiment in a customer service context that investigates how user perceptions differ when interacting with a chatbot that sends dynamically delayed responses compared to a chatbot that sends near-instant responses. The dynamic delay length was calculated based on the complexity of the response and complexity of the previous message. Our results indicate that dynamic response delays not only increase users' perception of humanness and social presence, but also lead to greater satisfaction with the overall chatbot interaction. Building on social response theory, we provide evidence that a chatbot's response time represents a social cue that triggers social responses shaped by social expectations. Our findings support researchers and practitioners in understanding and designing more natural human-chatbot interactions.

Keywords: Chatbot, response delay, social cue, perceived humanness, social presence.

1 Introduction

As with many promising technologies, conversational user interfaces are getting hyped as the "next big thing" (Gartner, 2017). These interfaces enable users to interact with information and communications technology (ICT) using natural language, just like engaging in a conversation with another human being (Dale, 2016; McTear et al., 2016). Well-known examples of systems that build on conversational user interfaces can be found in mobile devices (e.g., Apple's Siri) or on instant messaging platforms (e.g., Facebook Messenger chatbots). Recent technological advancements combined with the shift towards messaging as a primary channel for both personal and professional communication have contributed to the enormous increase in popularity of chatbots in particular (Dale, 2016; Følstad and Brandtzæg, 2017). Across all industries, companies are experimenting with or have already implemented chatbots in an effort to support users in finding relevant information about products and services as well as carrying out basic tasks to facilitate transactions (e.g., book flights). However, despite the hype, interacting with chatbots often feels unnatural and awkward (Ben Mimoun et al., 2012; Moore et al., 2017; Schuetzler et al., 2014). Due to the complexity of natural language, there are still difficulties in understanding ambiguous user input (Klopfenstein et al., 2017; Moore et al., 2017). However, it has been realized that human-chatbot interaction is also constrained by the fact that designing conversational user interfaces is quite different from designing graphical user interfaces (Følstad and Brandtzæg, 2017; Jenkins et al., 2007). While user interface design commonly focuses on graphical elements and site structures, the design of human-chatbot interaction critically rests on the way conversations are facilitated (Følstad and Brandtzæg, 2017). Specifically, the design of natural conversations has been identified as a key challenge for increasing user satisfaction and adoption of chatbots (Moore et al., 2017).

Researchers have studied chatbots and other types of conversational agents for decades, starting with the well-known ELIZA (Weizenbaum, 1966). We regard chatbots as a subclass of conversational agents that are designed to interact with users using written, natural language, typically in messaging applications or on websites (Følstad and Brandtzæg, 2017). Since chatbots can fulfill the role of service employees (Larivière et al., 2017) and are able to support consumers in their decision-making when searching for and selecting products, they can also act as recommendation agents (Qiu and Benbasat, 2009; Wang et al., 2016; Xiao and Benbasat, 2007). Several studies on recommendation agents (e.g., Qiu and Benbasat, 2009) and chatbots (e.g., Appel et al., 2012; Schuetzler et al., 2014) have investigated how human-like characteristics of such systems influence users' perception and behavior. Moreover, researchers in the field of human-computer interaction (HCI) have shown that users mindlessly apply social rules and expectations in their interaction with computers that use natural language or display other human characteristics (Nass et al., 1994; Nass and Moon, 2000). These "social cues" can substantially affect users' perception, which significantly impacts adoption and use of these systems (Candello et al., 2017; Kang and Gratch, 2014; Nass and Moon, 2000).

System response time has been identified as a critical factor for user satisfaction and productivity (Hoxmeier and DiCesare, 2000; Muylle et al., 2004). In human-chatbot interaction, response time can play an important role in how users perceive a chatbot (Fraser, 1997; Holtgraves et al., 2007; Moon, 1999). Generally, it is assumed that very fast responses make a chatbot appear unhuman-like (Holtgraves and Han, 2007) and do not give users the feeling of a natural conversation (Appel et al., 2012; Shechtman and Horowitz, 2003). Since current chatbots are able to respond almost instantly to a user's input, some of them delay their responses to simulate the time it would take a human to read messages and respond to them (Appel et al., 2012; Klopfenstein et al., 2017; Shechtman and Horowitz, 2003). The underlying assumption is that delayed responses increase a chatbot's perceived humanness and make conversations more natural (Appel et al., 2012; Klopfenstein et al., 2017). However, delayed responses may also negatively impact user satisfaction (Hoxmeier and DiCesare, 2000; Taylor et al., 2016), particularly in situations in which users expect fast response times, such as customer service (McLean and Wilson, 2016; Song and Zinkhan, 2008). For example, a slow response time of customer service agents in live chat negatively influences user satisfaction and website quality perceptions (Song and Zinkhan, 2008).

So far, only little empirical research has been conducted on the impact of response delays on users' perception of chatbots. Previous research has analyzed how users perceive the persuasiveness and personality of chatbots with different response times (Holtgraves et al., 2007; Moon, 1999). Although these studies have been conducted in different contexts, they provide inconsistent findings on the effect of response delays: While Moon (1999) finds that users negatively evaluate short response times (i.e., indicating a lack of cognitive effort), Holtgraves et al. (2007) argue that a chatbot that responds quickly may be perceived more positively than a chatbot sending delayed responses. Based on these inconsistent findings, we argue that there is a lack of understanding as to whether delayed responses result in more favorable perceptions of a chatbot. More specifically, it is unclear whether dynamically delayed responses (e.g., delays based on message characteristics, such as complexity), as compared to near-instant responses, can increase a chatbot's humanness and make the conversation feel more natural in a customer service context. Addressing this gap is not only important because of the growing use of chatbots in customer service (Gartner, 2018), but also because response delays are increasingly implemented in practice, despite the uncertainty about their impact on users' perceptions (Crozier, 2017; Klopfenstein et al., 2017). Therefore, we address the following research question:

How do dynamically delayed responses affect users' perception of a customer service chatbot as compared to near-instant responses?

In this paper, we present the results of a two-condition online experiment addressing the effect of dynamically delayed responses on users' perception of a customer service chatbot. Drawing on social response theory (Nass et al., 1994; Nass and Moon, 2000; Reeves and Nass, 1996), we investigate the effect of dynamic delays on perceived humanness and social presence as well as on user satisfaction with the chatbot interaction. Our results show that dynamically delayed responses, as compared to near-instant responses, not only increase a user's perception of a chatbot's humanness and social presence, but also lead to a greater satisfaction with the overall interaction. Overall, our research contributes to the current discussion on how to design conversational user interfaces and, particularly, chatbots to make the interaction feel more natural. Additionally, our formula for calculating dynamic response delays benefits practitioners, such as chatbot designers, by providing them with an easily applicable technique for improving human-chatbot interaction. Our main theoretical contribution is to show that response time is a social cue that generates social responses from users.

The remainder of this paper is organized as follows. Section two introduces related work on chatbots and response time as well as theoretical foundations. In section three, we derive three hypotheses concerning the effect of dynamically delayed responses. Subsequently, we describe our research method and present the results of our analysis. In section six, we discuss the implications of our results based on social response theory and provide design implications for chatbots and conversational user interfaces in general. We then conclude the paper with a summary of our main findings and contributions.

2 Related Work and Theoretical Background

The idea of humans interacting with intelligent systems using natural language has been around for decades and has been featured in many science fiction books and movies (McTear et al., 2016). With recent advances in technology, many ICTs now provide conversational user interfaces, ranging from smartphones (e.g., Apple's Siri, Samsung's Bixby) and smart speakers (e.g., Amazon's Alexa) to personal computers (e.g., Microsoft's Cortana). While these voice-based "personal assistants" have become very popular in the last years (Maedche et al., 2016), organizations are increasingly shifting their attention to chatbots that build on text-based conversational user interfaces (Gnewuch et al., 2017) and can be made available on instant messaging platforms or websites (Gartner, 2017; Klopfenstein et al., 2017). Chatbots have their origins in the ELIZA system developed by Weizenbaum (1966). While early chatbots were built to simulate human conversation using pattern matching algorithms, recent technological advances have enormously improved their capabilities (Knijnenburg and Willemsen, 2016; McTear et al., 2016). Consequently, many organizations are investigating how they can make use of chatbots, for example, as a cost-effective solution in customer service (Oracle, 2016). Examples can be found in industries ranging from travel (e.g., online check-in or flight booking) and retail (e.g., product selection) to financial services (e.g., money transfer).

Although a vast amount of research has been conducted on chatbots and conversational user interfaces in general, most studies focus on their technical aspects, such as by developing better natural language processing algorithms or new architectures (Chakrabarti and Luger, 2015; Sarikaya, 2017). Therefore, it has been largely neglected that other factors can also significantly influence the human-chatbot interaction (Følstad and Brandtzæg, 2017). In the following, we introduce related work on one such factor: a chatbot's response time. Subsequently, we provide an overview of social response theory that serves as the theoretical foundation for our research.

2.1 Response Time of Chatbots

Research has shown that a system's *response time* (sometimes also called *response latency* or *response speed*) is an important factor that influences user satisfaction and other aspects related to perceived system quality (Hoxmeier and DiCesare, 2000; Rushinek and Rushinek, 1986; Taylor et al., 2016;

Wixom and Todd, 2005). In the context of face-to-face and computer-mediated communication, response time has also been found to affect people's impressions of others (Ho et al., 2016; Moon, 1999). Here, response time refers to the amount of time it takes for a person to respond to the sender's input as well as the lag time between messages (Moon, 1999). Short response times are perceived as a lack of thought and cognitive effort, whereas long response times are perceived as an indication of deception (Ho et al., 2016; Moon, 1999).

In human-chatbot interaction, response time can be expected to affect conversation flow, message content, and a user's judgement of whether s/he is talking to a computer or human (Fraser, 1997). While only a few researchers have explicitly investigated how different response times influence human-chatbot interaction, several studies can be found that delayed the responses of a chatbot (or other systems with a conversational user interface) to simulate the time it would take a human to read and respond to a message (Appel et al., 2012; Holtgraves and Han, 2007; Shechtman and Horowitz, 2003; Skowron et al., 2011). For example, Appel et al. (2012) delayed the responses of a chatbot by 15-30 seconds to "give the participant a feeling of real-time conversation" (p. 4). Holtgraves and Han (2007) calculated a time delay for each response by a chatbot based on its number of characters (i.e., 50 milliseconds per character). Shechtman and Horowitz (2003) implemented a delay "which helped to maintain the illusion that a partner was taking the time to read and respond" (p. 4). However, only the studies by Moon (1999) and Holtgraves et al. (2007) have specifically varied a chatbot's response time using static delays to understand the effects on users' perception. Moon (1999) found that response time significantly influences the persuasiveness of a chatbot's messages. In a lab experiment in which participants had to solve the desert survival problem (Lafferty et al., 1974) with a chatbot, information by the chatbot was found more persuasive when responses were sent with a moderate (5-10 s) rather than a short (0-1 s) or long (13-18 s) response delay. However, in an online experiment conducted by Holtgraves et al. (2007), a chatbot that responded relatively quickly (1 s) was perceived as more conscientious and extraverted than a chatbot that responded more slowly (10 s). In this experiment, participants had a general conversation with an ALICE chatbot before being asked to rate the chatbot's personality (Goldberg, 1992). Although their experimental design did not include a condition with a medium length delay, they suggest that a "bot that responds quickly may be perceived more positively than a slowly-responding bot" (Holtgraves and Han, 2007, p. 2172). Based on the findings of both studies, we argue that it is not clear whether delayed responses will result in more favorable perceptions of a chatbot. Although the aforementioned studies have been conducted in different contexts, they provide inconsistent findings on the effect of response delays. However, despite this uncertainty, it is common practice for many chatbot designers to introduce delays in an effort to make the conversation with their chatbots more natural and human-like (Crozier, 2017; Klopfenstein et al., 2017). Therefore, we argue that there is a need to better understand the effect of response delays of chatbots (e.g., in form of dynamic delays).

2.2 Social Response Theory: Computers are Social Actors

Social response theory posits that social cues from computers, such as interacting with others, using natural language, or playing social roles, trigger mindless responses from humans, no matter how rudimentary those cues are (Moon, 2000, 2003; Nass et al., 1994; Nass and Moon, 2000; Reeves and Nass, 1996). This theory emerged from the "Computers are Social Actors" (CASA) paradigm (Nass et al., 1994) and many studies, particularly in the field of HCI, have used it as a theoretical foundation to explain the effects of different social cues on the perception of human-like technologies.

Following the CASA paradigm, researchers have also investigated how users respond to social cues from chatbots and other systems with conversational user interfaces (e.g., embodied conversational agents). The majority of studies has examined the effect of visual cues such as the presence of a virtual character (von der Pütten et al., 2010), smiling (Verhagen et al., 2014), or the typefaces used in the conversational user interface (Candello et al., 2017). However, other types of social cues have also been found to affect users' perception of a chatbot such as the degree of interactivity (Schuetzler et al., 2014), communication style (Verhagen et al., 2014), or assumed agency (i.e., whether users think they are interacting with a human or computer) (Appel et al., 2012). The common understanding in this area of research is that even minimal social cues can generate a wide range of social responses from users (Candello et al., 2017; Nass et al., 1994).

As illustrated in the previous section, a chatbot's response time can also trigger social responses that are shaped by users' social expectations. For example, response time has been found to influence the persuasiveness of a chatbot's messages (Moon, 1999) and perceptions of a chatbot's personality (Holtgraves et al., 2007). Therefore, we argue that response time represents a social cue that plays an important role in the interaction with a chatbot. Even though users know they are interacting with a chatbot (i.e., with a machine that does not need to read, think about, or enter a message on a keyboard), they still expect certain characteristics of human conversation in their interaction (Nass et al., 1994; Nass and Moon, 2000). More specifically, fast responses will be perceived as unnatural because humans would not be able to respond instantaneously, as their response time depends, for example, on the complexity of the message. Therefore, it can be argued that responses sent after a dynamic delay may "make the conversation more familiar to the user" (Klopfenstein et al., 2017, p. 559). Thus, *dynamically delayed responses* that, in some way, reflect the time a human would need to (1) read and process a message by another person as well as to (2) formulate and enter a response, may better match a user's social expectations and thus, positively influence their perception of a chatbot.

3 Hypotheses

Although little empirical research has been conducted on how response delays affect users' perception of chatbots (Holtgraves et al., 2007; Moon, 1999), many researchers as well as practitioners have used response delays in an effort to make their chatbots appear more human-like and make the interaction feel more natural (Appel et al., 2012; Crozier, 2017; Klopfenstein et al., 2017; Shechtman and Horowitz, 2003). Drawing on social response theory (Nass et al., 1994; Nass and Moon, 2000; Reeves and Nass, 1996), we investigate how *dynamically delayed* compared to *near-instant responses* affect users' perception of chatbots. In the following, we formulate three hypotheses regarding the effect on perceived humanness, social presence, and satisfaction.

According to social response theory, users respond to human characteristics in their interaction with a computer, even if they know that they are not interacting with a human being (Nass and Moon, 2000). Since previous research has found that response time influences the persuasiveness of a chatbot's messages (Moon, 1999) and perceptions of a chatbot's personality (Holtgraves et al., 2007), it can be argued that response time also serves as a social cue that triggers social responses. In contrast to human users who need some time to read and respond to a message (e.g., in instant messengers), chatbots can respond almost instantly. However, many researchers discovered that quick responses make a chatbot appear unhuman-like (Appel et al., 2012; Holtgraves et al., 2007; Shechtman and Horowitz, 2003). Consequently, we argue that a chatbot using dynamic delays calculated based on message complexity will be perceived as more human-like (for details on the calculation, see section 4.3). This is because dynamic delays correspond to the time that a human would need to read and respond (i.e., more complex messages take more time to read and more complex responses take more time to formulate) (Mentzer et al., 2007; Vronay et al., 1999). Thus, we propose that:

> **H_1**: *A customer service chatbot that sends dynamically delayed responses will yield a higher level of perceived humanness than a customer service chatbot that sends near-instant responses*.

The concept of social presence has long been employed to study the social aspects of technologies, such as websites (Cyr et al., 2009), recommendation agents (Qiu and Benbasat, 2009), or avatars (von der Pütten et al., 2010). Social presence was originally defined as "the degree of salience of the other person in a mediated communication and the consequent salience of their interpersonal interactions" (Short et al., 1976, p. 65). However, Gefen and Straub (2004) suggest that "the perception of social presence can still be created despite the lack of actual human contact" (p. 410). Following this perspective on social presence, many studies have shown that social cues from technology create perceptions of social presence (Gefen and Straub, 2003; Hassanein and Head, 2007; Hess et al., 2009). Therefore, we argue that feelings of a chatbot's social presence can also be installed through dynamically delayed responses because they simulate a sense of interacting with another human. In contrast, near-instant responses are less able to convey feelings of social presence. Thus, we hypothesize that:

H$_2$: *A customer service chatbot that sends dynamically delayed responses will yield a higher level of perceived social presence than a customer service chatbot that sends near-instant responses*.

In the context of customer service, satisfaction is an important indicator of how customers feel about their interaction with a service provider (Barger and Grandey, 2006; Oliver, 1997). According to Bitner et al. (2000), technology is a key enabler of service encounter satisfaction and can be used to provide customers with pleasing experiences. Current technologies are able to incorporate a range of social cues (e.g., friendliness, smiling) that are important for delivering successful service encounters (van Doorn et al., 2017; Verhagen et al., 2014). Assuming that a chatbot's response time also represents a social cue, we argue that it also influences how satisfied users are with the chatbot interaction. Moreover, in human-human communication, too short response times are perceived as a lack of thought and cognitive effort (Kang et al., 2013; Moon, 1999). Therefore, dynamically delayed responses can convey the impression to users that the chatbot puts more cognitive effort into its responses, instead of just quickly "firing off" a response. Thus, we propose that:

H$_3$: *Users will be more satisfied with the interaction with a customer service chatbot that sends dynamically delayed responses than with the interaction with a customer service chatbot that sends near-instant responses.*

Figure 1 depicts our research model and hypotheses.

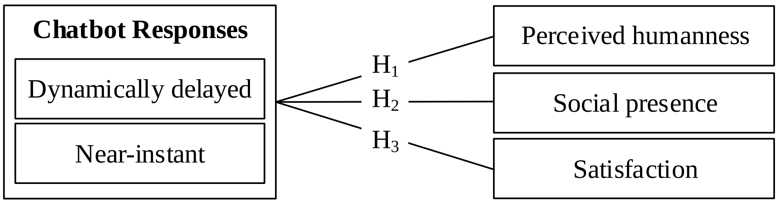

Figure 1: Research model

4 Method

To investigate whether dynamic response delays influence users' perception of chatbots, we conducted an online experiment in a customer service context. In the following, we describe our experimental design, our formula for calculating dynamic delays, and the measures used in the post-experiment questionnaire.

4.1 Experimental Design

We used a between-subjects design with two experimental conditions: near-instant responses (Control) and dynamically delayed responses (Treatment). In the near-instant condition, participants interacted with a chatbot that sent responses almost instantly. This means that its responses were only delayed by the time required by the network to send user input to the chatbot and return its response (i.e., caused by physical limits of data transmission). Therefore, the chatbot's response time was relatively consistent with an average of one second. In the dynamic delay condition, participants interacted with a chatbot that sent responses after an additional, dynamically calculated delay. This delay was calculated based on message complexity and is described in detail in section 4.3. In both conditions, participants were informed before the experiment that they were interacting with a chatbot, not a human. Our experimental platform randomly assigned the participants to one of the two conditions.

To conduct the experiment, we developed two chatbots using Microsoft's Bot Builder software development kit. They were integrated into the experimental platform via a simple chat window (see Figure 2) and a custom-built interface that communicated with the chatbots' messaging endpoint. Microsoft's

Language Understanding Intelligent Services (LUIS) was used to process natural language user input (i.e., to recognize user intentions and extract entities, such as the names of different mobile phone plans). Both chatbots used the same language model and had identical dialogs implemented. The only difference between them was their respective response time. All conversation data was stored in a log file for subsequent analysis.

4.2 Experimental Task

The online experiment was conducted in a customer service context. As illustrated in Figure 2, participants were shown a fictitious copy of last month's mobile phone bill, which indicated that their current plan did not fit their actual usage patterns (e.g., their data usage was much higher than the amount included in their plan, resulting in high additional costs). Therefore, participants were asked to interact with a chatbot to find out whether they could save money by switching to a better mobile phone plan. The chatbot was introduced as an expert on the plans of all mobile phone providers (i.e., analogous to online price comparison portals). During the conversation, the chatbot asked about the participant's usage patterns (e.g., how much data was used). Some of the information was given on the fictitious bill, but participants could freely choose additional features (e.g., international calling) and decide how much they would be willing to pay for a new plan. After collecting all information, the chatbot recommended a randomly generated plan that better met the participant's requirements. We selected this experimental task because it represents a relatively realistic scenario for a human-chatbot interaction and similar scenarios have been used by previous studies (e.g., Jenkins et al., 2007; Verhagen et al., 2014). Although our chatbot was able to understand and process most of the participant's input regardless of the precise wording used, we decided to implement a relatively structured dialog to ensure a high level of comparability across the two treatment conditions (i.e., all participants had a similar conversation with the chatbot).

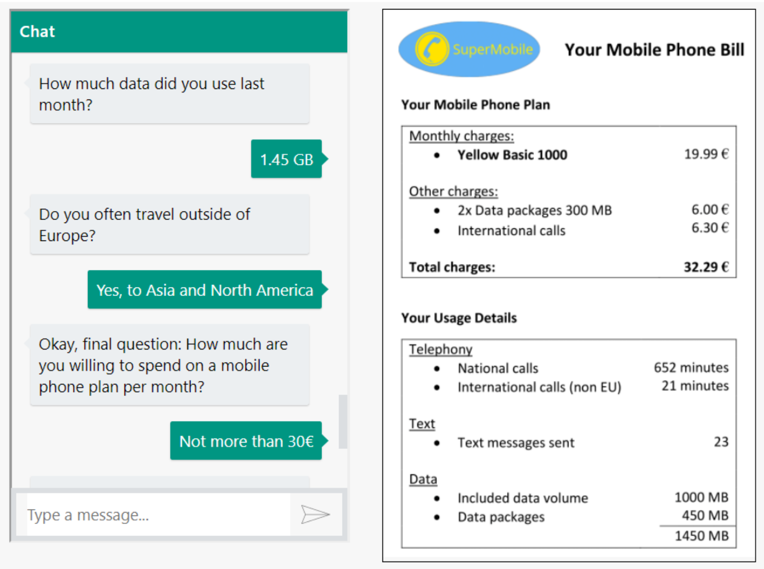

Figure 2. Experimental task and setup

4.3 Treatment Configuration: Dynamic Response Delays

To derive a formula for calculating dynamic response delays, we reviewed literature on chatbots and conversational agents in information systems (IS), HCI, and related domains. Interestingly, previous work has primarily used static or random time delays, independent of the characteristics of the response or previous message (e.g., Appel et al., 2012; Holtgraves et al., 2007; Shechtman and Horowitz, 2003). However, human response time depends on both the message received (e.g., reading and making sense of it) and the subsequent response (e.g., entering it in the chat box) (Derrick et al., 2013). Holtgraves and Han (2007) calculated a response delay based on the number of characters of the chatbot's response (i.e., 50 milliseconds per character), but did not consider the message sent by a user before or other factors that affect the response time. Literature on face-to-face and computer-mediated communication states that a person's response time is primarily made up of two components: (1) the time required to read and process another person's message, and (2) the time required to formulate and write a response (Derrick et al., 2013; Mentzer et al., 2007; Vronay et al., 1999). Both parts depend on the complexity of the individual message, such that more complex messages take longer to read and more complex responses longer to formulate (Mentzer et al., 2007). Semantic difficulty and syntactic complexity have been used as indicators of the complexity of a text (Lennon and Burdick, 2004). While semantic difficulty describes the use of words, their structure and length, syntactic complexity primarily reflects the sentence length. Several complexity measures have been used in research (Khawaja et al., 2010). We selected the Flesch-Kincaid grade level (Kincaid et al., 1975) because it has been used before to determine the complexity of a message in computer-mediated communication (Walther, 2007). It calculates the language complexity (C) of a message (m) using average sentence lengths and average syllables per word according to the following formula (Kincaid et al., 1975):

$$C(m) = 0.39 * \left(\frac{total\ words}{total\ sentences}\right) + 11.8 * \left(\frac{total\ syllables}{total\ words}\right) - 15.59$$

The complexity values range from -3.40 to positive infinity. Based on that, we calculate a time delay (D) in milliseconds for any given message (m) using the complexity value ($C(m)$) as input:

$$\begin{array}{ll} D(m) = 0.5 * \ln(C(m) + 0.5) + 1.5, & C(m) > 0 \\ D(m) = 0 & C(m) \leq 0 \end{array}$$

We calibrated this formula using feedback from pretests with four researchers not involved in this study. For short messages with a low complexity ($C(m) \leq 0$), the delay was set to zero. Using this formula, we calculated a delay for (1) the previous message (m_{n-1}) that was sent either by the user or by the chatbot and (2) the response sent by the chatbot (m_n). This was because some of the chatbot's responses were segmented into multiple, consecutive messages and not sent as a single, aggregated one. Finally, both delays were summed up to calculate the total delay (D_{total}) in milliseconds:

$$D_{total}(m_n) = D(m_{n-1}) + D(m_n)$$

Thus, responses from the chatbot were sent with an additional, dynamically calculated delay. More complex responses following a more complex message from the user were noticeably delayed (for examples, see Table 1). Less complex responses were only minimally delayed (if at all). The result of our calculation cannot be regarded as the "optimal" delay because many other factors might influence the perception of a response time (e.g., familiarity with chatbots). Nevertheless, we argue that the formula provides a sufficient approximation for this experiment. Table 1 exemplary illustrates how the total dynamic delay lengths correspond to the complexity of two interrelated messages.

Messages exchanged between human and chatbot	**Message complexity**	**Total response delay**
Human: "*Hey*" Chatbot: "*Hey*"	0 0	0ms + 0ms = 0ms
Human: „*Help me!*" Chatbot: *"If you need a new mobile phone plan, try asking me things like: 'new mobile phone plan'"*	0 4.922	0ms + 2351ms = 2351ms
Human: „*Yes, I often travel on business to Asia and North America."* Chatbot: *„Okay, final question: How much are you willing to spend on a mobile phone plan per month?"*	2.504 2.995	2049ms + 2126ms = 4175ms

Table 1. Examples of how the dynamic response delay was calculated

It must be noted that the dynamic delay length does not include the network delay of approximately one second that is caused by physical limits of data transmission over the Internet. For the chatbots in both conditions, this was the time required to send user input to the chatbot's messaging endpoint, to make a call to the natural language processing service LUIS, and to return the response to the user.

4.4 Measures

After interacting with the chatbot, participants were asked to complete a questionnaire about their perception of the chatbot and their opinion of their conversation. All measures in the questionnaire were adapted from established measurement instruments, namely perceived humanness (Holtgraves et al., 2007; Holtgraves and Han, 2007), social presence (Gefen and Straub, 1997), and service encounter satisfaction (Verhagen et al., 2014). Perceived humanness was measured on a 9-point semantic differential scale, while all other items were measured on a 7-point Likert scale. Table 2 lists all measurement items used as well as the composite reliability (CR) and average variance extracted (AVE) for each construct.

Measures	**Factor loading**
Perceived humanness (CR = .891, AVE = .673)	
extremely inhuman-like - extremely human-like	.842
extremely unskilled - extremely skilled	.852
extremely unthoughtful - extremely thoughtful	.861
extremely impolite - extremely polite	*dropped (.570)*
extremely unresponsive - extremely responsive	*dropped (.496)*
extremely unengaging - extremely engaging	.719
Social presence (CR = .952, AVE = .800)	
I felt a sense of human contact with the chatbot.	.896
I felt a sense of personalness with the chatbot.	.854
I felt a sense of sociability with the chatbot.	.912
I felt a sense of human warmth with the chatbot.	.895
I felt a sense of human sensitivity with the chatbot.	.914
Service encounter satisfaction (CR = .814, AVE = .595)	
How satisfied are you with the chatbot's advice?	.702
… the way the chatbot treated you?	.756
… the overall interaction with the chatbot?	.849
CR = Composite Reliability, AVE = Average Variance Extracted	

Table 2. Measures used in the post-experiment questionnaire

We dropped two items (polite and responsive) from the perceived humanness construct because of factor loadings below .60 (Gefen and Straub, 2005). As shown in Table 2, all constructs display sufficient CR above .80 and an AVE above a suggested level of .50 (Urbach and Ahlemann, 2010).

In addition, we collected demographic information (age, gender, and education) and asked participants about their frequency of using personal assistants (e.g., Siri, Alexa) and chatbots (e.g., on Facebook Messenger, Telegram, or websites), ranging from "never" to "daily". Finally, we asked the participants for general feedback with an open-ended question ("What did you like about the chatbot / experiment?").

4.5 Participants

With an a priori power analysis using G*Power (Faul et al., 2007), we determined a sample size of at least 70 subjects (effect size = .80, alpha = .05, and power = .95). We recruited participants using mailing lists and personal networks. Participants were not compensated for their participation. Of the total of 163 participants invited, 84 participants completed the online experiment and post-experiment questionnaire (response rate = 52%). Five responses were discarded because these participants either did not follow the experimental task or provided straight-lined responses in the questionnaire. Hence, the dataset contains 79 observations (25 females and 54 males; mean age = 28.835, SD = 6.388). Most of the participants had a university education (Bachelor = 17, Master = 39, and PhD = 9). Moreover, most of the participants reported that they either never use voice-based personal assistants (n = 41) or use them only about once a month (n = 17). Similarly, they stated that they either never use chatbots (n=43) or use them only about once a month (n = 21). In total, 44 participants were in the dynamic delay condition and 35 were in the near-instant condition. The unequal distribution across the two conditions was due to the random assignment of participants by our experimental platform that did not account for participants who did not complete the study.

5 Results

Table 3 shows descriptive results and test statistics for all measures used in our questionnaire. Additionally, Figure 3 depicts the bar charts with error bars of our results. Before conducting the analysis, we checked for the homogeneity of variance of all measures. Next, we tested for a significant difference between both conditions using Student's t-tests (for perceived humanness and social presence; homogeneous variances) and Welch's t-tests (for satisfaction; non-homogenous variances). All tests were performed one-sided to examine whether dynamically delayed responses (= Treatment) positively affect users' perception of chatbots (i.e., in comparison to near-instant responses = Control).

Condition	n	Perceived humanness[1]			Social presence[2]			Satisfaction[2]		
		Mean	SD	SE	Mean	SD	SE	Mean	SD	SE
Dyn. Delayed (Treatment)	44	5.534	1.674	0.252	3.695	1.507	0.227	5.174	0.974	0.147
Near-instant (Control)	35	4.457	1.878	0.317	2.954	1.425	0.241	4.610	1.344	0.227
Test statistic		$t(77)=-2.691$, $p=0.004$, r=.293			$t(77)=-2.224$, $p=0.015$, r=.246			$t(60.04)=-2.088$, $p=0.021$, r=.260		
Hypothesis		H1 confirmed			H2 confirmed			H3 confirmed		

[1] measured on a 9-point semantic differential scale | [2] measured on a 7-point Likert scale | SD = standard deviation | SE = standard error

Table 3. Descriptive results and test statistics for both conditions

Our results reveal that there is a significant difference in perceived humanness and social presence of the chatbot between both conditions, which supports H_1 and H_2. We also found a significant difference between the conditions in overall user satisfaction with the chatbot interaction, thus supporting H_3. Subsequently, we performed a post-hoc power analysis using G*Power (Faul et al., 2007). Our results suggest that all tests have sufficient power given the relatively low sample size ($power_{perceived\_humanness}$ = .894, $power_{social\_presence}$ = .720, and $power_{satisfaction}$ = .753).

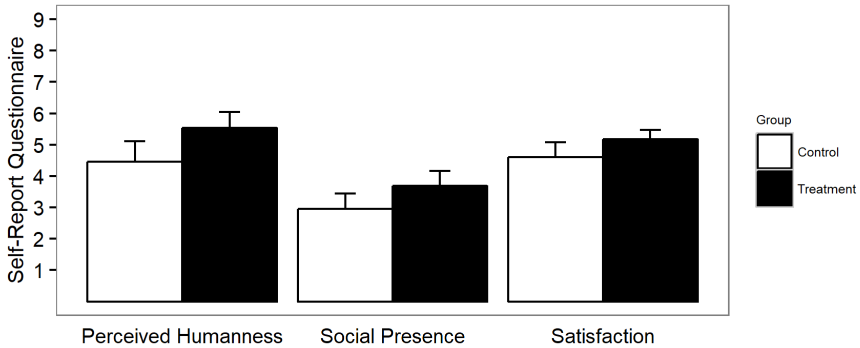

Figure 3. Differences in perceived humanness, social presence, and satisfaction between the two conditions. Note: The error bars indicate the 95% confidence interval.

Next, we performed additional robustness tests. To test whether the results of our experiment might be explained by the messages that participants sent to the chatbots, we conducted complementary analyses on the number of the messages sent by participants and on the overall interaction time. In total, 3590 messages were sent: 1415 from users (M=17.91 [SD=11.98]) and 2175 from the chatbots (M=27.53 [SD=17.10]). Importantly, the number of messages in the dynamic delay condition (M = 15.61 [SD = 6.65]) is not significantly different from those in the near-instant condition (M = 20.80 [SD = 16.06]), $t(77)=1.214$, $p=.228$). Similarly, there is no significant difference in the participant's average interaction time with the chatbot between the dynamic delay (M=12.03 min [SD=19.82 min]) and the near-instant condition (M=10.98 min [SD=11.42 min]), $t(77)=0.280$, $p>.780$. Furthermore, we tested for a difference in the demographics between the two conditions and found no significant difference in age ($t(46.4)=-1.699$, $p=0.096$), gender ($\chi^2(1)=.275$, $p=.6$), education ($\chi^2(5)=9.04$, $p=.107$), usage frequency of voice assistants ($\chi^2(4)=5.502$, $p=.24$), or usage frequency of chatbots ($\chi^2(4)=3.558$, $p=.469$). In summary, these results indicate that the differences in perceived humanness, social presence, and satisfaction, are not explained by differences in the number of messages sent, the overall interaction time, or demographics between the two conditions.

6 Discussion

The results of our online experiment suggest that dynamic response delays positively affect users' perception of customer service chatbots. The chatbot that sent dynamically delayed responses was perceived to be more human-like and to have a higher social presence than a chatbot sending near-instant responses. This finding supports the assumption that dynamically delaying responses is an effective way to "humanize" a chatbot and make conversations more natural to the user. In line with social response theory (Nass et al., 1994; Nass and Moon, 2000), we demonstrate that minimal social cues, such as different response times, trigger social expectations and processes and substantially affect users' perception of chatbots. Interestingly, our analysis also shows that response delays increase user satisfaction with the overall chatbot interaction. However, previous studies on the impact of response delays (Hoxmeier and DiCesare, 2000; Rushinek and Rushinek, 1986; Taylor et al., 2016) suggest that delays in a system's response time result in lower user satisfaction. Therefore, our results may initially appear to contradict their conclusions and seem somewhat counterintuitive: Why would anyone want to *wait longer* than necessary for a response from a chatbot? However, these studies have investigated the response times of systems with graphical user interfaces (e.g., web-based applications), in which interaction occurs through button clicks, scrolling, or swiping. By using natural language, chatbots may trigger more or other social responses than most web-based applications. We argue that the "human element" of chatbots significantly shapes their perception. Therefore, users might apply a different standard when they evaluate a chatbot's response time, which would explain the differences to studies on graphical user interfaces. Because people are used to chat with other humans who are not able to respond instantly, they appear to automatically apply the same rules and expectations when interacting with a chatbot. This effect could be further enhanced by the fact that many chatbots reside in messaging platforms, such as Facebook Messenger, Slack, or Telegram, which are commonly used to interact with other humans.

The assumption that chatbots are perceived as "social actors" (Nass et al., 1994) is further supported by the answers to the open-ended question at the end of our questionnaire that asked for general feedback from participants. Interestingly, several participants from the group that received near-instant responses stated that they felt irritated by the chatbot's fast response time (*"It was really fun, but the extremely quick answering rate is kind of confusing"*) or made a comparison with human-human communication *("No human is able to write a sentence in a fraction of a second"*). Hence, even though participants knew that they were interacting with a chatbot (i.e., a machine), they still applied the same social expectations as when interacting with a human being. Participants who interacted with the chatbot sending dynamically delayed responses did not mention such aspects and rather commented on other aspects of the experiment (e.g., the content of their conversation with the chatbot or the chatbot's limited ability to recommend plans with additional features). This finding may indicate either that they did not notice the delays or that these delays "felt right" for them. Similar results can be seen in research on humanoid robots. For example, Kanda et al. (2007) found that a humanoid robot's reaction in its body movements should be delayed to make the interaction more natural. In summary, we argue that *faster is not always better* in human-chatbot interaction. Although more research is needed, our results indicate that, in this particular context, users prefer to receive dynamically delayed than near-instant responses. Specifically, users seem to be irritated by too fast responses of the chatbot, which negatively affects their perception of the chatbot because it contradicts with their social expectations.

6.1 Implications for Designing Chatbots and Conversational User Interfaces

Our results have several implications for the design of chatbots and, to a certain extent, of conversational user interfaces. First, we show that design features, such as response delays, can have a significant impact on users' perception of chatbots. When designing chatbots, attention to details is required to make conversations feel natural and human-like. This particularly applies to design features that could be perceived as social cues (e.g., human-like appearance, language style, or personality (c.f., Fogg, 2002)) because they could unintentionally trigger social responses and processes. Moreover, the provision of inappropriate social cues to humanize a chatbot might create unrealistic user expectations and lead to misunderstandings, particularly when these cues overplay the chatbot's actual capabilities (Culley and Madhavan, 2013; Knijnenburg and Willemsen, 2016; Ben Mimoun et al., 2012). Although the provision of social cues in technologies has been generally linked to positive outcomes (e.g., Hess et al., 2009; Qiu and Benbasat, 2009), research also points out that people may respond quite negatively when technologies too closely resemble human beings (MacDorman and Ishiguro, 2006; Mori, 1970). For example, the "uncanny valley" hypothesis states that human-like technologies are perceived as more agreeable up until they become so human that people find their nonhuman imperfections unsettling (MacDorman and Ishiguro, 2006; Mori, 1970).Therefore, it can be argued that chatbots, just as humanoid robots, may reach a point of human-likeness that makes users uncomfortable. Consequently, human-like features that represent social cues need to be designed carefully to limit possible negative outcomes (Candello et al., 2017; Klopfenstein et al., 2017).

Furthermore, our results show the importance of "conversations as the object of design" (Følstad and Brandtzæg, 2017, p. 40). Specifically, we argue that the design of a conversation should not only consider its content but also other factors that are known to influence human-computer interaction (e.g., response time). While chatbots are often touted as "easy to build" from a technical perspective (Moore et al., 2017), we argue that designing the social and human elements of chatbots represents a major challenge for creating natural interactions with these systems. Second, we found that users might evaluate a chatbot's response time differently than, for example, a website's response time. This finding is in line with other researchers (Følstad and Brandtzæg, 2017; Moore et al., 2017) who suggest that designing conversational user interfaces is different from designing graphical user interfaces. The user interface of a chatbot (e.g., a chat history and text box) not only hides the complexity of the underlying technology, but also provides less opportunities to apply established usability principles that have been developed for graphical user interfaces (Følstad and Brandtzæg, 2017; Moore et al., 2017). As there is a growing trend from graphical towards conversational user interfaces, particularly in the interaction with smart machines and devices (Brynjolfsson and McAfee, 2016), more research is needed to understand how the interaction between humans and machines should be designed when natural language is involved.

6.2 Limitations and Future Work

We are aware that our work comes with limitations. First, our experiment consisted of only two conditions: (1) dynamically delayed and (2) near-instant responses. Future research should investigate additional conditions such as static time delays of different lengths (c.f., Holtgraves et al., 2007; Moon, 1999). Thus, extending the experiment could provide further insights into the effect of different delay types. Future research could also combine the investigation of response delays with related design features such as "typing indicators". These indicators are implemented on many messenger platforms and are also increasingly being used by chatbots (Gnewuch et al., 2018; Klopfenstein et al., 2017), without knowing exactly whether and how they affect users' perception of a chatbot. Second, even though our formula for calculating dynamic response delays based on message complexity builds on the established Flesch-Kincaid grade level, the calibration of the formula in our study is based on pretests and may require adjustment to specific contexts. In particular, we acknowledge that message complexity does not only depend on syntactic complexity (e.g., number of sentences, words, and syllables), but also on its context-specific meaning (e.g., computer-mediated negotiation versus answering frequently asked questions). Hence, the calibration and potential extensions of the formula should consider complexity associated with context-specific meaning. Moreover, individual characteristics (e.g., familiarity with chatbots, personality traits) and expectations may also influence the subjective judgement of how "good" or "natural" a delay is and could therefore be included as moderators in future studies. While our results suggest that certain dynamic delays positively impact users' perception of chatbots, future research is needed to better understand how these delays should be optimally calculated. This also offers the opportunity to make use of machine learning algorithms to dynamically adapt delays based on a user's response time during the conversation. Third, we conducted our experiment online and not in a laboratory environment. Because of physical limits of data transmission over the Internet, there was an unavoidable network delay of approximately one second in both conditions. However, this limitation applies to all current chatbots running on instant messaging platforms or websites. Therefore, we argue that our condition in which responses were sent near-instantly (i.e., in one second) comes very close to the minimum response time that is possible with current technology. Furthermore, our results could have been biased by a participant's slow Internet connection or browser speed. Although we controlled for this bias by analyzing the timestamps in the log files of both chatbots, future research could validate our results in a controlled laboratory experiment. This will ensure equal conditions for all participants as well as reduce the minimal response time in the control condition to less than one second when the chatbot is running locally.

7 Conclusion

This paper provides evidence that dynamic response delays positively affect users' perception of customer service chatbots. Specifically, we found that when responses are dynamically delayed, users perceive chatbots as more human-like and more socially present, and are more satisfied with the overall interaction than when responses are sent near-instantly. From a theoretical perspective, we demonstrate that a chatbot's response time represents a social cue that elicits social responses from users. By triggering social scripts and expectations, users perceive chatbots differently when they send near-instant than when they send dynamically delayed responses. Since dynamic response delays can positively shape users' perception of chatbots, they should be accounted for in their design. Although we focus on customer service chatbots, we believe that our findings can be valuable for the design of chatbots in other contexts as well. While more research on the specific characteristics of dynamic response delays is necessary, we believe that our findings provide an important first step towards making human-chatbot interaction more natural. Our formula for dynamically calculating response delays and our design implications can inform practitioners that wish to better understand and design human-chatbot interactions.

References

Appel, J., von der Pütten, A., Krämer, N.C. and Gratch, J. (2012), “Does Humanity Matter? Analyzing the Importance of Social Cues and Perceived Agency of a Computer System for the Emergence of Social Reactions during Human-Computer Interaction”, *Advances in Human-Computer Interaction,* Vol. 2012, pp. 1–10.

Barger, P.B. and Grandey, A.A. (2006), “Service with a smile and encounter satisfaction: Emotional contagion and appraisal mechanisms”, *Academy of Management Journal,* Vol. 49 No. 6, pp. 1229–1238.

Bitner, M.J., Brown, S.W. and Meuter, M.L. (2000), “Technology Infusion in Service Encounters”, *Journal of the Academy of Marketing Science,* Vol. 28 No. 1, pp. 138–149.

Brynjolfsson, E. and McAfee, A. (2016), *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies,* WW Norton & Company, New York.

Candello, H., Pinhanez, C. and Figueiredo, F. (2017), “Typefaces and the Perception of Humanness in Natural Language Chatbots”, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’17),* ACM, Denver, CO, USA, pp. 3476–3487.

Chakrabarti, C. and Luger, G.F. (2015), “Artificial Conversations for Customer Service Chatter Bots: Architecture, Algorithms, and Evaluation Metric”, *Expert Systems with Applications,* Vol. 42 No. 20, pp. 6878–6897.

Crozier, R. (2017), “Lufthansa delays chatbot’s responses to make it more ‘human’”, *iTnews Australia,* available at: https://www.itnews.com.au/news/lufthansa-delays-chatbots-responses-to-make-it-more-human-462643 (accessed 20 December 2017).

Culley, K.E. and Madhavan, P. (2013), “A note of caution regarding anthropomorphism in HCI agents”, *Computers in Human Behavior,* Vol. 29, pp. 577–579.

Cyr, D., Head, M., Larios, H. and Pan, B. (2009), “Exploring Human Images in Website Design: A Multi-Method Approach”, *MIS Quarterly,* Vol. 33 No. 3, pp. 539–566.

Dale, R. (2016), “The return of the chatbots”, *Natural Language Engineering,* Vol. 22 No. 5, pp. 811–817.

Derrick, D.C., Meservy, T.O., Jenkins, J.L., Burgoon, J.K. and Nunamaker, J.F. (2013), “Detecting Deceptive Chat-Based Communication Using Typing Behavior and Message Cues”, *ACM Transactions on Management Information Systems,* Vol. 4 No. 2, pp. 1–21.

van Doorn, J., Mende, M., Noble, S.M., Hulland, J., Ostrom, A.L., Grewal, D. and Petersen, J.A. (2017), “Domo Arigato Mr. Roboto: Emergence of Automated Social Presence in Organizational Frontlines and Customers’ Service Experiences”, *Journal of Service Research,* Vol. 20 No. 1, pp. 43–58.

Faul, F., Erdfelder, E., Lang, A.-G. and Buchner, A. (2007), “G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences”, *Behavior Research Methods,* Vol. 39 No. 2, pp. 175–191.

Fogg, B.J. (2002), “Computers as Persuasive Social Actors”, *Persuasive Technology: Using Computers to Change What We Think and Do,* Morgan Kaufmann Publishers, San Francisco, CA, USA, pp. 89–120.

Følstad, A. and Brandtzæg, P.B. (2017), “Chatbots and the new world of HCI”, *Interactions,* Vol. 24 No. 4, pp. 38–42.

Fraser, N. (1997), “Assessment of Interactive Systems”, in Gibbon, D., Moore, R. and Winski, R. (Eds.), *Handbook on Standards and Resources for Spoken Language Systems,* Mouton de Gruyter, New York, NY, USA, pp. 564–614.

Gartner. (2017), “Top Trends in the Gartner Hype Cycle for Emerging Technologies”, available at: https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/ (accessed 20 December 2017).

Gartner. (2018), “Gartner Says 25 Percent of Customer Service Operations Will Use Virtual Customer Assistants by 2020”, available at: https://www.gartner.com/newsroom/id/3858564 (accessed 11 April 2018).

Gefen, D. and Straub, D.W. (1997), “Gender Differences in the Perception and Use of E-Mail: An Extension to the Technology Acceptance Model”, *MIS Quarterly,* Vol. 21 No. 4, pp. 389–400.

Gefen, D. and Straub, D.W. (2003), “Managing User Trust in B2C e-Services”, *E-Service Journal*, Vol. 2 No. 2, pp. 7–24.

Gefen, D. and Straub, D.W. (2005), “A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example”, *Communications of the Association for Information Systems*, Vol. 16, pp. 91–109.

Gnewuch, U., Morana, S., Heckmann, C. and Maedche, A. (2018), “Designing Conversational Agents for Energy Feedback”, *Proceedings of the 13th International Conference on Design Science Research in Information Systems and Technology (DESRIST 2018)*, Chennai, India.

Gnewuch, U., Morana, S. and Maedche, A. (2017), “Towards Designing Cooperative and Social Conversational Agents for Customer Service”, *Proceedings of the 38th International Conference on Information Systems (ICIS)*, Seoul, South Korea.

Goldberg, L.R. (1992), “The development of markers for the Big-Five factor structure”, *Psychological Assessment*, Vol. 4 No. 1, pp. 26–42.

Hassanein, K. and Head, M. (2007), “Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping”, *International Journal of Human-Computer Studies*, Vol. 65 No. 8, pp. 689–708.

Hess, T., Fuller, M. and Campbell, D. (2009), “Designing Interfaces with Social Presence: Using Vividness and Extraversion to Create Social Recommendation Agents.”, *Journal of the Association for Information Systems*, Vol. 10 No. 12, pp. 889–919.

Ho, S.M., Hancock, J.T., Booth, C. and Liu, X. (2016), “Computer-Mediated Deception: Strategies Revealed by Language-Action Cues in Spontaneous Communication”, *Journal of Management Information Systems*, Vol. 33 No. 2, pp. 393–420.

Holtgraves, T. and Han, T.-L. (2007), “A procedure for studying online conversational processing using a chat bot”, *Behavior Research Methods*, Vol. 39 No. 1, pp. 156–163.

Holtgraves, T., Ross, S.J., Weywadt, C.R. and Han, T.-L. (2007), “Perceiving artificial social agents”, *Computers in Human Behavior*, Vol. 23 No. 5, pp. 2163–2174.

Hoxmeier, J.A. and DiCesare, C. (2000), “System Response Time and User Satisfaction : An Experimental Study of Browser-based Applications”, *Proceedings of the 6th Americas Conference on Information Systems (AMCIS)*, Long Beach, CA, USA, pp. 140–145.

Jenkins, M.-C., Churchill, R., Cox, S. and Smith, D. (2007), “Analysis of User Interaction with Service Oriented Chatbot Systems”, in Jacko, J. (Ed.), *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments*, Springer, Berlin, Heidelberg, pp. 76–83.

Kanda, T., Kamasima, M., Imai, M., Ono, T., Sakamoto, D., Ishiguro, H. and Anzai, Y. (2007), “A humanoid robot that pretends to listen to route guidance from a human”, *Autonomous Robots*, Vol. 22 No. 1, pp. 87–100.

Kang, L., Tan, C.H. and Zhao, L.J. (2013), “The impact of intra-transaction communication on customer purchase behaviour in e-commerce context”, *Proceedings of the 24th Australasian Conference on Information Systems (ACIS)*, Melbourne, Australia, pp. 1–12.

Kang, S.-H. and Gratch, J. (2014), “Exploring users’ social responses to computer counseling interviewers’ behavior”, *Computers in Human Behavior*, Vol. 34, pp. 120–130.

Khawaja, M.A., Chen, F. and Marcus, N. (2010), “Using language complexity to measure cognitive load for adaptive interaction design”, *Proceedings of the 15th International Conference on Intelligent User Interfaces (IUI ’10)*, ACM, pp. 333–336.

Kincaid, J.P., Fishburne, R.P., Rogers, R.L. and Chissom, B.S. (1975), “Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) For Navy Enlisted Personnel”, *Branch Report*, Vol. 8–75.

Klopfenstein, L.C., Delpriori, S., Malatini, S. and Bogliolo, A. (2017), “The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms”, *Proceedings of the 2017 Conference on Designing Interactive Systems*, pp. 555–565.

Knijnenburg, B.P. and Willemsen, M.C. (2016), “Inferring Capabilities of Intelligent Agents from Their External Traits”, *ACM Transactions on Interactive Intelligent Systems*, Vol. 6 No. 4, pp. 1–25.

Lafferty, J.C., Eady, P.M. and Elmers, J. (1974), “The desert survival problem”, *Experimental Learning Methods*, Plymouth, MI, USA.

Larivière, B., Bowen, D., Andreassen, T.W., Kunz, W., Sirianni, N.J., Voss, C., Wünderlich, N. V., et al. (2017), "'Service Encounter 2.0': An Investigation into the Roles of Technology, Employees and Customers", *Journal of Business Research*, Vol. 79, pp. 238–246.

Lennon, C. and Burdick, H. (2004), *The Lexile Framework as an Approach for Reading Measurement and Success*, available at: http://www.lexile.com.

MacDorman, K.F. and Ishiguro, H. (2006), "The uncanny advantage of using androids in cognitive and social science research", *Interaction Studies*, Vol. 7 No. 3, pp. 297–337.

Maedche, A., Morana, S., Schacht, S., Werth, D. and Krumeich, J. (2016), "Advanced User Assistance Systems", *Business & Information Systems Engineering*, Vol. 58 No. 5, pp. 367–370.

McLean, G. and Wilson, A. (2016), "Evolving the online customer experience ... is there a role for online customer support?", *Computers in Human Behavior*, Vol. 60, pp. 602–610.

McTear, M., Callejas, Z. and Griol, D. (2016), *The Conversational Interface: Talking to Smart Devices*, Springer.

Mentzer, G., Cryan, J. and Teclehaimanot, B. (2007), "Two Peas in a Pod? A Comparison of Face-to-Face and Web Based Classrooms", *Journal of Technology and Teacher Education*, Vol. 15 No. 2, pp. 233–246.

Ben Mimoun, M.S., Poncin, I. and Garnier, M. (2012), "Case study-Embodied virtual agents: An analysis on reasons for failure", *Journal of Retailing and Consumer Services*, Vol. 19 No. 6, pp. 605–612.

Moon, Y. (1999), "The effects of physical distance and response latency on persuasion in computer-mediated communication and human-computer communication.", *Journal of Experimental Psychology: Applied*, Vol. 5 No. 4, pp. 379–392.

Moon, Y. (2000), "Intimate Exchanges: Using Computers to Elicit Self-Disclosure From Consumers", *Journal of Consumer Research*, Vol. 26 No. 4, pp. 323–339.

Moon, Y. (2003), "Don't Blame the Computer: When Self-Disclosure Moderates the Self-Serving Bias", *Journal of Consumer Psychology*, Vol. 13 No. 1–2, pp. 125–137.

Moore, R.J., Arar, R., Ren, G.-J. and Szymanski, M.H. (2017), "Conversational UX Design", *Proceedings of the SIGCHI Conference: Extended Abstracts on Human Factors in Computing Systems (CHI '17)*, ACM, Denver, CO, USA, pp. 492–497.

Mori, M. (1970), "The uncanny valley", *Energy*, Vol. 7 No. 4, pp. 33–35.

Muylle, S., Moenaert, R. and Despontin, M. (2004), "The conceptualization and empirical validation of web site user satisfaction", *Information and Management*, Vol. 41 No. 5, pp. 543–560.

Nass, C. and Moon, Y. (2000), "Machines and Mindlessness: Social Responses to Computers", *Journal of Social Issues*, Vol. 56 No. 1, pp. 81–103.

Nass, C., Steuer, J. and Tauber, E.R. (1994), "Computers are social actors", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Boston, MA, USA, pp. 72–78.

Oliver, R.L. (1997), *Satisfaction: A Behavioral Perspective on the Consumer*, McGraw-Hill, New York, NY, USA.

Oracle. (2016), "Can Virtual Experiences Replace Reality?", available at: https://www.oracle.com/webfolder/s/delivery_production/docs/FY16h1/doc35/CXResearchVirtualExperiences.pdf (accessed 14 December 2017).

von der Pütten, A.M., Krämer, N.C., Gratch, J. and Kang, S.-H. (2010), "'It doesn't matter what you are!' Explaining social effects of agents and avatars", *Computers in Human Behavior*, Vol. 26 No. 6, pp. 1641–1650.

Qiu, L. and Benbasat, I. (2009), "Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems", *Journal of Management Information Systems*, Vol. 25 No. 4, pp. 145–182.

Reeves, B. and Nass, C. (1996), *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*, CSLI Publications, Cambridge, UK.

Rushinek, A. and Rushinek, S.F. (1986), "What makes users happy?", *Communications of the ACM*, Vol. 29 No. 7, pp. 594–598.

Sarikaya, R. (2017), "The Technology Behind Personal Digital Assistants: An Overview of the System Architecture and Key Components", *IEEE Signal Processing Magazine*, Vol. 34 No. 1, pp. 67–81.

Schuetzler, R.M., Grimes, G.M., Giboney, J.S. and Buckman, J. (2014), “Facilitating Natural Conversational Agent Interactions: Lessons from a Deception Experiment”, *Proceedings of the 35th International Conference on Information Systems (ICIS ’14)*, Auckland, NZ, pp. 1–16.

Shechtman, N. and Horowitz, L.M. (2003), “Media Inequality in Conversation: How People Behave Differently When Interacting with Computers and People”, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’03)*, ACM, Ft. Lauderdale, FL, USA, pp. 281–288.

Short, J., Williams, E. and Christie, B. (1976), *The Social Psychology of Telecommunications*, London, UK.

Skowron, M., Pirker, H., Rank, S., Paltoglou, G., Jung Hyun, A. and Gobron, S. (2011), “No peanuts! Affective Cues for the Virtual Bartender”, *Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference*, pp. 117–122.

Song, J.H. and Zinkhan, G.M. (2008), “Determinants of Perceived Web Site Interactivity”, *Journal of Marketing*, Vol. 72 No. 2, pp. 99–113.

Taylor, B., Dey, A.K., Siewiorek, D. and Smailagic, A. (2016), “Using Crowd Sourcing to Measure the Effects of System Response Delays on User Engagement”, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’16)*, ACM, San Jose, CA, USA, pp. 4413–4422.

Urbach, N. and Ahlemann, F. (2010), “Structural equation modeling in information systems research using partial least squares”, *JITTA: Journal of Information Technology Theory and Application*, Vol. 11 No. 2, p. 5.

Verhagen, T., van Nes, J., Feldberg, F. and van Dolen, W. (2014), “Virtual Customer Service Agents: Using Social Presence and Personalization to Shape Online Service Encounters”, *Journal of Computer-Mediated Communication*, Vol. 19 No. 3, pp. 529–545.

Vronay, D., Smith, M. and Drucker, S. (1999), “Alternative interfaces for chat”, *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology (UIST ’99)*, ACM, Asheville, NC, USA, pp. 19–26.

Walther, J.B. (2007), “Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition”, *Computers in Human Behavior*, Vol. 23 No. 5, pp. 2538–2557.

Wang, W., Qiu, L., Kim, D. and Benbasat, I. (2016), “Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust”, *Decision Support Systems*, Vol. 86, pp. 48–60.

Weizenbaum, J. (1966), “ELIZA - A Computer Program for the Study of Natural Language Communication between Man and Machine”, *Communications of the ACM*, Vol. 9 No. 1, pp. 36–45.

Wixom, B.H. and Todd, P.A. (2005), “A theoretical integration of user satisfaction and technology acceptance”, *Information Systems Research*, Vol. 16 No. 1, pp. 85–102.

Xiao, B. and Benbasat, I. (2007), “E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact”, *MIS Quarterly*, Vol. 31 No. 1, pp. 137–209.