Karlsruhe Institute of Technology

Institute of Information Systems and Marketing (IISM)
Fritz-Erler-Strasse 23
76133 Karlsruhe - Germany
http://iism.kit.edu

Karlsruhe Service Research Institute (KSRI)
Kaiserstraße 89
76133 Karlsruhe – Germany
http://ksri.kit.edu

"The Chatbot is typing …" – The Role of Typing Indicators in Human-Chatbot Interaction

Ulrich Gnewuch
Karlsruhe Institute of Technology, Germany
ulrich.gnewuch@kit.edu

Stefan Morana
Karlsruhe Institute of Technology, Germany
stefan.morana@kit.edu

Marc T. P. Adam
The University of Newcastle, Australia
marc.adam@newcastle.edu.au

Alexander Maedche
Karlsruhe Institute of Technology, Germany
alexander.maedche@kit.edu

ABSTRACT

Chatbots have attracted considerable interest in recent years. A key design challenge to increase their adoption is to make the interaction with them feel natural and human-like. Therefore, it is suggested to incorporate social cues in the chatbot design. Drawing on the Computers are Social Actors paradigm and the "uncanny valley" hypothesis, we study the effect of one specific social cue (i.e., typing indicators) on social presence of chatbots. In an online experiment, we investigate the effect of two specific designs of typing indicators. Our preliminary results indicate that graphical typing indicators increase social presence of chatbots, but only for novice users. Therefore, our results suggest that the relationship between typing indicators and perceived social presence of chatbots depends on the design of these indicators and user's prior experience. We contribute with empirical insights and design knowledge that support researchers and practitioners in understanding and designing more natural human-chatbot interactions.

Keywords

Chatbot, conversational agent, typing indicator, social presence, computers are social actors, online experiment.

INTRODUCTION

Text-based conversational agents, commonly referred to as chatbots, have received a lot of attention in recent years. They are designed to interact with humans using natural language and are increasingly used on messaging platforms and websites (Dale 2016). As advances in artificial intelligence continue, many organizations are beginning to implement chatbots to automate customer service and reduce costs (Gartner 2018). Despite the large interest in chatbots, their adoption and use is growing much slower than expected (Liedtke 2017). One key barrier to the adoption and use of chatbots is that the interaction with them often does not feel natural and human-like (Schuetzler et al. 2014). However, established design principles for creating chatbot interactions that feel natural to the user are scarce (McTear 2017). Previous research on the design of websites and recommendation agents suggests that incorporating social cues, such as natural language and human-like appearance, makes the interaction more natural and enhances users' perceived social presence (e.g., Qiu and Benbasat 2009).

Much of this research builds on the Computers are Social Actors (CASA) paradigm to explain why users apply social rules and expectations in their interaction with information systems (IS) that incorporate social cues, such as natural language or human-like appearance (Nass et al. 1994). Many studies have found that social cues can positively affect users' perceived social presence, trust, enjoyment, and usage intentions (e.g., Qiu and Benbasat 2009; Wakefield et al. 2011). However, it has also been shown that social cues may backfire, particularly when they irritate users or overplay the system's actual capabilities (Louwerse et al. 2005). In the context of chatbots, a design feature that has been used by both researchers and practitioners to make chatbot interactions appear more natural and familiar to the user is the so-called *typing indicator* (Appel et al. 2012; Klopfenstein et al. 2017). This typing indicator (e.g., three animated dots or the message "[Person X] is typing") was originally developed for text-based computer-mediated communication (CMC) systems to create awareness that the other person is typing in order to support turn-taking (Auerbach 2014). Today, most messaging applications have implemented different forms of these indicators. Consequently, it is suggested to incorporate them in the design of chatbots in order to make the interaction with them feel more like interacting with a real human being (Appel et al. 2012).

Existing research on typing indicators indicates that they may also serve as a social cue as they increase the feeling of being closer to another person (Shin et al. 2018) and are used to make the interaction appear more natural (Appel et al. 2012). However, due to the fact that a chatbot is a machine, which does not type responses on a keyboard, "faking human responses" using these indicators may adversely affect users' perceptions of chatbots (Klopfenstein et al. 2017, p. 559). To the best of our knowledge, little empirical research has examined the role of typing indicators in

human-chatbot interaction. Moreover, there is a lack of knowledge on whether different typing indicator designs influence user perceptions of chatbots differently. To address this gap, we focus on the concept of social presence, which has been identified as a key factor in the design of IS in online environments. Moreover, it is an important driver of trust, enjoyment, and usage intentions (Cyr et al. 2007), particularly in the domain of customer service in which chatbots are increasingly used (Gartner 2018). Thus, we address the following research question:

How do typing indicators influence users' perceived social presence of chatbots in customer service?

To address our research question, we conducted a three-condition, between-subjects online experiment to investigate the effect of typing indicators in human-chatbot interaction. Drawing on the CASA paradigm (Nass et al. 1994) and the uncanny valley hypothesis (Mori 1970), we examined how the existence and two different designs of typing indicators influence users' perceived social presence of chatbots. The results of our preliminary analysis show that typing indicators positively influence novice users' perceived social presence of a chatbot, while they make no difference for experienced users. Moreover, there is a significant difference between the two designs of typing indicators for novice users. By investigating the design and outcomes of a specific social cue (i.e., typing indicators), our research contributes to existing literature on the design of text-based conversational agents.

RELATED WORK AND THEORETICAL FOUNDATION

Chatbots are Social Actors

Previous research has used a variety of terms to describe systems that allow users to interact with them using natural language (e.g., conversational agent, chatbot, or virtual assistant). It has been shown that humans respond socially to computers exhibiting human-like characteristics (Nass et al. 1994). The CASA paradigm posits that, when users are confronted with these so-called social cues from a computer (e.g., natural language or human-like appearance), they automatically apply social rules and expectations in their interaction with it (Nass et al. 1994). Even rudimentary social cues are sufficient to generate a wide range of social responses from users. Following the CASA paradigm, many studies have investigated how users react to various social cues from computers, robots, and other technologies. Examples in IS research include recommendation agents (e.g., Qiu and Benbasat 2009) and websites (e.g., Wakefield et al. 2011). A few studies were also conducted in the context of chatbots to examine the effect of visual (e.g., Appel et al. 2012) or verbal cues (e.g., Schuetzler et al. 2014). Across these studies, social cues have been found to positively affect user perceptions of chatbots (e.g., social presence, trust, or engagement). However, the provision of inappropriate social cues to overly humanize chatbots may also backfire when they too closely resemble human beings (Gnewuch et al. 2017; Louwerse et al. 2005). 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 (Mori 1970). Therefore, design features, such as typing indicators, that represent social cues need to be designed carefully to limit possible negative impacts (Fogg 2002).

Typing Indicators and Turn-Taking

Turn-taking is a fundamental mechanism for the organization of turns in human-human interaction (Sacks et al. 1974). It basically describes the rules by which participants in a conversation manage who speaks when and for how long in order to avoid overlaps and minimize silence between turns (Sacks et al. 1974). In face-to-face communication, turn-taking is facilitated by a multitude of social cues such as gesticulations, eye contact, and facial expressions (Wiemann and Knapp 1975). Since these cues are missing in text-based CMC, such as instant messaging, users are usually not aware of another person's turn because their messages do not appear on the users' screen until the person typing them hits the return key. To prevent overlaps and increase turn-taking awareness, developers of one of the first messaging applications, Microsoft's MSN Messenger, invented the typing indicator (Auerbach 2014). Once a user started typing a message, this indicator displayed "[Person X] is typing" on the other person's screen, which facilitated turn-taking and substituted the missing social cues (e.g., facial expressions) used in human-human interaction.

Nowadays, typing indicators have been implemented in all major messaging applications. Looking at the most widely-used messaging applications, primarily two different types of designs can be identified: (1) a graphical typing indicator (*3 dots*) and (2) a textual typing indicator (*Typing*). Both typing indicators share some similarities with filler interfaces found on websites (e.g., loading screens, progress bars, or "Please wait" messages), which have been found to influence users' waiting experience and perceived waiting time by directing attention away from the wait (Lee et al. 2012).

As chatbots are often implemented in the same text-based CMC channels (e.g., Facebook Messenger), they also increasingly use typing indicators (Klopfenstein et al. 2017). Although practitioners frequently highlight their potential benefits, empirical research on their role in human-chatbot interaction is scarce. More specifically, it is not clear whether typing indicators positively affect user perceptions of chatbots (i.e., increase social presence) and whether different typing indicator designs influence these perceptions differently.

RESEARCH MODEL

Although typing indicators are implemented in most messaging applications and also used by chatbots (Klopfenstein et al. 2017), little empirical research has been conducted to understand how they impact human-chatbot interaction. Therefore, we decided to adopt a two-step

approach to derive our hypotheses and develop our research model (c.f., Lee et al. 2012). Before exploring different typing indicator designs, it needs to be examined whether the existence of the typing indicator itself influences users' perceived social presence of a chatbot. Drawing on the CASA paradigm and the uncanny valley hypothesis, we formulate four hypotheses on the effect of the existence and two designs of typing indicators.

Social presence has been identified as a key factor in the design of IS in online environments and an important driver of trust, enjoyment, and usage intentions (Cyr et al. 2007). The concept of social presence is used to understand how feelings of warmth, sociability, and human contact can be created without actual human contact (Gefen and Straub 2004). Previous research has shown that many social cues incorporated in websites, recommendations agents, and other technologies, create perceptions of social presence (e.g., Qiu and Benbasat 2009). According to the CASA paradigm, these perceptions are the result of an unconscious process, in which users respond to technologies as though they were human, despite knowing that they are interacting with a machine (Nass et al. 1994). Drawing on CASA, we argue that typing indicators serve as a social cue in the interaction with a chatbot. More specifically, we believe that when users interact with a chatbot with a typing indicator, their perceptions of the chatbot are shaped by their social expectations from interacting with other human beings (i.e., using messaging applications such as WhatsApp). Although they know that the chatbot does not type responses on a keyboard, they will subconsciously apply the social rules practiced in their daily life, which in turn generates perceptions of social presence similar to those that would be generated if the user were interacting with another human. Therefore, we propose that in human-chatbot interaction, the mere existence of a typing indicator, regardless of its design (i.e., graphical/3 Dots or textual/Typing), will lead to higher levels of perceived social presence. Hence, we argue that:

$H_{1a,b}$: *Users exhibit higher levels of perceived social presence when interacting with a chatbot with **(a)** a graphical typing indicator and **(b)** a textual typing indicator, compared to the same chatbot without a typing indicator.*

As mentioned above, there are primarily two different designs of typing indicators used in major messaging applications: a graphical typing indicator (*3 Dots*) and a textual typing indicator (*Typing*). Although we argue that the existence of typing indicators itself influences users' perceived social presence of chatbots, we assume that there are differences in user perceptions between the two identified designs. However, based on theory, it is not clear how these differences will manifest themselves. Therefore, we formulate two contrasting hypotheses on the effect of different typing indicator designs (graphical vs. textual) on social presence.

First, the graphical typing indicator has a rather functional design, similar to loading screens or progress bars of websites. Thus, while it indicates that "something" is happening, it does not explicitly state what happens. Since users are familiar with the graphical typing indicator from their use of messaging applications, they subconsciously associate a human action (i.e., typing) with it. According to the CASA paradigm, even such minimal social cues can trigger a wide range of social responses (Nass et al. 1994), resulting in increased social presence as compared to no typing indicators. However, the textual typing indicator explicitly states that the chatbot is "typing" while users are waiting for a response. Therefore, it can be argued that this explicit message is a more salient social cue because it imitates human action. Research based on CASA has found that when users are exposed to more or stronger social cues, their social responses become stronger as well (e.g., von der Pütten et al. 2010). Therefore, we propose that a textual typing indicator generates a higher level of perceived social presence than a graphical typing indicator:

H_{2a}: *Users exhibit higher levels of perceived social presence when interacting with a chatbot with a textual typing indicator compared to a graphical typing indicator.*

However, previous research also points out that "turn[ing] up the volume on the social element" of technologies can have some undesirable side effects (Fogg 2002, p. 114) and may even backfire (Louwerse et al. 2005). For example, Groom et al. (2009) found that users were less comfortable interacting with a very realistic and human-like conversational agent than with an agent with lower realism. The uncanny valley hypothesis states that when technologies become more human-like (e.g., by incorporating stronger social cues), they are perceived as more agreeable and social, until a point beyond which the reaction is reversed (Mori 1970). Therefore, we argue that textual typing indicators explicitly stating that the chatbot is "typing" approach the edge of the uncanny valley. Instead of increasing users' perceptions of social presence, these indicators might backfire because users feel like they are "faking" human action. Thus, we propose that a textual typing indicator generates a *lower* level of social presence than a graphical typing indicator:

H_{2b}: *Users exhibit lower levels of perceived social presence when interacting with a chatbot with a textual typing indicator compared to a graphical typing indicator.*

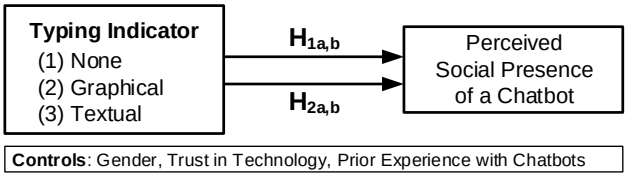

Figure 1. Research Model

METHOD

To test our research model, we conducted an online experiment, in which participants interacted with a chatbot in a customer service context. In the experiment, participants were provided with a hypothetical scenario of using the chatbot to search for a cheaper mobile phone plan that fit their individual needs. After they received a recommendation by the chatbot and ended their conversation, they were asked to complete a survey about

their perceptions of the chatbot and their interaction with it. Overall, 256 student subjects participated in the experiment. They were randomly assigned to one of the experimental conditions.

Treatment Configuration and Chatbot Scenario

The online experiment employed a between-subjects design with three conditions (*typing indicator*: none, graphical, textual) to avoid potential carry-over effects. In all conditions, participants were told that they were interacting with a chatbot. In the control condition (CTRL) condition, participants interacted with a chatbot that did not display a typing indicator before sending a message. In the first treatment condition (3DOTS), the chatbot displayed a graphical typing indicator before sending a message (see Table 1). In the second treatment condition (TYPING), the chatbot displayed a textual typing indicator before sending a message. In each condition, the chatbots' responses were delayed by 2.3 seconds to ensure that participants were sufficiently exposed to the typing indicators. Since response time can also serve as a social cue (Gnewuch et al. 2018), responses of the chatbot without typing indicators were also delayed to keep the chatbots' response time identical and hence comparable across all conditions.

Condition	Typing Indicator	Description
CTRL	None	-
3DOTS	Graphical	Three animated dots fading in one after another and then fading out, placed above the user input field.
TYPING	Textual 	A textual status message "typing…" placed above the user input field.

Table 1. Treatment Configuration

Measures and Manipulation Check

All measures used in the survey were adapted from established scales. Social presence was assessed using the items from Gefen and Straub (1997). Moreover, we measured control variables, such as disposition to trust technology (Lankton et al. 2015) and prior experience with chatbots (ranging from "never" to "daily"), as well as collected demographic information (e.g., age, gender).

Condition	Social Presence		
	Mean	SD	SE
CTRL (n=63)	3.448	1.442	0.182
3DOTS (n=63)	3.902	1.451	0.183
TYPING (n=63)	3.663	1.408	0.177
SD = standard deviation | SE = standard error			

Table 2. Descriptive Statistics

We included a manipulation check to test whether the typing indicator manipulation was successful. Participants were asked to rate whether the chatbot indicated that a response was being prepared/generated, using a 7-point Likert scale (1 = "strongly disagree"; 7 = "strongly agree). A one-way ANOVA showed a significant influence of the experimental conditions on perceived indication that a response was prepared/generated ($F(2, 186)=439.3$, $p<.001$).

PRELIMINARY RESULTS

In our preliminary analysis conducted so far, we estimated two regression models to analyze the effect of the treatment conditions on perceived social presence and the interaction between users' prior experience with chatbots and the treatment condition along with two user-related factors (i.e., gender, disposition to trust technology). We differentiated between novice users who have never used a chatbot and experienced users who use a chatbot at least 1-2 times per year. Our preliminary analysis shows mixed results. For novice users, we find that the existence of a graphical typing indicator (i.e., three animated dots) significantly impacts their perceived social presence of a chatbot. However, this relationship is not significant for experienced users. Contrary to our expectations, we did not observe a significant effect for the textual typing indicators (i.e., "typing…"), neither for novice nor experienced users. However, novice users perceived a chatbot with a graphical typing indicator significantly more socially present than a chatbot with a textual typing indicator. Our results indicate that the relationship between typing indicators and user perceptions of chatbots is more complex than assumed. Moreover, individual user characteristics, such as experience with chatbots, seem to play an important role in this relationship.

CONCLUSION AND EXPECTED CONTRIBUTIONS

This paper provides first insights on the role of different typing indicators in human-chatbot interaction. In particular, our findings suggest that the relationship between typing indicators and perceived social presence of chatbots depends on the design of these indicators and user's prior experience with chatbots. Therefore, our findings contribute to existing literature on the design and evaluation of text-based conversational agents. In our future research, we intend to continue our data analysis and plan to supplement our survey responses with findings from a qualitative analysis of the chatbot conversations.

REFERENCES

1. 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* (2012), pp. 1–10.
2. Auerbach, D. 2014. "I Built That 'So-and-So Is Typing' Feature in Chat." (http://www.slate.com/articles/technology/bitwise/2014/02/typing_indicator_in_chat_i_built_it_and_i_m_not_sorry.html, accessed August 8, 2018).
3. Cyr, D., Hassanein, K., Head, M., and Ivanov, A. 2007. "The Role of Social Presence in Establishing Loyalty in

E-Service Environments," *Interacting with Computers* (19:1), pp. 43–56.

4. Dale, R. 2016. "The Return of the Chatbots," *Natural Language Engineering* (22:05), pp. 811–817.
5. Fogg, B. J. 2002. "Computers as Persuasive Social Actors," in *Persuasive Technology: Using Computers to Change What We Think and Do*, San Francisco, CA, USA: Morgan Kaufmann Publishers, pp. 89–120.
6. Gartner. 2018. "Gartner Says 25 Percent of Customer Service Operations Will Use Virtual Customer Assistants by 2020." (https://www.gartner.com/newsroom/id/3858564, accessed September 11, 2018).
7. 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* (21:4), pp. 389–400.
8. Gefen, D., and Straub, D. W. 2004. "Consumer Trust in B2C E-Commerce and the Importance of Social Presence: Experiments in e-Products and e-Services," *Omega* (32:6), pp. 407–424.
9. Gnewuch, U., Morana, S., Adam, M. T. P., and Maedche, A. 2018. "Faster Is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction," in *Proceedings of the European Conference on Information Systems (ECIS)*, Portsmouth, UK.
10. Gnewuch, U., Morana, S., and Maedche, A. 2017. "Towards Designing Cooperative and Social Conversational Agents for Customer Service," in *Proceedings of the 38th International Conference on Information Systems (ICIS)*, Seoul, South Korea.
11. Groom, V., Nass, C., Chen, T., Nielsen, A., Scarborough, J. K., and Robles, E. 2009. "Evaluating the Effects of Behavioral Realism in Embodied Agents," *International Journal of Human-Computer Studies* (67:10), pp. 842–849.
12. Klopfenstein, L. C., Delpriori, S., Malatini, S., and Bogliolo, A. 2017. "The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms," in *Proceedings of the 2017 Conference on Designing Interactive Systems*, pp. 555–565.
13. Lankton, N. K., Mcknight, D. H., and Tripp, J. 2015. "Technology, Humanness, and Trust: Rethinking Trust in Technology," *Journal of the Association for Information Systems* (16:10), pp. 880–918.
14. Lee, Y., Chen, A., and Ilie, V. 2012. "Can Online Wait Be Managed? The Effect of Filler Interfaces and Presentation Modes on Perceived Waiting Time Online.," *Management Information Systems Quarterly* (36:2), pp. 365–394.
15. Liedtke, T. 2017. "Facebook's Perfect, Impossible Chatbot," *MIT Technology Review*. (https://www.technologyreview.com/s/604117/facebooks-perfect-impossible-chatbot/, accessed August 9, 2018).
16. Louwerse, M. M., Graesser, A. C., Lu, S., and Mitchell, H. H. 2005. "Social Cues in Animated Conversational Agents," *Applied Cognitive Psychology* (19:6), pp. 1–12.
17. McTear, M. F. 2017. "The Rise of the Conversational Interface: A New Kid on the Block?," in *Future and Emerging Trends in Language Technology. Machine Learning and Big Data. FETLT 2016*, pp. 38–49.
18. Mori, M. 1970. "The Uncanny Valley," *Energy* (7:4), pp. 33–35.
19. Nass, C., Steuer, J., and Tauber, E. R. 1994. "Computers Are Social Actors," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Boston, MA, USA, pp. 72–78.
20. 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* (26:6), pp. 1641–1650.
21. Qiu, L., and Benbasat, I. 2009. "Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems," *Journal of Management Information Systems* (25:4), pp. 145–182.
22. Sacks, H., Schegloff, E. A., and Jefferson, G. 1974. "A Simplest Systematics for the Organization of Turn-Taking for Conversation," *Language* (50:4), p. 696.
23. Schuetzler, R. M., Grimes, G. M., Giboney, J. S., and Buckman, J. 2014. "Facilitating Natural Conversational Agent Interactions: Lessons from a Deception Experiment," in *Proceedings of the 35th International Conference on Information Systems (ICIS '14)*, Auckland, NZ, pp. 1–16.
24. Shin, I.-G., Seok, J.-M., and Lim, Y.-K. 2018. "Too Close and Crowded: Understanding Stress on Mobile Instant Messengers Based on Proxemics," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, Montreal, Canada, pp. 1–12.
25. Wakefield, R. L., Wakefield, K. L., Baker, J., and Wang, L. C. 2011. "How Website Socialness Leads to Website Use," *European Journal of Information Systems* (20:1), pp. 118–132.
26. Wiemann, J. M., and Knapp, M. L. 1975. "Turn-taking in Conversations," *Journal of Communication* (25:2), pp. 75–92.