

## **HUMAN OR BOT?**

### **THE HUMANIZING POWER OF (CORRECTED) ERRORS**

#### **ABSTRACT**

Today more than ever before, online text-based interactions have become a common means of communication between consumers and companies. The current research investigates how one common feature in text communication—typographical errors (“typos”)—can humanize a communicator (e.g., a customer service agent). Across six experiments ( $N = 2,727$ ) that used ambiguously humanlike conversational counterparts (i.e., customer service agents who were either chatbots or real humans), participants perceived agents who made and subsequently corrected a typo to be more human than agents who made no typos or made but did not correct a typo. Moreover, perceiving an agent as more human led participants to form a more favorable perception of the agent (e.g., as being warmer and more helpful) and to be more willing to reward and engage with the agent. These findings provide novel insights into how conversational features influence customers’ perceptions of online agents, with implications for the agents’ corresponding brands. The current research also provides theoretical implications for anthropomorphism and human-computer interaction, as well as practical and ethical considerations for companies wishing to humanize their online customer service agents.

*Keywords:* humanization, chatbot, artificial intelligence, error, communication, customer service

## HUMAN OR BOT?

### THE HUMANIZING POWER OF (CORRECTED) ERRORS

On July 5, 1993, the *New Yorker* published a cartoon by Peter Steiner that quickly became its most reproduced single-panel cartoon in the 20th century, and perhaps one of the most iconic cartoons about the Internet in history: A dog seated in front of a computer turns to its canine companion and says, “*On the Internet, nobody knows you’re a dog.*” (Fleishman 2000). Although this quote perfectly captured the novel anonymity of the early Internet, the prevalence of chatbots and automated messaging nowadays demands a revision to this adage: *On the Internet, nobody knows you’re a bot.*

Chatbots are constantly evolving to become more humanlike conversation partners (Gottsegen 2016; Metz & Collins 2018). Indeed, recently developed chatbots, such as those powered by large language models (e.g., ChatGPT), have achieved impressive abilities to imitate human speech patterns. At the same time, text-based communication inherently lacks the “richness” of other types of human interactions, meaning that consumers have reduced capacity to receive information simultaneously via nonverbal and individuating cues, to assess their chat partners’ mental capacities and intentions, and to experience synchrony with their chat partners (Chafe 1982; Kiesler, Siegel, and McGuire 1984; Kruger et al. 2005; Pinker and Bloom 1990; Schroeder and Epley 2015, 2016; Schroeder, Kardas, and Epley 2017). As a result, Internet users frequently struggle to recognize whether they are communicating with a human or a machine (Adams 2018). Identifying whether a Twitter user, a customer service agent, or a YouTube commenter is a bot or a bona fide human is engaging billions of Internet users in an everyday version of the Turing Test (Brown 2017; Turing 1950). Moreover, companies and consumers are

likely to have different priorities regarding the use of chatbots: While automating consumer service via chatbots promises efficiency and scalability and can therefore yield considerable financial returns for companies (Dal Porto 2017; Frankel 2020; Reddy 2017), consumers may prefer to receive service from human agents who can empathize with them, understand their issues, and potentially help them more effectively (LivePerson 2019; Press 2019).

In the face of such pervasive ambiguity in text-based communications, which cues do consumers rely on to determine whether they are interacting with a human or a bot? This question is particularly important because consumers' perception of an agent's humanness may impact how they perceive and interact with their agent as well as their satisfaction with the agent, which has meaningful consequences for the company.

Whereas prior research has examined how the aesthetics of text (like typeface) or the content of what is expressed (like coherency of the language) can influence perceptions of humanness, it has failed to examine the central way in which written communication actually unfolds in real time—how writing is generated—affects perceived humanness. This paper seeks to fill this gap by investigating a common yet overlooked factor in how writing occurs, specifically whether the writer makes, and subsequently corrects, typographical errors (“typos”). Across six experiments using a multimethod approach, including both message displays and real-time interactions across different contexts, we demonstrate how the correction of typos can lead consumers to perceive a service agent as more human. Furthermore, we examine potential downstream consequences of this humanization process, including how it impacts consumers' perceptions of the agent, their intention to collaborate, and even perceptions of the company using the service agent.

## *Perceiving Humanness*

While *being* human is a biological fact, *perceiving* humanness is a psychological process. Prior work has focused on three ways to influence how people perceive humanness in other agents. First, several nonverbal and verbal cues that tend to be closely associated with people, such as their appearance, voice, and identity, have been shown to impact perception of humanness in nonhuman agents and actual humans. For instance, people tend to attribute more humanlike characteristics or mental capacities to robots and avatars that project more humanlike facial expressions, body configurations, or movements, and those that speak with a humanlike voice (e.g., Heider and Simmel 1944; Looser and Wheatley 2010; Morewedge, Preston, and Wegner 2007; Nass and Brave 2005; Schroeder and Epley 2016; Zhao, Phillips, and Malle 2019). They also attribute more humanness to a communicator with demographic or identity information, such as a gender, name, or nationality (Araujo 2018; Kuchenbrandt et al. 2013; Nass & Moon 2000; Waytz, Heafner, and Epley 2014). However, most of these cues are either absent from text-based communications or are not inherent to written language.

Second, people can infer a communicator's mental capacities from the semantic content of communication—that is, through *what* is expressed. This includes both the extent to which a communicator expresses coherent, sophisticated, and relevant thought (Lortie and Guitton 2011; Christian 2012), as well as the linguistic features and social desirability of what they express (McCoy and Ullman 2018). For instance, when guessing whether self-presentational materials are generated by human or AI, people tend to associate first-person pronouns, spontaneous self-expression, or family topics with humans (Jakesch, Hancock, and Naaman, 2022).

Third, people sometimes infer a human mind by considering the style in which content is communicated. For instance, textual paralinguistics—written manifestations of audible, tactile,

and visual elements that mimic nonverbal cues in face-to-face interaction (e.g., exclamation points, emojis, handwritten-like typefaces, and vocalizations)—have been shown to humanize communicators or products (Candello, Pinhanez, and Figueiredo 2017; Luangrath, Peck, and Barger 2017; Schroll, Schnurr, and Grewal 2018). From adding multiple exclamation points to adopting handwritten-like typefaces, this work has predominantly focused on how static text — the final product of a communication effort — influences perception of humanness toward the communicator.

What is missing in this prior research is an understanding of how the very process of generating text communication itself — via the act of writing — can convey humanness in particular. Indeed, most written conversations with customer agents are now synchronous and instant (e.g., online chat) rather than asynchronous (e.g., email). This provides unprecedented access into the mind of the communicator, because consumers can watch each line of conversation unfold and get insight into how the communicator’s thoughts are getting translated into language, much as they naturally do in a spoken conversation. The rise of synchronous AI-based conversations thus opens a new and important research space of investigating how the *process* of writing can reveal (or conceal) humanness. For example, whether the communicator repeats themselves, presses backspace in revising their language, changes their mind, and so on are all aspects of the process of writing that may convey humanness. The current paper makes inroads on this new set of research questions by studying a typical component of synchronous writing, which is whether a communicator makes, and potentially corrects, typographical errors. We selected this feature because it is common in writing and has precedent in prior research of being associated with humanness in paradoxical ways.

## *To Err is Human? To Correct, Divine*

Common wisdom suggests that “to err is human,” yet mistakes are often associated with “stupid” and ignorant AI systems not seen as measuring up to human standards. Is making mistakes seen as an inevitable part of being human, or does making mistakes lead to negative, dehumanizing perceptions? Examining how these seemingly contradictory intuitions are manifested in consumers’ perception of online chat agents promises both theoretical clarity and practical relevance.

On the one hand, it is possible that errors will be humanizing and create positive social perceptions. For instance, previous research findings (dubbed the “Pratfall Effect”) have shown that making mistakes can, under certain conditions, increase a person’s likeability (e.g., Aronson, Willerman, and Floyd 1966; Helmreich, Aronson, and LeFan 1970). In one experiment (Aronson et al. 1966), a contestant who performed exceptionally well on a difficult quiz but then later made a blunder (i.e., spilling a cup of coffee) was rated as more likable by observers than the same contestant who made no blunder. Moreover, in the specific case of typos, hitting a wrong button on a keyboard when typing quickly seems to be a common human experience that most people can relate to (Norman, 2002). Capitalizing on this intuition, some programmers have inserted typos in an attempt to make their chatbots appear more humanlike in the annual Turing test (i.e., the Loebner Prize; Christian 2011), and social psychologists have even embraced typos as a tool to help them create written materials allegedly generated by another person (Pierce et al. 2013; Wilson et al. 2005).

However, despite its intuitive appeal, the potentially humanizing effect of typos lacks empirical support; to our knowledge, no research has ever documented this effect in a controlled setting. In fact, some research has even found typos to be *dehumanizing* and to lead to negative

social perceptions. For example, in one set of studies, people perceived a communicator who made typos to be less socially attractive and less intelligent than a communicator who did not make typos (Westerman, Cross, and Lindmark 2018). In other studies, participants rated individuals who produced a large volume of spelling and grammatical errors in emails or essays as less competent, less trustworthy, and as having weaker mental capacities, which is a dehumanizing attribution (Figueredo and Varnhagen 2005; Kreiner et al. 2002; Lea and Spears 1992; Vignovic and Thompson 2010). Moreover, most people can also relate to the experience of attempting to converse with a chatbot yet only getting disappointed by their erroneous responses which a sensible human would never produce. Evidently, errors do not always humanize a communicator.

In light of these seemingly contradictory findings, we contend that an error is neither inherently humanizing nor dehumanizing, but what happens *after* the error is key. In fact, both humans and machines make errors: Machines make mistakes due to limited algorithmic capacity, whereas people make mistakes due to limitations of the human mind or body (Carlson & Murphy 2005; Tian and Oviatt 2021). However, in experiments documenting the negative impact of errors, the errors were presented as part of a final product—such as an email or an essay—and were never corrected (e.g., Figueredo and Varnhagen 2005; Lea and Spears 1992; Vignovic and Thompson 2010); by contrast, studies showing positive effects of errors almost always presented errors within a stream of actions, wherein the actors either corrected or at least acknowledged their blunders as the interaction unfolded (e.g., Aronson et al., 1966; Mirnig et al., 2017). Inspired by this critical difference, we propose that written typos during a text-based interaction may humanize a communicator when they are *subsequently corrected* by the communicator.

We believe that correcting an error is essential for appearing human for at least two reasons. First, correcting one's mistake signals that the communicator is actively monitoring, assessing, and reacting to each moment of the ongoing conversation. Even though the initial occurrence of an error may result from an absent or inattentive mind, correcting one's error shows that the mind is engaged and intentional. Second, correcting one's mistake also signals that one cares about how one is being perceived. People are often aware that others may form negative impressions of those who make typos, and they prefer to correct their own errors rather than be corrected by others (Schegloff, Jefferson, and Sacks 1977). Therefore, communicators who correct their own errors signal that they have social awareness and a desire to fix their errors so as to avoid being perceived as mentally "inferior." By contrast, people tend to believe that machines lack the capacity for such self-awareness or conscious intent (Gray, Gray, and Wegner 2007) and do not care about how they will be perceived by others (Boden 2018).

Drawing on the above reasoning, we hypothesize the following:

*H1a-b: People will perceive an online chat agent who makes and subsequently corrects their typographical errors as more human compared to an agent who does not make errors (H1a) and compared to an agent who makes but does not correct their errors (H1b).*

### ***Consequences of Perceiving Humanness***

Humanizing an interaction partner can lead to a multitude of downstream consequences (for reviews, see Epley and Waytz 2010; Złotowski et al. 2015), but little is known about how making and correcting typos in text-based communication affects consumers' perceptions of the communicator. In this research, we primarily focus on consumers' perceptions of the warmth and



helpfulness of an ambiguous customer service agent because these metrics are critical for successful customer service experiences (e.g., Güntürkün, Haumann and Mikolon 2020).

*Perceived warmth.* Warmth is one of the primary dimensions of social evaluation toward interaction partners (Fiske, Cuddy, and Glick 2007). Given that online communication is often experienced as cold and impersonal, and that warmth has been shown to influence consumer satisfaction (Güntürkün, Haumann, and Mikolon 2020; Smith, Martinez, and Sabat 2016), it seems particularly important to investigate how a communicator's text messages can influence how warm they appear.<sup>1</sup> According to people's lay beliefs, machines lack the capacity to experience emotions, whereas humans have the full range of emotional capabilities, such as being able to feel joy, pleasure, hunger, and love (Gray et al. 2007; Waytz and Norton 2014; Weisman, Dweck, and Markman 2017). As a result, whereas machine(-like) interaction partners are often seen as cold, detached, and impersonal, human(-like) partners are perceived as conveying qualities linked to interpersonal warmth, such as being responsive, sincere, sociable, good-natured, and likeable (Broadbent et al. 2013; Haslam 2006; Kim, Schmitt, and Thalmann 2019; Reeves, Hancock, and Liu, 2020). Hence, actions that increase a communicator's perceived humanness should increase their perceived warmth as well. As such, we hypothesize the following:

*H2a-b: People will perceive a chat agent who makes and corrects typographical errors to be warmer (H2a) than an agent who does not make errors (or makes but does not correct errors). This effect will be mediated by perceived humanness (H2b).*

---

<sup>1</sup> To be thorough in assessing consumers' interpersonal impressions, we also examine how typos influence perception of the agent's overall competence, which is the other primary dimension of social evaluation along with warmth (Fiske et al., 2007). We had no specific hypotheses regarding how corrected typos might influence perception of competence, but note that some prior research has suggested that a large number of uncorrected typos are generally associated with lack of education and mental capacity (e.g., Lea and Spears 1992). See results and further discussion on this measure in the studies.

***Perceived helpfulness.*** Nowadays, many companies provide online chat options to customers who want to obtain information or assistance (Adam, Wessel and Benlian 2021). However, communicating with customers over text can be challenging, because it takes a high level of social and communicative skills—skills that typically only another human would have—to understand each customer’s unique situation and identify effective solutions to help them (Weisz, Jain, Joshi, Johnson and Lange 2019). Indeed, consumers sometimes indicate a preference for engaging with humans over chatbots because they believe a human can better understand them and solve their problems (Araujo 2018; LivePerson 2019). This suggests that people may perceive humans (vs. chatbots) to be more helpful—specifically more capable of understanding and solving their problem—in customer service contexts. As such, we hypothesize the following:

*H3a-b: People will perceive a chat agent who makes and corrects typographical errors to be more helpful (H3a) than an agent who does not make errors (or makes but does not correct errors). These effects will be mediated by perceived humanness (H3b).*

In addition to exploring how typos change perceptions of agents’ warmth and helpfulness, we explore three possible downstream behavioral consequences that have marketing implications. These consequences can reveal whether perceiving humanness is related to consumers’ behavioral intentions in current and future interactions.

***Willingness to reward the agent.*** Businesses and service agents need to develop and maintain positive relationships with their customers. One indicator of a positive relationship is consumers’ willingness to endorse the agent’s service immediately after an interaction (e.g., by providing positive ratings or reviews). Accumulating evidence suggests that more humanlike agents can elicit higher levels of cooperation from consumers, such as greater acceptance of their

offers in an online transaction (Schanke, Burtch, and Ray 2021) and more prosocial behavior toward them when they need help (Oliveira et al. 2021). Furthermore, given that humanness is associated with the capacity to feel emotions (Gray et al. 2007; Waytz and Norton 2014), customers may recognize that a human(-like) agent is more likely to care about and enjoy a reward (e.g., service endorsement) for their hard work.

***Responsiveness to the agent's recommendation.*** Customers' willingness to follow a service agent's sale or product recommendation is a critical metric for online service providers because it can enable providers to reach more customers and create more value. Prior research suggests that perceiving online recommendation agents as "social actors" with human characteristics can increase consumers' intentions to comply with their suggestions (Adam et al., 2021; Benbasat and Wang 2005), although the effectiveness of persuasion also depends on characteristics of a consumer, such as their need for connection and level of interpersonal trust (Tam, 2015; Touré-Tillery and McGill, 2015). Thus, it is possible that people would be more interested in checking out a product recommended by a more human(-like) agent.

***Interest in future interactions with the agent.*** User retention and customer loyalty are essential for the long-term success of a company (Gupta and Lehmann, 2003; Reichheld and Sasser, 1990). When customers are interested in engaging in future interactions with the same service agent, it can be a useful early indicator of their customer loyalty. Along this line, research shows that introducing human-like features in a personal intelligence agent (e.g., Siri) increases customers' desire to engage with that agent in the future (Moussawi, Koufaris, and Benbunan-Fich 2021).

Based on the above reasoning, we hypothesize the following:

*H4a: People will be more likely to reward a chat agent (e.g., endorsing their service) who makes and corrects typographical errors than an agent who does not make errors.*

*H4b: People will be more likely to adopt a recommendation from a chat agent who makes and corrects typographical errors than an agent who does not make errors.*

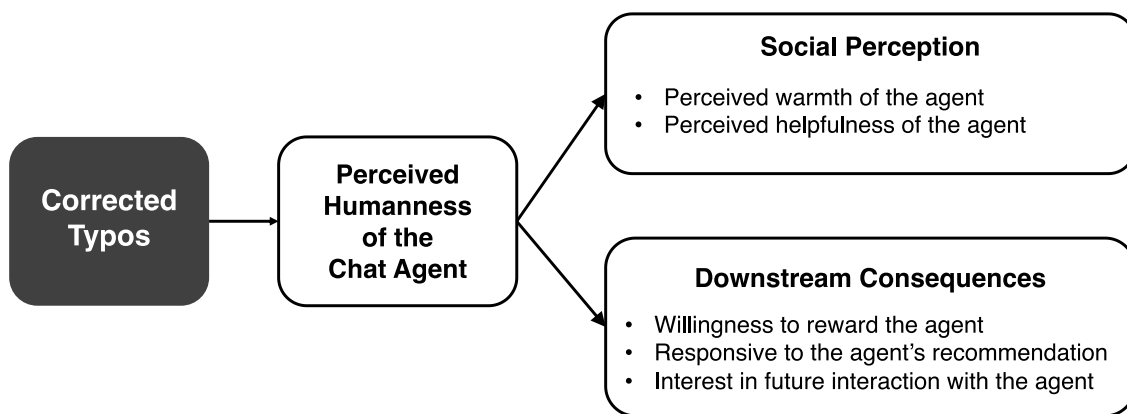
*H4c: People will be more interested in using a chat agent again in the future who makes and corrects typographical errors than an agent who does not make errors.*

*H4d: The above effects (H4a-c) will be mediated by perceived humanness.*

The full theoretical model we propose is shown in Figure 1.

**FIGURE 1**

**PROPOSED THEORETICAL MODEL**



**OVERVIEW OF STUDIES**

We test our predictions across five experiments with a variety of research paradigms. But before testing the specific hypotheses articulated above, we first explore whether people intuitively recognize the value of typos in customer service interactions in a Pilot Study. Given

that typos are usually seen as negative or undesirable features in human-to-human interactions, we think people may hold the *expectation* that typos will harm customer service interactions and fail to recognize their potential humanizing value.

Studies 1-3 test our primary hypotheses that corrected typos, when observed in text-based online chats that unfold over time, can humanize a customer service agent (compared to no typo) and have systematic consequences for how people perceive and interact with the agent. In particular, Study 1 employed an interactive paradigm where the service agent who conversed with participants in an online chat was (unbeknownst to participants) always a human, allowing us to examine the humanizing effect of corrected typos and their impact on perceptions of an agent's warmth (H1a, H2a-b). Study 2 similarly employed real-time conversations between participants and a chat agent with ambiguous identity, but different from Study 1, the agent was always a pre-programmed chatbot. Study 2 tested the same measures as Study 1 but additionally examined the perceived helpfulness of the agent and participants' willingness to reward the agent (H1a, H2a-b, H3a-b, H4a). To compare the humanizing effect of a corrected typo with other potentially humanizing features of text, Study 3 employed a factorial experimental design that include three other cues common in text-based communication—i.e., the agent's name, a human photo, and gender—along with corrected typo, and measured the same consequences as in Study 2. To test the critical role of correcting one's error in humanizing the agent, Study 4 employed the same interactive paradigm as in Study 2 but included an additional, typo-only condition in which the chat agent's typos were left unaddressed (H1a-1b). Finally, Study 5 introduced different consumer service scenarios (i.e., billing, shipment) and explored additional downstream consequences, including participants' responses to the agent's persuasive appeals and their interest in future interactions with the same agent (H4b-4d).

For all studies, we report how we determined our sample size, all data exclusions, all manipulations, and all primary measures. All studies were preregistered on AsPredicted.org. Stimuli, measures, data, and preregistration forms for all studies are available on the Open Science Framework: [[https://osf.io/6s8wm/?view\\_only=c2e1d7bb0efb455fa97deb11a844f6d9](https://osf.io/6s8wm/?view_only=c2e1d7bb0efb455fa97deb11a844f6d9)]. In some studies, we introduced secondary measures to probe other potential effects of corrected typos, which we describe and fully report in the Web Appendix.

### **PILOT STUDY: LAY BELIEFS ABOUT TYPOS**

To investigate people's both general and managerial intuitions about typos in customer service contexts, we asked participants to take the perspective of a customer or a company manager, and measured their expectations about customers' responses to corrected typos in a service agent's text. We predicted that participants would fail to recognize the potential humanizing power of corrected typos, regardless of whether they take the perspective of a customer or a manager.

#### ***Method***

*Participants and design.* We recruited a total of 413 adults ( $M_{age} = 37.64$ ,  $SD_{age} = 13.04$ ; 47.3% female) from Prolific who completed our survey in exchange for \$0.28. We aimed to collect 50 participants in each of the eight experimental conditions, thus 400 participants across conditions. Ten participants were excluded from the analysis—one declined the consent form and thus never started the study, and nine started but did not complete the study—leaving a sample of 403 participants for data analysis.

This study employed a 2 (error: no-typo, typo) × 2 (perspective: customer, manager) × 2 (service context: billing issue, service issue) between-subjects design. Participants were randomly assigned to one of the eight conditions.

*Procedure.* Our primary manipulation of interest was whether participants were told that the customer service agent made no typos (no-typo condition) or made and corrected several typos (corrected-typo condition) when interacting with customers. We added two additional manipulations for enhanced generalizability. First, we manipulated participants' perspective by asking them to imagine either that they wanted to chat with a customer service agent over text to ask for assistance (customer perspective) or that they oversaw customer service agents who assisted customers on such issues (manager perspective). Second, we manipulated whether the customer had a billing issue with their phone plan or a service issue with their Internet plan (see the scenarios in Web Appendix).

*Measures.* After reading their respective scenario, participants reported their prediction of their own intention (customer perspective) or their prediction of their customers' intention (manager perspective) *to interact with the agent* on a composite measure with two items (e.g., "How much [do you/will customers] want to work with this customer service agent?" and "Upon being assigned to this customer service agent, how likely are [you/customers] to log out and try to get a different agent?" (reverse-scored);  $r = .71$ ). Ratings on both items ranged from 1 ("not at all") to 7 ("very much/likely").

To measure general preferences for seeing typos, the last page of the survey asked participants to indicate whether they could think of a situation in which they would prefer a customer service agent who made typos in the text (yes/no) and to describe that situation (if they chose "yes") before reporting their age and gender.

## **Results**

A three-way ANOVA on participants' prediction of customers' intention to interact with the agent revealed the expected significant main effect of typo,  $F(1, 393) = 233.89, p < .001, \eta^2 = .373$ , such that participants reported (or predicted) lower intention to interact with the agent when the agent made a typo ( $M = 3.80, SD = 1.53$ ) than when it did not ( $M = 5.76, SD = 1.03$ ). Not only did participants see little value in using typos in customer service interactions, they also believed that typos would actually *harm* customer engagement. The main effect of perspective was also significant,  $F(1, 393) = 7.84, p = .005, \eta^2 = .020$ , such that participants who took the perspective of a customer predicted greater intention to work with the agent ( $M = 4.98, SD = 1.51$ ) compared to those who took the perspective of a manager ( $M = 4.60, SD = 1.72$ ). The effect of service context was not significant ( $F(1, 393) = .32, p = .569, \eta^2 = .001$ ).

Among all interaction terms, only the interaction between typo and perspective was significant ( $F(1, 393) = 9.07, p = .003, \eta^2 = .023$ ). Specifically, those who took the perspective of a manager expected customers to be even less willing to work with an agent with typos than an agent without typos (typo:  $M = 3.43, SD = 1.40$ ; no-typo:  $M = 5.78, SD = 1.09$ ),  $t(198) = -13.13, p < .001, d = 1.87$ , compared to those who took the perspective of a consumer (typo:  $M = 4.18, SD = 1.57$ ; no-typo:  $M = 5.75, SD = .96$ ),  $t(199) = -8.56, p < .001, d = 1.20$ .

Finally, we examined the proportions of participants who indicated that they could (versus could not) think of a situation where they would prefer a customer service agent who made typos in their text. Overall, only 18% of the participants reported that they could think of a situation where typos might be preferable, which was significantly lower than the chance level of 50%,  $\chi^2 = p < .001$ .

## **Discussion**



The majority of participants in this study believed that agents who make typos will harm customer service interactions. Whether taking the perspective of a consumer or a manager in a customer service interaction, participants reported or predicted that they/customers would be less interested in engaging with customer service agents who made and corrected typos, compared to those who did not make typos. Moreover, most participants struggled to come up with *any* situation where making typos would be preferable for customer service agents. These results suggest that people do not have the intuition that typos could be desirable in online customer service—if anything, people’s general belief is that typos should be avoided.

In the subsequent studies, we examine participants’ *actual* responses to typos when they are encountered in real-time conversations with service agents. In particular, we wondered whether the negative lay beliefs about typos observed in the Pilot Study might be prediction errors, and if seeing an agent actually make and correct typos might result in surprisingly positive reactions. While imagining online chats in the abstract, it is easy to overlook how the dynamic process of generating text could shape people’s perceptions of the communicator; in contrast, we predicted that actually observing or experiencing a written conversation unfold over time would allow for the humanizing power of typos to emerge.

### **STUDY 1: CAN TYPOS HUMANIZE A HUMAN COMMUNICATOR?**

As text-based communication blurs the line between humans and bots and makes it more difficult to discern the humanness of an online interaction partner, we first examine the humanizing effect of typos when customers actually interact with another human service agent. To this end, we hired a team of research assistants who interacted with participants via a private

online chatroom and asked questions to collect factual personal information about the participants. Research assistants in the typo condition made and corrected two typos at the beginning of their conversation, whereas those in the control condition made no typos.

### ***Method***

*Participants and design.* We employed a between-subjects design with two experimental conditions—typo and no-typo—and aimed to recruit 100 participants in each condition.<sup>2</sup> In total, 200 participants were recruited from either a university campus or from a metropolitan area with a more diverse population.<sup>3</sup> Participants completed this experiment in exchange for \$4.00.

Following our predetermined exclusion criteria, we excluded a total of 10 participants because they experienced prolonged disconnection from the chat due to Internet disruption or experimenter error ( $N = 4$ ), lacked sufficient English proficiency to have an online chat ( $N = 2$ ), provided inattentive responses ( $N = 1$ ), or because the human chat agent made unintended errors other than the prescribed typos ( $N = 3$ ). This resulted in a final sample of 190 participants (campus-based sample:  $N = 74$ ,  $M_{\text{age}} = 24.5$ ,  $SD_{\text{age}} = 7.7$ , 52.7% female; downtown city-based sample:  $N = 126$ ,  $M_{\text{age}} = 42.7$ ,  $SD_{\text{age}} = 14.5$ , 37.3% female).

*Human confederates.* Six experienced research assistants served as human chat agents and received rigorous standardized training based on a predetermined script to ensure identical conversations across all participants (see OSF folder for the confederate training protocol and script). The script was pretested to eliminate potential ambiguity in the agent's language and to

---

<sup>2</sup> According to a power analysis, this sample size was sufficient to detect a small-to-medium effect size of  $d = .40$  given power of .80 and an alpha level of .05. We chose the minimum sample size we would need due to the sheer amount of resources required to coordinate a live interaction paradigm that involved recruiting participants to a laboratory to interact with live human confederates.

<sup>3</sup> Additional analyses confirmed that recruitment location did not interact with any of our key dependent variables.

keep participants on the same course of conversation. We trained the human confederates to time their responses to convey the impression of reading and typing.

*Conversation interface.* To enable real-time conversations with human chat agents, we employed the ChatPlat platform (<https://www.chatplat.com/>), a web application that resembles an actual online chatroom and allows multiple users to send and receive messages in a private chatroom in real time. This application has been used in previous research to study human social interactions (e.g., Huang et al 2017; Logg et al 2019). ChatPlat also displays a prompt indicating “someone else is typing” when it detects activity in the other user’s entry box.

*Procedure.* We informed participants that this study was part of “a nationwide project to compare research practices and participant populations across behavioral labs,” and that they would be interacting with a “research agent” who could be either a human or a chatbot (when in fact the agent was a human across conditions). Next, we randomly assigned participants to either the typo or the no-typo condition before matching them with an agent to begin the chat. Participants then saw a loading animation and the sentence, “You will be matched with a research agent in a moment.” In both conditions, they were first greeted by the chat agent, who introduced herself as “Angela”<sup>4</sup>. In the typo condition, Angela made and corrected a typo (“helo /“\*help””; see Figure 2), whereas in the no-typo condition, Angela did not make a typo. Next, Angela provided a brief task instruction and confirmed that the participant was ready to start before proceeding to the first question. To strengthen the manipulation, we then included a second typo and correction in Angela’s first question in the typo condition (“talking” /“\*taking””; see Figure 2). No further typos were presented for the rest of the chat.

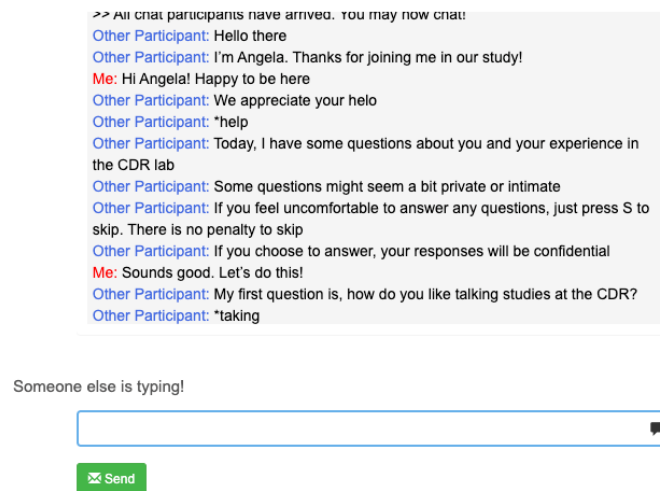
---

<sup>4</sup> We selected Angela as a common, generic female name.

The agent asked a total of eight personal questions that were intended to elicit factual personal information from this sample of participants (e.g., “What is the most sensitive issue that you ever encountered at the [lab name]?”, “What is your primary source of income?”, and “What religion do you believe in, if any?”). Participants could skip any question that they did not wish to respond. When a participant asked a clarifying question, Angela provided either a brief answer (if the question was a simple yes/no question) or a pre-scripted response like “*Thanks for asking, but I can’t comment on that.*” Finally, Angela concluded the chat by indicating that the time was up, thanked the participant, and instructed participants to proceed to the next page.

## FIGURE 2

### EXAMPLE OF THE INTERACTIVE CHAT PLATFORM IN STUDY 1 (CHATPLAT)



*Measures.* First, participants reported their *perception of the agent’s humanness* on a four-item composite measure that was developed based on prior literature and administered across all studies. The first two items focused on the uncertainty commonly experienced by online users (Adam 2018), thus providing face validity to the humanness perception construct:

“To what extent did the customer service agent seem human on the previous screen?” (1 – *not at all human*; 7 – *extremely human*); “In your opinion, how likely is it that the customer service agent is a bot versus a human?” (0 – *very likely a bot*; 100 – *very likely a human*).<sup>5</sup> The next two items were developed to capture people’s lay belief that robots are unable to experience emotions or bodily sensations: “If you told a joke, do you think the agent would get it?” and “Do you think this customer service agent ever gets tired when working?” (1 – *not at all likely*; 7 – *extremely likely*). These four items were highly correlated ( $\alpha = .85$ ) and were combined to form our primary measure of *perceived humanness*. For all statistical analyses in this and subsequent studies, results hold when examining the single, most direct humanness item alone (i.e., “To what extent did the customer service agent seem human in the previous screen?”).

Participants then reported their *perceptions of the agent’s warmth* (friendly/good-natured/warm/sincere;  $\alpha = .95$ ) and *competence* (confident/competent/independent/intelligent;  $\alpha = .90$ ; scale adapted from Fiske et al. 2007). Finally, participants answered a manipulation check item asking them to indicate whether the agent made a typo and, if so, to try to recall they typo. Finally, participants reported their gender, age, and any comments they had about the chat experience. Then, we thanked, debriefed, and compensated participants.

## **Results**

*Manipulation check.* Among all participants, 86.32% passed the manipulation check (92 out of 93 in the no-typo condition, 72 out of 97 in the corrected-typo condition). Results remained the same when we analyzed the data including or excluding participants who failed the manipulation checks. As a more conservative test of our hypotheses, we report results including

---

<sup>5</sup> We linearly transformed ratings on this item to a range of [1, 7] in our subsequent analyses.

all participants regardless of whether they passed manipulation checks in this and subsequent studies.

*Perceived humanness.* Supporting H1a, a chat agent who made and corrected typos was perceived to be more human ( $M = 4.31$ ,  $SD = 1.51$ ) than an agent who did not make any typos ( $M = 3.58$ ,  $SD = 1.58$ ;  $F(1, 188) = 10.48$ ,  $p = .001$ ,  $\eta^2 = .053$ ).

*Perceived warmth.* Supporting H2a, participants in the typo condition perceived the chat agent to be warmer ( $M = 5.07$ ,  $SD = 1.59$ ) than those in the no-typo condition ( $M = 4.52$ ,  $SD = 1.73$ ;  $F(1, 188) = 5.10$ ,  $p = .025$ ,  $\eta^2 = .027$ ). A mediation analysis (SPSS Macro PROCESS, Model 4; 10,000 bootstrap samples) showed that perceived humanness mediated the effect of typo on perceived warmth (supporting H2b;  $b = .41$ ,  $SE = .14$ ; 95% CI = [.16, .70]).<sup>6</sup>

## ***Discussion***

Using a live interaction paradigm with an actual human agent, we found that correcting a typographical error, compared to making no errors, humanized a chat agent (H1a). Moreover, we observed that people perceived the agent to be warmer when they made and corrected their typos, and that these effects were mediated by perceived humanness of the agent (H2a-b). Overall, by implementing a realistic real-time conversation paradigm, similar to what consumers regularly experience when chatting with online service agents, Study 1 provides initial evidence that correcting typos can lead people to perceive an ambiguous communicator—who is actually a human—as more human.

## **STUDY 2: CAN TYPOS HUMANIZE A CHATBOT?**

---

<sup>6</sup> For the sake of thoroughness, we also examined perceived competence and found no significant difference between the typo condition and the no typo condition ( $M_s = 5.10$  and  $4.87$ ,  $SD_s = 1.50$  and  $1.49$ , respectively;  $F(1, 188) = 1.10$ ,  $p = .30$ ,  $\eta^2 = .006$ ). This non-significant pattern was replicated in several subsequent studies and is reported in Web Appendix.

While some customer service agents are humans, an increasing number of agents are powered by chatbots. In Study 2, we investigated participants' evaluations when they interacted with a pre-programmed chatbot, rather than a human, during a live conversation. To this end, we programmed a dynamic chatbot paradigm that enabled a prolonged and real-time engagement between participants and the chat agent, ensuring the psychological realism of our method and the ecological validity of our findings. We hypothesized that the humanizing effect of a typo would be robust even when users were interacting with a non-human bot.

### ***Method***

*Participants and design.* To ensure sufficient statistical power due to the innovative live chat platform, we aimed to recruit 200 participants in each of two experimental conditions, thus 400 participants in total. As planned, 400 participants on Prolific completed our study in exchange for \$2.00 each. According to our pre-registered exclusion criteria, we excluded 14 participants who either provided nonsensical and irrelevant responses or inadvertently asked questions that revealed the program's technological limitations (see next section for detail), and one participant due to suspicious spam activity, resulting in 385 participants in the final analysis ( $M_{\text{age}} = 33.90$ ,  $SD_{\text{age}} = 12.93$ , 53.2% female).

*Chat platform.* Instead of deploying marketplace AI-based chatbots, we developed our own rule-based chatbot so that we could precisely manipulate the presence or absence of typos. Learning from experts who wrote scripts for commercially successful personal assistant products (e.g., Siri and Alexa), we created a script for the chatbot where the chatbot played the role of a research assistant named "Angela" interviewing the participant with a list of personal questions. We developed the chat program using JavaScript, CSS, and HTML codes and embedded the

interactive chat interface in a Qualtrics survey.<sup>7</sup> The program detected keywords in participants' responses and provided minimal, pre-determined responses to convey listening and understanding (see a chat example between the chat program and a participant in Figure 3; the full script and computer codes are available in our OSF folder). For example, when a participant's response contained a question mark, the program would interpret it as a question and respond, "*Please try your best to answer this question.*" This response worked in most circumstances, yet occasionally participants used question marks for rhetorical or clarification purposes (e.g., "What is your next question?"), which triggered the automatic response and therefore exposed that the agent was an algorithm, warranting data exclusion based on our pre-registered exclusion criteria.

Besides scripting the conversation to make the chatbot appear at least somewhat "intelligent," we also implemented a few design features in the chat interface—modeled after a private chat room—that simulated a typical chat experience with a customer service agent. First, before each utterance, the chatbot would pause for a brief moment proportional to the number of words in a participant's previous message, which was meant to convey the impression that the agent spent time reading the message before responding. Second, when the agent was in typing mode, the screen would display "*Angela is typing! Please wait.*" Above the chat box to indicate ongoing activity from the agent and thereby mimic the experience of a chat conversation with service agents. Third, the chatbot was set to be "typing" at a reasonably fast human speed (approximately 7 characters per second, spaces included), creating a realistic waiting time before each message appeared on the screen. As we developed these design features, we also conducted

---

<sup>7</sup> Our OSF folder includes a demonstration video of this chatbot and a Qualtrics survey file (.qsf), to allow interested readers to download the code and create basic chatbots for their own research.



multiple pilot studies to ensure that participants on average perceived the agent to be moderately human.

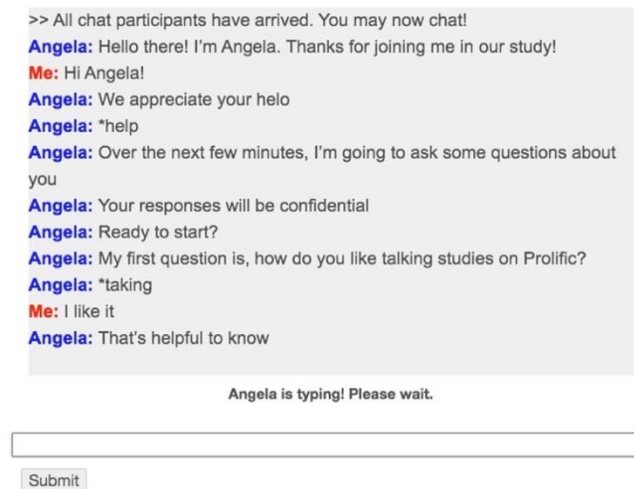
*Procedure.* We informed participants that this study was part of “a nationwide project on participant well-being across behavioral labs during COVID-19,” and that they would be connected to a “research agent” via a chat platform who would collect their information. Next, participants saw a loading animation and the sentence, “You will be matched with a research agent in a moment.” Participants were randomly assigned to either the typo or the no-typo condition. In both conditions, they were first greeted by the chat agent, who introduced herself as Angela.<sup>8</sup> In the typo condition, Angela made and corrected a typo (“helo /“\*help””; see Figure 3), whereas in the no-typo condition, Angela did not make a typo. Next, Angela provided a brief task instruction and confirmed that the participant was ready to start. To strengthen the typo manipulation, we then included a second typo and correction in Angela’s first question in that condition (“talking” /“\*taking””; see Figure 3). No further typos were present for the rest of the chat. After receiving a response, Angela continued with four questions designed to elicit self-disclosure (i.e., “What’s one big goal in your life right now?”, “What’s an important life lesson you learned in the past month?”, “What does a perfect day look like to you nowadays?” and “Would you tell me a little bit about your family background?”). Finally, Angela concluded the chat by indicating that the time was up, thanked the participant, and instructed them to proceed to the next page.

---

<sup>8</sup> Interested readers are invited to try the chatbot at: <https://tinyurl.com/humantypo>. For those who wish to deploy similar chatbots in their research, please contact the lead author for more information.

### FIGURE 3

#### EXAMPLE OF THE INTERACTIVE CHAT PLATFORM IN STUDY 2 (CHATBOT)



*Measures.* Following the chat, participants filled out a survey measuring their evaluation of the agent and the conversation. Similar to Study 1, participants indicated their *perceived humanness* of the agent ( $\alpha = .80$ ), as well as *perceptions of agent warmth* ( $\alpha = .90$ ) and *competence* ( $\alpha = .90$ ). Next, participants reported their *perceptions of the agent's helpfulness* (“How likely do you think it is that this agent could help you solve your problem?” and “How likely do you think it is that this agent would be able to understand you?”;  $r = .87$ ). We further tested a potential behavioral consequence—participants’ *willingness to reward the agent* (“To what extent do you think that this agent should be considered to receive a reward (full week of Starbucks coffee supply) from the company?”). All items were rated on scales from 1 (“*not at all*”) to 7 (“*very much*”).

Finally, participants elaborated on their humanness judgment in an open-ended question, indicated what typo the agent made (if any), and answered demographic questions on their age, gender, highest education level, and race/ethnicity.

#### **Results**

*Manipulation check.* Among all participants, 94.3% (169 out of 191 in the typo condition, 194 out of 194 in the no-typo condition) identified their condition correctly. Results remained the same when we analyzed the data including or excluding participants who failed the manipulation checks. Hence, below we report results including all participants.

*Perceived humanness.* Supporting H1a and replicating the humanizing effect of typos in Study 1, participants who engaged in a live chat with a service agent perceived the agent who made and corrected their typos to be significantly more human ( $M = 3.99$ ,  $SD = 1.72$ ) than one who made no typos ( $M = 3.27$ ,  $SD = 1.34$ ;  $F(1, 383) = 20.96$ ,  $p < .001$ ,  $\eta^2 = .052$ ).

*Perceived warmth and helpfulness.* Regarding the agent's perceived warmth, participants perceived the agent to be relatively warm on average and the direction of the effect was consistent with our hypothesis (H2a; typo:  $M = 5.39$ ,  $SD = 1.27$ ; no typo:  $M = 5.28$ ,  $SD = 1.23$ ), but the effect was not significant ( $F(1, 383) = 0.73$ ,  $p = .39$ ,  $\eta^2 < .001$ ). Although H2a was not supported, the data revealed a significant mediation effect supporting the role of humanness perception (supporting H2b): perceived humanness mediated the relationship between the typo manipulation and perceptions of agent warmth ( $b = 0.34$ ,  $SE = .08$ , 95%  $CI = [.19, .51]$ ).

Regarding the agent's perceived helpfulness, participants in both conditions believed that the agent would be somewhat able to understand and help them solve their issues. The direction was once again consistent with our hypothesis (H3a; typo:  $M = 4.91$ ,  $SD = 1.62$ ; no typo:  $M = 4.70$ ,  $SD = 1.46$ ), but the effect was not significant ( $F(1, 383) = 1.86$ ,  $p = .17$ ,  $\eta^2 = .005$ ). However, supporting H3b, perceived humanness mediated the effect of typo on the agent's perceived helpfulness ( $b = .51$ ,  $SE = .11$ , 95%  $CI = [.28, .74]$ ),

*Willingness to reward the agent.* Supporting H4a, participants who saw the agent make and correct typos were significantly more willing to reward the agent ( $M = 5.34$ ,  $SD = 1.81$ ) than

those who saw an agent that made no typos ( $M = 4.76$ ,  $SD = 1.84$ ;  $F(1, 383) = 9.63$ ,  $p = .002$ ,  $\eta^2 = .035$ ). Supporting H4b, perceived humanness mediated the effect of typo on reward endorsement ( $b = .51$ ,  $SE = .12$ ,  $95\% \text{ CI} = [.28, .75]$ ).

### ***Discussion***

Study 2 employed an interactive paradigm where participants engaged in a live interaction with an ambiguous communicator powered by our pre-programmed chatbot. Replicating Study 1, we observed that making and correcting typos led the chat agent to be perceived as more human (H1a). Furthermore, participants were more inclined to endorse a reward for the typo-correcting chat agent (H4a) than a chat agent that did not make a typo.

Contrary to our predictions, we did not observe significant effects on participants' perceptions of the agent's warmth (H2a) and helpfulness (H3a). We suspect that the attenuated effects for these measures might have at least partly resulted from the complexity of the conversation topics, which ranged from the meaning of life to emotional reactions to the COVID-19 pandemic. Different from the more factual questions in Study 1 and most online service contexts, these open-ended questions about subjective experiences in Study 2 introduced more variance to people's conversation experiences and may have revealed the technological limitations of our chatbot. Importantly, however, assessments of the agent's humanness still mediated all of the above measures (supporting H2b, H3b, and H4d).

### **STUDY 3: TYPO VS. OTHER HUMANIZING CUES**

To compare the effect of typo to other potentially humanizing cues, Study 3 examined not just typos but also two common and possibly humanizing features in online chats: whether an

agent's profile photo portrayed a real human, and whether the agent introduced themselves by name. In addition, we manipulated the agent's gender (male or female) to test whether agent gender might impact perceived humanness or interact with the effect of typo.

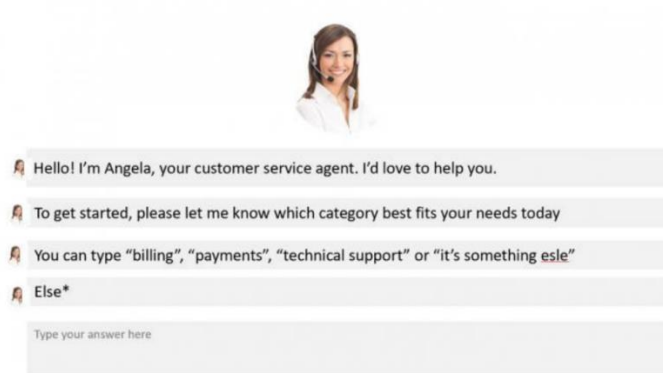
### ***Method***

*Participants and design.* We aimed to collect 50 participants in each of the sixteen experimental conditions, thus 800 participants across conditions. A total of 929 adults ( $M_{\text{age}} = 32.62$ ,  $SD_{\text{age}} = 12.02$ ; 53.1% female) from Prolific started our survey in exchange for \$0.64. Among them, 114 participants never completed the survey and were thus excluded, resulting in 815 participants in the final analysis ( $M_{\text{age}} = 32.60$ ,  $SD_{\text{age}} = 11.95$ ; 51.8% female). This study employed a 2 (error: typo, no-typo)  $\times$  2 (profile photo: human, avatar)  $\times$  2 (name: present vs. absent)  $\times$  2 (agent gender: male, female) between-subjects design (16 separate conditions).

*Procedure.* Participants learned that their task was to read and evaluate online chat messages written by a service agent of a cellular company and then saw a screenshot displaying a customer service agent's greeting messages. The typo condition included a typo in one of the messages (misspelling "else" as "esle") that was corrected in a subsequent message by the same agent ("Else\*"; see Figure 4 and OSF for stimuli presentation). Messages in the no-typo condition included no typo or correction, but were otherwise identical. Along with the messages, participants saw either an avatar or a human photo (i.e., photo manipulation) depicting either a male or a female (i.e., gender manipulation). The agent also either introduced him/herself ("I am Angela [Michael], your customer service agent") or did not ("I am your customer service agent"; i.e., name manipulation). After viewing the agent's greeting messages, participants completed a survey evaluating the agent.

## FIGURE 4

### STIMULI OF TYPO EXAMPLE IN STUDY 3



First, participants reported their *perception of the agent's humanness* on a four-item scale as in previous studies ( $\alpha = .89$ ). Similar to Study 2, participants then reported their *perceptions of the agent's warmth* ( $\alpha = .92$ ) and *competence* ( $\alpha = .88$ ), their *perceptions of the agent's helpfulness* ( $r = .87$ ), and their *willingness to reward the agent*. Finally, participants completed the study by responding to manipulation checks regarding whether the agent had a name, what their profile picture looked like, and what typo the agent made (if any), and then reported their demographic information.

### **Results**

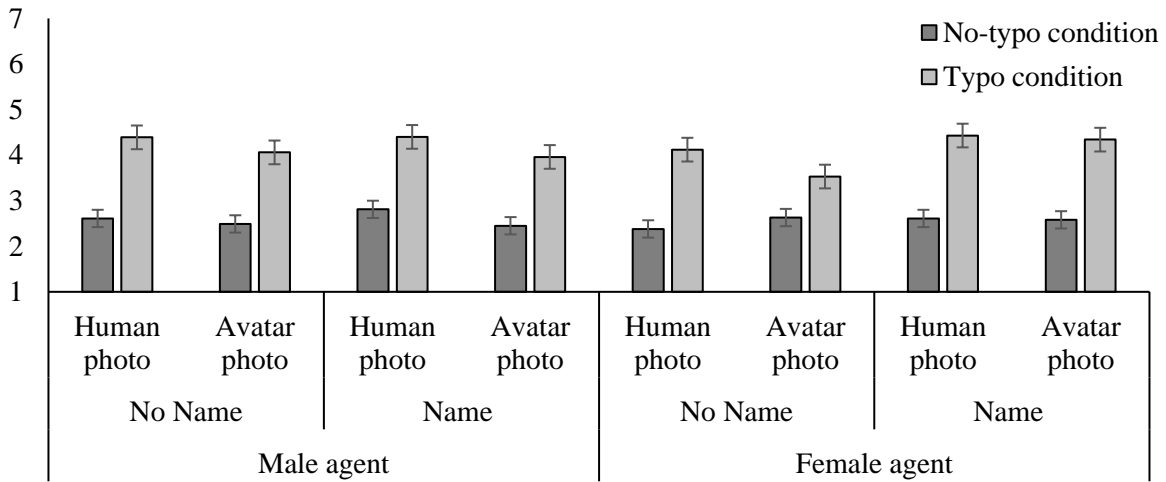
*Manipulation checks.* Among all participants, 89.1% passed the typo manipulation check (406 out of 411 in the no-typo condition, 328 out of 412 in the typo condition), 98.9% passed the photo attention check (407 out of 411 in the no-typo condition, 407 out of 412 in the typo condition), and 77% passed the name attention check (328 out of 411 in the no-typo condition, 306 out of 412 in the typo condition). All findings remained the same when our analyses included or excluded those who failed the manipulation checks. Therefore, below we report results including all participants.

*Perceived humanness.* A four-way ANOVA on the composite score of perceived humanness revealed a main effect of typo, supporting H1a: Participants perceived the agent to be more human when the agent made and corrected a typo in the messages ( $M = 4.08$ ,  $SD = 1.81$ ) than when the agent did not make a typo ( $M = 2.44$ ,  $SD = 1.21$ ;  $F(1, 798) = 227.33$ ,  $p < .001$ ,  $\eta^2 = .22$ ). We also found a significant, although notably smaller, main effect of profile photo ( $F(1, 798) = 4.19$ ,  $p = .041$ ,  $\eta^2 = .005$ ), such that participants perceived the agent with a human photo to be slightly more human ( $M = 3.36$ ,  $SD = 1.75$ ) than one with an avatar photo ( $M = 3.18$ ,  $SD = 1.74$ ). By contrast, mentioning the agent's name only led to a marginally significant increase in perceived humanness (name:  $M = 3.34$ ,  $SD = 1.74$ ; no name:  $M = 3.20$ ,  $SD = 1.74$ ;  $F(1, 798) = 2.84$ ,  $p = .092$ ,  $\eta^2 = .004$ ), and the agent's gender showed no impact on perceived humanness ( $F(1, 798) = .274$ ,  $p = .601$ ,  $\eta^2 < .001$ ).

Of all possible interaction effects, only two emerged as marginally statistically significant: The interaction between typo and profile photo ( $F(1, 798) = 3.04$ ,  $p = .081$ ,  $\eta^2 = .004$ ) reflected a slightly larger humanizing effect of typo when the profile photo depicted a human (typo:  $M = 4.29$ ,  $SD = 1.72$ ; no typo:  $M = 2.27$ ,  $SD = 1.25$ ) than when it depicted an avatar (typo:  $M = 3.88$ ,  $SD = 1.88$ ; no typo:  $M = 2.43$ ,  $SD = 1.19$ ), and the interaction between gender and name ( $F(1, 798) = 2.93$ ,  $p = .087$ ,  $\eta^2 = .004$ ) reflected a slightly larger humanizing effect of name when the agent was a female (typo:  $M = 4.36$ ,  $SD = 1.73$ ; no typo:  $M = 2.45$ ,  $SD = 1.04$ ) than when it was male (typo:  $M = 4.09$ ,  $SD = 1.82$ ; no typo:  $M = 2.48$ ,  $SD = 1.28$ ). No other two-way, three-way, or four-way interactions reached statistical significance ( $F_s < 2.13$ ,  $p_s > .144$ ,  $\eta^2_s < .001$ ). See Figure 5 for graphical representation of the results.

**FIGURE 5**

PARTICIPANTS' PERCEPTIONS OF AN AGENT'S HUMANNESS IN STUDY 3



NOTE.—Error bars represent  $\pm 1$  standard error around the means.

*Perceived warmth and helpfulness.* Regarding the agent's perceived warmth, supporting H2a, participants perceived the agent who made and corrected a typo to be warmer ( $M = 4.42$ ,  $SD = 1.36$ ) than one who did not ( $M = 3.85$ ,  $SD = 1.54$ ;  $F(1, 798) = 33.24$ ,  $p < .001$ ,  $\eta^2 = .040$ ). Participants also perceived an agent who introduced itself by name to be slightly warmer ( $M = 4.24$ ,  $SD = 1.48$ ) than one who did not ( $M = 4.04$ ,  $SD = 1.48$ ;  $F(1, 798) = 4.01$ ,  $p = .045$ ,  $\eta^2 = .005$ ). We found no main effect on warmth of the agent's profile photo ( $F(1, 798) = .481$ ,  $p = .488$ ,  $\eta^2 = .001$ ) or gender ( $F(1, 798) = .419$ ,  $p = .518$ ,  $\eta^2 = .001$ ). Finally, we observed a small and unexpected three-way interaction among typo, name, and gender ( $F(1, 798) = 4.51$ ,  $p = .034$ ,  $\eta^2 = .006$ ). No other interactions were significant ( $F_s < 2.14$ ,  $p_s > .14$ ). Supporting H2b, perceived humanness mediated the effect of typos on warmth perceptions ( $b = 0.50$ ,  $SE = 0.04$ ,  $95\% \text{ CI} = [.41, .59]$ )



Regarding the agent's perceived competence, we found that participants actually perceived the agent in the typo condition to be even *more* competent ( $M = 4.30$ ,  $SD = 1.27$ ) than that in the no-typo condition ( $M = 4.07$ ,  $SD = 1.38$ ;  $F(1, 798) = 6.96$ ,  $p = .008$ ,  $\eta^2 = .009$ ). No other main effects ( $F_s < 2.16$ ,  $p_s > .14$ ) or interactions were significant ( $F_s < 2.34$ ,  $p_s > .13$ ) besides a small three-way interaction among typo, name, and gender ( $F(1, 798) = 6.07$ ,  $p = .014$ ,  $\eta^2 = .008$ ).

Regarding the agent's perceived helpfulness, supporting H3a, participants in the typo condition believed that the agent would be able to better understand them and solve their problems ( $M = 4.39$ ,  $SD = 1.59$ ) than did those in the no-typo condition ( $M = 3.43$ ,  $SD = 1.65$ ;  $F(1, 798) = 69.99$ ,  $p < .001$ ,  $\eta^2 = .081$ ). We found no main effects for other factors ( $F_s < 2.15$ ,  $p_s > .14$ ), and only a marginally significant two-way interaction between typo and name ( $F(1, 798) = 3.73$ ,  $p = .054$ ,  $\eta^2 = .005$ ). Specifically, when an agent introduced themselves by name, their typo led to an even greater increase in perceived understanding and problem-solving ability (typo:  $M = 4.51$ ,  $SD = 1.57$ ; no typo:  $M = 3.33$ ,  $SD = 1.71$ ) than when they did not mention their name (typo:  $M = 4.28$ ,  $SD = 1.60$ ; no typo:  $M = 3.53$ ,  $SD = 1.62$ ). No other two-, three- or four-way interaction effects between factors were significant ( $F_s < 2.33$ ,  $p_s > .127$ ,  $\eta^2_s < .003$ ). Supporting H2d, mediation analysis showed that perceived humanness mediated the relationship between the typo manipulation and participants' expectations of the agent's helpfulness ( $b = .59$ ,  $SE = 0.04$ , 95% CI = [.50, .68])

*Willingness to reward the agent.* Supporting H4a, participants in the typo condition were more inclined to endorse a reward for the agent ( $M = 3.76$ ,  $SD = 1.74$ ) than those in the no-typo condition ( $M = 3.43$ ,  $SD = 1.79$ ;  $F(1, 798) = 29.25$ ,  $p < .001$ ,  $\eta^2 = .035$ ). In addition, when the agent mentioned its name, participants were also marginally more likely to endorse a reward for

the agent (name:  $M = 3.54$ ,  $SD = 1.80$ ; no name:  $M = 3.31$ ,  $SD = 1.77$ ;  $F(1, 798) = 3.57$ ,  $p = .059$ ,  $\eta^2 = .004$ ). No other main effects or interaction effects reached statistical significance ( $F_s < 2.68$ ,  $p_s > .10$ ). Supporting H4b, the relationship between the typo manipulation and endorsing a reward for the agent ( $b = .50$ ,  $SE = 0.04$ , 95%  $CI = [.42, .59]$ ).

### ***Discussion***

Study 3 showed once again that observing a corrected typo in an online chat, compared to observing no errors, led people to perceive the agent as more human. Further, the increase in humanness perception was associated with a host of positive outcomes, such as perceiving the agent as warmer, more helpful, and more deserving of a reward. By contrast, seeing a human profile photo had a notably smaller effect on humanness perception and inconsistent effects on the downstream consequences, and neither the presence of the agent's name nor the agent's gender influenced people's perception of the agent's humanness at all.

## **STUDY 4: CORRECTED VS. UNCORRECTED TYPOS**

To investigate whether it is correcting a typo, rather than just making a typo, that humanizes the chat agent (H1b), Study 4 included three experimental conditions: one in which the chat agent made and corrected typos, another in which the agent just made the typos without correcting them, and the third in which the agent made no typo at all. We used the chat platform we programmed for Study 2 to create a real-time conversation experience, except that in this case the chat agent asked factual questions about participants' personal information (e.g., their date of birth) instead of open-ended life experience questions to make the conversation flow more predictable and reduce variance in participants' experiences.

## ***Method***

*Participants and design.* This experiment employed a between-subjects design with three conditions: corrected-typo, uncorrected-typo and no-typo. We aimed to recruit 200 participants in each of the three experimental conditions, thus 600 participants in total. A total of 603 participants were recruited and completed our study in exchange for \$2.00 each. According to our pre-registered exclusion criteria, we excluded participants whose responses to the open-ended questions were unintelligible or lacked minimal effort ( $N = 24$ ), whose responses to the exploratory 18-item information-disclosure index seemed improbable ( $N = 4$ ), whose chat responses exposed the mechanics of the chatbot ( $N = 51$ ; 44 participants used a question mark that triggered an out-of-context automatic response from the agent and 7 participants attempted to ask the agent multiple questions during the greeting phrase that the agent could not answer), and who experienced technical errors during the chat ( $N = 6$ ), resulting in a final sample of 518 participants for final analysis ( $M_{\text{age}} = 36.94$ ,  $SD_{\text{age}} = 11.94$ , 42.5% female).

*Procedure.* Prior to the study, participants learned that we were “constructing a new participant database to study consumer purchases” and were randomly assigned to the no-typo condition, the corrected-typo condition, or the uncorrected-typo condition prior to the conversation. After entering a virtual chatroom identical to that in Study 2, participants were greeted by a chat agent named Angela. In both the corrected-typo and the uncorrected-typo conditions, the agent made two typos during the conversation identical to those in Studies 1 and 2 (i.e., “*helo*” / “\**help*”, “*talking*”/ “\**taking*”; see Figures 2 and 3). However, in the uncorrected-typo condition, the agent continued the conversation without addressing the typos. During the course of the conversation, the agent asked a total of ten personal questions (e.g., “What is your date of birth?”, “What is your zip code?”, “How much money do you make each month?”, and

“What are the last four digits of your phone number?”), and were again allowed to skip questions by simply pressing a skip button on their interface.

*Measures.* Participants reported the agent’s *perceived humanness* ( $\alpha = .92$ ) and their *perceptions of the agent’s warmth* ( $\alpha = .93$ ) and *competence* ( $\alpha = .88$ ) using the same items as in previous studies. And once again, participants elaborated on their humanness perceptions in an open-ended question and indicated what typo the agent made, if any, in a manipulation check question. This study did not test H3 or H4 because it predated the generation of those hypotheses.

## **Results**

*Manipulation check.* Among all participants, 17.5% (89 participants) failed the manipulation check (66 out of 173 in the uncorrected-typo condition, 23 out of 156 in the corrected-typo condition, and 1 out of 180 in the no-typo condition). Results remained the same when we analyzed the data including or excluding participants who failed the manipulation checks. Therefore, below we report results including all participants.

*Perceived humanness.* A one-way ANOVA revealed a significant effect of the typo manipulation on perceived humanness of the chat agent ( $F(2, 515) = 6.20, p < .001, \eta^2 = .024$ ). Specifically, when the chat agent corrected the typos, participants perceived the agent to be significantly more human ( $M = 4.07, SD = 1.76$ ) compared to when the agent did not make a typo ( $M = 3.45, SD = 1.67; t(340) = 3.32, p = .001, d = 0.36$ ) and when the typos were made but not corrected ( $M = 3.50, SD = 1.85; t(331) = 2.83, p = .005, d = 0.31$ ). In fact, when the agent did not make any typos or did not correct the typos it did make, people perceived similar levels of humanness in the agent ( $t(359) = 0.30, p = .764, d = 0.02$ ). Supporting H1a and H1b, it was

the correction of a typo, rather than the mere presence of a typo, that led participants to perceive a chat agent as more human.

*Perceptions of agent warmth.* A one-way ANOVA on the agent's perceived warmth revealed a significant difference across three conditions ( $F(2, 515) = .3.96, p = .020, \eta^2 = .015$ ). Supporting H2a, participants in the corrected-typo condition perceived the agent to be warmer ( $M = 5.12, SD = 1.29$ ) than those in the no-typo condition ( $M = 4.68, SD = 1.50; t(340) = 2.85, p = .005, d = 0.31$ ), whereas the difference between the uncorrected-typo ( $M = 4.90, SD = 1.48$ ) and the no-typo conditions was nonsignificant ( $t(359) = 1.39, p = .16, d = 0.14$ ). Supporting H2b, mediation analysis showed that perceived humanness of the agent mediated the difference in warmth perception between the corrected-typo condition and the no-typo condition ( $b = .22, SE = .08; 95\% CI = [.06, .38]$ ). In addition, one-way ANOVA on the agent's perceived competence showed no significant effect of the typo manipulation across conditions ( $F(2, 515) = .709, p = .493, \eta^2 = .003$ ).

## ***Discussion***

Using a live chat paradigm, Study 4 supports H1a and H1b, indicating that it is correcting a typo, rather than merely making a typo, that leads to greater perceptions of humanness and warmth. In other words, the presence of errors *per se* does not necessarily humanize a communication, but the act of *correcting* errors seems to convey a mindful agent. To examine the robustness of this finding, we replicated this result in Supplemental Study S1 using a different set of typos (“*meat*” → “*meet*”; “*plajiarisim*” → “*plagiarism*”) in the chat (see Web Appendix). Study S1 also employed a novel button-based chatbot interface whereby the participant selected their response from a set of options, rather than typing it themselves, making

it impossible for participants to “break” the chat program and allowing us to keep 100% of participants in our analysis.

## **STUDY 5: DOWNSTREAM CONSEQUENCES ACROSS SERVICE CONTEXTS**

We conducted Study 5 to fulfill three goals. First, we sought to conceptually replicate how typos impact perceived humanness (H1a) and helpfulness (H3a) of a chat agent in different customer service contexts. Second, we explored additional behavioral consequences particularly relevant to customer service (H4b, H4c). Finally, we counterbalanced the order of humanness and other consequence measures to rule out the possibility that explicitly asking participants to first consider the humanness of the chat agent might have led them to form more positive reactions toward the typo-correcting agent afterwards. We expected that participants would have more favorable perceptions of the agent in the typo condition than in the no-typo condition regardless of whether perceptions of humanness were measured first or last.

### ***Method***

*Participants and design.* This study employed a 2 (error: typo, no-typo) × 3 (service context: billing, product, shipping) between-subjects design. We aimed to recruit 50 participants in each of the six experimental conditions, thus 300 participants in total. A total of 310 participants were recruited in exchange for \$0.28 each ( $M_{\text{age}} = 34.19$ ,  $SD_{\text{age}} = 12.16$ , 47.1% female). Among them, nine participants did not complete the survey and were thus excluded, resulting in 301 participants in the final analysis.

*Procedure.* Participants were randomly assigned to one of six conditions. For the context manipulation, participants imagined that they were experiencing a billing, shipping, or product

issue with their cellphone company and that they logged into the company's website and started to chat with a customer service agent. Similar to Study 3, participants then saw a screenshot displaying a customer service agent's greeting message. The typo condition included a typo in one of the messages (misspelling "else" as "esle") that was corrected in a subsequent message by the same agent, whereas the no-typo condition included identical messages without typo or correction. After viewing the agent's greeting messages, participants were asked to answer a few questions.

*Measures.* To shorten this survey, we did not measure impressions of the agent's warmth or competence. We also shortened our measure of the agent's *perceived humanness*, using a single item with clear face validity: "To what extent does this agent seem human?" (from 1 – not at all human to 7 – very human). Importantly, we counterbalanced the order of the *perceived humanness* item with other consequence measures, such that half of the participants answered the humanness item first, and the other half answered the following consequence measures first.

To examine potential consequences of the perception of humanness, we first measured participants' *perceptions of agent helpfulness* with the same items as in previous studies ( $r = .84$ ). This was followed by a new item measuring participants' *interest in future interactions with the agent* (i.e., "How much would you like to work with this agent again in the future?"). Next, participants were asked to imagine that this agent told them about a new unlimited high-speed wifi service and reported their likelihood of *considering the agent's recommendation* on a single item, "How likely would you be to look into this special offer?" All ratings ranged from 1 ("not at all") to 7 ("very much/likely"). Finally, participants indicated what typo the agent made (if any) as a manipulation check question and reported their age and gender.

## ***Results***

*Manipulation check.* Among all participants, 92.5% (282 participants) passed the manipulation check (130 out of 152 in the corrected-typo condition, 152 out of 153 in the no-typo condition). Results remained the same when we analyzed the data including or excluding participants who failed the manipulation check. Therefore, below we report results including all participants.

*Perceived humanness.* A two-way ANOVA revealed a significant effect of the typo manipulation on perceived humanness of the chat agent ( $F(1, 295) = 155.18, p < .001, \eta^2 = .345$ ). Specifically, supporting H1a, when the chat agent corrected the typos, participants perceived the agent to be significantly more human ( $M = 5.15, SD = 1.78$ ) compared to when the agent did not make a typo ( $M = 2.73, SD = 1.56$ ). The main effect of scenario was not significant ( $F(2, 295) = .394, p = .674, \eta^2 = .003$ ), nor was the interaction effect ( $F(2, 295) = .211, p = .810, \eta^2 = .001$ ).

*Perceived helpfulness.* A two-way ANOVA revealed a significant main effect of typo ( $F(1, 295) = 55.06, p < .001, \eta^2 = .157$ ). Supporting H3a, participants in the corrected-typo condition expected the agent to be more helpful ( $M = 5.12, SD = 1.25$ ) than those in the no-typo condition ( $M = 4.02, SD = 1.31$ ). The main effect of scenario was not significant ( $F(2, 295) = .569, p = .567, \eta^2 = .004$ ), nor was the interaction effect ( $F(2, 295) = .371, p = .690, \eta^2 = .003$ ).

*Consideration of agent's recommendation.* A two-way ANOVA on customers' likelihood to consider the agent's recommendation revealed a significant main effect of typo ( $F(1, 295) = 4.99, p = .026, \eta^2 = .017$ ). Supporting H4b, participants in the corrected-typo condition reported a higher likelihood to look into the agent's recommendation ( $M = 3.88, SD = 1.77$ ) than those in the no-typo condition ( $M = 3.41, SD = 1.85$ ). The main effect of scenario was not significant ( $F(2, 295) = .062, p = .940, \eta^2 < .001$ ), nor was the interaction ( $F(2, 295) = .551, p = .577, \eta^2 = .004$ ).



*Interest in future interactions.* A two-way ANOVA on the willingness to use the agent in future interactions revealed a significant main effect of typo ( $F(1, 295) = 32.51, p < .001, \eta^2 = .099$ ). Supporting H4c, participants in the corrected-typo condition were more interested in interacting with the same agent again in the future ( $M = 4.66, SD = 1.35$ ) than those in the no-typo condition ( $M = 3.72, SD = 1.48$ ). The main effect of scenario was not significant ( $F(2, 295) = .758, p = .470, \eta^2 = .005$ ), nor was the interaction ( $F(2, 295) = .264, p = .768, \eta^2 = .002$ ).

*Order effects.* We examined whether explicitly considering the agent's humanness impacted the humanizing effect of corrected typos and other downstream consequences. Across all outcome variables, no significant interactions emerged between the typo manipulation and the order manipulation ( $ps > .070$ ), suggesting that measurement order did not moderate the effects of corrected typos.

*Mediation effects.* Supporting H3b, perceived humanness of the agent mediated the difference in perceived helpfulness between the corrected-typo condition and the no-typo condition ( $b = -1.02, SE = .12; 95\% CI = [-1.27, -.78]$ ). This mediation was significant both when humanness was measured first ( $b = -1.22, SE = .18; 95\% CI = [-1.60, -.87]$ ) and when it was measured second ( $b = -.81, SE = .15; 95\% CI = [-1.13, -.53]$ ).

Finally, supporting H4d, perceived humanness of the agent mediated the difference in agent's recommendation considerations between the corrected-typo condition and the no-typo condition ( $b = -.73, SE = .15; 95\% CI = [-1.04, -.42]$ ). This mediation was significant both when humanness was measured first ( $b = -.52, SE = .24; 95\% CI = [-.98, -.05]$ ) and when it was measured last ( $b = -.89, SE = .21; 95\% CI = [-1.33, -.50]$ ). In addition, perceived humanness also mediated the difference in participants' interest in future interaction with the agent between the corrected-typo condition and the no-typo condition ( $b = -1.16, SE = .13; 95\% CI = [-1.44, -.91]$ ),

which was also significant both when humanness was measured first ( $b = -1.47$ ,  $SE = .21$ ; 95%  $CI = [-1.91, -1.08]$ ) and when it was measured last ( $b = -.88$ ,  $SE = .17$ ; 95%  $CI = [-1.25, -.57]$ ).

### *Discussion*

Supporting H1a and H3-4, Study 5 showed that across different service contexts, online service agents who made and corrected a typo in text messages were seen as more human than agents who made no typos. Furthermore, making and correcting a typo increased participants' interest in the agent's recommended product and future interactions with the same agent. Finally, Study 5 indicated that the effect of typo on these behavioral consequences was robust to whether participants explicitly considered the agent's perceived humanness before or after reporting their behavioral intentions.

## **GENERAL DISCUSSION**

Online text-based communication is an essential means for companies to provide efficient and scalable customer service experience. As customer service agents in online communications are increasingly pre-programmed AI chatbots (rather than humans), how do consumers discern the identity of those agents, and what consequences do inferences about an agent's humanness have on consumer judgments and behaviors? In this paper, we explored the impact of a common, yet rarely researched, phenomenon in online communication—making and correcting typographical errors (i.e., typos)—on perceptions of humanness.

Although people generally do not expect typos to be desirable in customer service (as seen in the Pilot Study), across five experiments using a variety of stimuli and paradigms—from static or dynamic message displays (Studies 3 and 5) to live chats with a human agent (Study 1)

or a programmed chatbot (Studies 2 and 4)—our research consistently demonstrates that making and correcting one’s typos in a live chat can humanize a customer service agent and elicit favorable perceptions from customers, such as making them appear warmer and more helpful. Crucially, we found that it is the correcting of one’s typo, rather than merely making a typo, that humanizes the agent (Study 4 and Supplemental Study S1). Furthermore, consumers reported greater willingness to reward an agent who made and corrected typos (vs. one who did not make typos or made uncorrected typos), greater likelihood to adopt the agent’s marketing recommendation, and greater interest in interacting with the agent in the future.

### ***Theoretical Contributions***

By emphasizing the dynamic process of how text is generated, our research fills an important gap in the literature on anthropomorphism and mind perception. Different from the majority of prior work that focuses on the outcome of a communication (e.g., typefaces, emojis; Candello et al. 2017; Jakesch et al. 2022; Luangrath et al. 2017; Schroll et al. 2018) or nonverbal cues of the communicator (e.g., appearance, voice; Morewedge et al. 2007; Schroeder and Epley 2016; Zhao et al. 2019) as humanizing forces, our research demonstrates that people also perceive the humanness of communicators based on how their ostensible writing is generated and altered. Given the ubiquity of online chats in consumers’ lives nowadays, which puts the dynamic course of writing on display, our research reveals a critical new dimension of how people perceive humanness in these contexts. Our consistent results highlight how the very process of generating text communication itself — via the act of writing — can provide a unique window into the “mind” of an online agent.

Moreover, our research expands the empirical literature on the interpersonal consequences of making mistakes. Specifically, we help to reconcile an intriguing discrepancy between research suggesting that small errors can increase the perceived likeability of people and products (e.g., Aronson et al. 1966; Ein-Gar et al. 2011) and a separate literature showing that errors can lead to negative impressions of the actor (e.g., Figueredo and Varnhagen 2005). We show that acknowledging and correcting one's error, as opposed to simply making an error, is a key factor for increasing perceived humanness. We theorize that this is because addressing one's error signals an engaged and caring mind—it requires some degree of intentionality (e.g., desire to avoid being seen as careless) and meta-cognitive capacities (e.g., self-awareness), which are generally considered to be lacking in machines, algorithms, and other artificial intelligence. In addition, the fact that perceived competence did not show consistent results across studies (see Web Appendix) suggests that errors do not always lower competence perceptions, at least when they are less central to the product or service per se (Mende et al. 2019).

Finally, our research also highlights a variety of novel consequences of perceiving humanness in chat agents. While text-based communication can feel inherently cold and mechanical, we provide consistent evidence that perceiving an agent as more human can lead to an increase in perceived warmth of the agent. In addition, participants expected an agent that seemed more human to be more helpful in service interactions, likely because people tend to assume that machines, at least in their current capacities, struggle to understand others' needs and inquiries. Several behavioral consequences of perceived humanness also emerged in our studies, including customers' greater willingness to reward the agent, greater responsiveness to the agent's recommendation, and greater interest in future interactions with the agent.

### ***Practical Implications***

Knowledge about what humanizes a communicator can be a double-edged sword. Although it might help service agents and consumers convey their humanness in text-based communication, this knowledge can also allow companies and AI systems to trick users into believing they are interacting with a human when it is really a bot (Jakesch et al. 2022). Keeping such ethical considerations in mind, our research provides practical implications to consumers, service representatives, companies, and policy-makers.

For consumers, our research highlights a cognitive heuristic affecting the way they perceive communicators. Shedding light on how people respond to their conversational counterparts in online settings can help them understand how dynamic cues in text communications affect their consumption experiences. This is especially important since, based on our Pilot Study, consumers are not always aware of how these forces impact their preferences. In the new age of digital communication, where consumers often struggle to discern the identity of their online interaction partners, providing them with knowledge and tools to distinguish humans from bots may help them gain a greater sense of control over their digital environment.

For human customer service representatives as well as their managers, our research provides the reassuring news that, contrary to the intuitions of most people, making a few typos when interacting with customers over text is unlikely to create permanent damage in customer relationships. In fact, typos that are subsequently corrected are humanizing, increasing agents' perceived warmth and helpfulness in the eyes of customers.

For companies that deploy online chat in customer service, our research suggests that humanizing chat agents can create a variety of favorable perceptions. Regardless of whether the company employs a human representative or a chatbot, finding ways to humanize the company's

service agents can signal the company's dedication to connecting with customers, potentially offsetting the impersonal and dehumanizing nature of text-based interactions (while still reaping the benefits of their efficiency). Our research shows that one easy and cost-efficient way to humanize a chat agent is to allow, if not to actively encourage, making and correcting occasional typos. However, it is also important to recognize that humanizing a chat agent (especially a chatbot) can also backfire. As we will discuss below, companies should exercise caution and recognize potential pitfalls when humanizing their chatbots.

Our work also has implications for policymakers on key issues related to transparency and online security. Given that humanizing a chatbot and/or keeping its identity ambiguous can lead to a variety of benefits, our findings illustrate why policymakers need to step in to provide more guidance on what should and should not be allowed when companies deploy increasingly humanlike chatbots. Indeed, recent regulation in California has made it unlawful for a bot to fail to disclose its identity when interacting with consumers or voters ("Bolstering Online Transparency" Act, California Business and Professional Code, Division 7, Part 3, § 17940). The fact that the home state of many tech giants passed such a law highlights the need for public policy to regulate rapidly developing yet potentially controversial technologies.

### ***Limitations and Future Directions***

One big remaining question is whether the humanizing effect of errors is limited to corrected typos compared with the many other types of social errors that can occur during a customer service interaction (e.g., Tian and Oviatt, 2021). Typos are relatively harmless errors that everyone makes and can easily relate to, which might be why they were not perceived as diagnostic for the agent's competence. It is unclear, however, whether other types of errors, such

as grammatical errors or factual inconsistency, would yield similar outcomes. When a customer service agent provides a piece of incorrect information, we suspect that consumers may be less forgiving toward the agent than when merely seeing a typo.

Relatedly, the number of errors, and when they occur in a conversation, may moderate their humanizing effect. Although we showed that correcting one or two typos can humanize a chat agent without creating negative consequences for their perceived competence, it is possible that a large number of typos—even when corrected right away—could ultimately cause negative impressions. We also suspect that typos that occur earlier in a conversation may have more impact on impressions than those that occur later in the conversation (as first impressions tend to be “sticky”).

Future research could also investigate what other downstream consequences may result from observing typos or errors more broadly. Beyond positive attitudes (e.g., perceptions of the chat agent’s warmth and helpfulness) and behavioral intentions (willingness to reward the agent, take their recommendations, and engage with them in the future) toward the agent, it is also possible that customers might form more favorable impressions toward the company or brand (see Web Appendix for some preliminary evidence). Several of our studies explored another potential consequence of making typos – customers’ willingness to disclose their personal information to agents. Because we did not find a consistent main effect of experimental condition on participant disclosure, we relegated these results to the Web Appendix (see full results in Web Appendix). However, we did observe a robust mediation effect of humanness perception across all studies with information disclosure, such that participants were more willing to share personal information with a chat agent when they perceived the agent to be more human. In light of these early and inconclusive results, future researchers might want to

investigate how humanness perception can impact self-disclosure and other conversation behaviors, as well as the underlying psychological mechanisms. For example, a human(like) communicator may highlight the possibility of social evaluation and social surveillance (Epley and Waytz 2010), thus promoting compliance and socially desirable behaviors (e.g., Oliveira et al. 2021), while also discouraging information disclosure that one deems embarrassing or socially inappropriate. Such findings will have important implications for online privacy in numerous contexts where consumers interact with chatbots, from online counseling to conversations on social media.

Though humanizing artificial agents can create a variety of positive outcomes, overly humanizing an agent may introduce several unintended side effects and potential pitfalls. First, while reaping positive effects of perceived humanness can be tempting (see also Svenningsson and Faraon 2019), greater humanness perception can also mean higher expectations for the agent's performance (Malle et al. 2021). People expect human customer service agents to be reliable, empathic, and capable of understanding complex issues (LivePerson 2017), which is often beyond the capacities of current conversational AI systems. Therefore, to avoid setting too high expectations for a human(like) service agent, and later disappointing customers due to limited capabilities, companies may consider humanizing chatbots only for tasks where the performance of AI systems is proven to be reliable.

Moreover, misleading people about the identity of a chat agent could create ethical concerns and even lead to backlash. When consumers discover a company's acts of commission (i.e., actively lying about a chatbot's identity) or omission (i.e., keeping the identity of a chatbot ambiguous), they may feel angry and see the company as manipulative or untrustworthy. Future research should investigate how people react to and perceive a convincingly humanlike chatbot



when its true identity is revealed *after* the interaction. It is possible that the more convincing the human façade a chatbot puts on, the more consumers rely on increasingly sophisticated tools to identify robot imposters such as GPT Zero (<https://gptzero.me>) or Botometer (<https://botometer.iuni.iu.edu/>), and the more negatively they may react when the imposter is exposed. Given that text-based customer service—either powered by human agents or chatbots—is a common way for consumers to interact with companies nowadays and can strongly impact consumer experiences, this work highlights an important avenue for future research in customer service.

The potential for backlash when lying to customers elicits a new and important question: Can corrected typos even have the power to humanize a communicator whose identity is already known to be either a human or a bot? We tested this important question in Supplemental Study S2 by randomly assigning participants to engage with an agent whose identity was described as a human, a chatbot, or was left ambiguous (and who made a corrected typo or not; i.e., a  $2 \times 3$  between-subjects design). Critically, regardless of identity condition, participants who saw the agent make and correct a typo rated the agent as significantly more human than those who saw an agent without a typo, suggesting that typos can be humanizing even for agents who are known to be a human or a bot (for full details, see Web Appendix).

## ***Conclusion***

In the current digital world, consumers commonly use text-based communication channels to interact with companies. Although texting with a customer service agent can be efficient, the inherent lack of nonverbal cues present in these conversations also makes it difficult for customers to experience a “human connection” with such an agent. Our research

suggests that making and correcting typos can humanize an online chat agent, which in turn leads consumers to form more positive impressions of, and behave more positively toward, the agent. Evidently, perceptions of humanness come not only from what is written but also from the process of how the conversation unfolds.

## REFERENCES

- Adams, Tim (2018), "The charge of the chatbots: How do you tell who's human online?" *The Guardian* (November 18), <https://www.theguardian.com/technology/2018/nov/18/how-can-you-tell-who-is-human-online-chatbots>.
- Adam, Martin, Michael Wessel, and Alexander Benlian (2021), "AI-based chatbots in customer service and their effects on user compliance," *Electronic Markets*, 31, 427-445.
- Aggarwal, Pankaj and Ann L. McGill (2007), "Is that car smiling at me? Schema congruity as a basis for evaluating anthropomorphized products," *Journal of Consumer Research*, 34 (4), 468-479.
- Appel, Jana, Astrid von der Pütten, Nicole C. Krämer, and Jonathan Gratch (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, 1–10.
- Araujo, Theo (2018), "Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions," *Computers in Human Behavior*, 85, 183–189.
- Arechar, Antonio A., and David G. Rand (2021), "Turking in the time of COVID." *Behavior research methods*: 1-5.
- Aronson, Elliot, Ben Willerman, and Joanne Floyd (1966), "The effect of a pratfall on increasing interpersonal attractiveness," *Psychonomic Science*, 4, 227-228.
- Benbasat, Izak, and Weiquan Wang (2005), "Trust in and adoption of online recommendation agents," *Journal of the association for information systems* 6(3),4.

- Boden, Margaret (2018), "The robots won't take over because they couldn't care less," *Aeon Essays*, (August 13), <https://aeon.co/essays/the-robots-wont-take-over-because-they-couldnt-care-less>.
- Broadbent, Elizabeth, Vinayak Kumar, Xingyan Li, John Sollers, Rebecca Q. Stafford, Bruce A. MacDonald, and Daniel M. Wegner (2013), "Robots with Display Screens: A Robot with a More Humanlike Face Display Is Perceived to Have More Mind and a Better Personality," *PLoS ONE*, 8(8), 1–10.
- California Bots: disclosure bill, Cal. S. B. 1001 (2017-2018), Chapter 892 (Cal. Stat. 2018).
- Candello, Heloisa, Claudio Pinhanez, and Flavio Figueriredo (2017), "Typefaces and the perception of humanness in natural language chatbots," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, New York, NY: Association for Computing Machinery, 3476–3487.
- Chafe, Wallace L. (1982), "Integration and involvement in speaking, writing, and oral literature," in *Spoken and written language: Exploring orality and literacy*, Deborah Tannen, ed. Norwood, NJ: Ablex, 35-53.
- Christian, Brian (2012), *The Most Human Human: What artificial intelligence teaches us about being alive*, New York, NY: Anchor Books.
- Dal Porto, Lindsay (2017), "Chatbot chatter," *Quality Progress*, 50 (7), 6-9.
- DiSalvo, Carl F., Francine Gemperle, Jodi Forlizzi, and Sara Kiesler (2002), "All robots are not created equal: the design and perception of humanoid robot heads," in *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, New York, NY: Association for Computing Machinery, 321-326.
- Ein-Gar, Danit, Baba Shiv, and Zakary L. Tormala (2011), "When blemishing leads to blossoming: The positive effect of negative information," *Journal of Consumer Research*, 38, 846-859.
- Epley, Nicolas and Juliana Schroeder (2014), "Mistakenly seeking solitude," *Journal of Experimental Psychology: General*, 143 (5), 1980.
- Epley, Nicolas and Adam Waytz (2010), "Mind perception" in *Handbook of Social Psychology*, S. T. Fiske, D. T. Gilbert, and G. Lindzey, eds. Hoboken, NJ: John Wiley & Sons Inc, 498-541.

- Epley, Nicolas, Adam Waytz, and John T. Cacioppo (2007), "On seeing human: a three-factor theory of anthropomorphism" *Psychological Review*, 114 (4), 864.
- Fiske, Susan T., Amy J. Cuddy, and Peter Glick (2007), "Universal dimensions of social cognition: Warmth and competence," *Trends in Cognitive Sciences*, 11, 77-83.
- Figueredo, Lauren and Connie K. Varnhagen (2005), "Didn't you run the spell checker? Effects of type of spelling error and use of a spell checker on perceptions of the author," *Reading Psychology*, 26, 441-458.
- Fulmer, Russ, Angela Joerin, Breanna Gentile, Lysanne Lakerink and Michiel Rauws (2018), "Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial," *JMIR Mental Health*, 5 (4), e64.
- Gottsegen, Gordon (2016), "Are you there, chatbot? It's me, human," *CNET*, (December 21), <https://www.cnet.com/news/youre-talking-to-ai-chatbots-and-its-just-the-beginning/>.
- Grand View Research (2021), "Chatbot Market Size, Share & Growth Report, 2021-2028," *Grand View Research*, <https://www.grandviewresearch.com/industry-analysis/chatbot-market>.
- Gray, Heather M., Kurt Gray and Daniel M. Wegner (2007), "Dimensions of mind perception," *Science*, 315, 619.
- Güntürkün, Pascal, Till Haumann, and Sven Mikolon (2020). "Disentangling the differential roles of warmth and competence judgments in customer-service provider relationships." *Journal of Service Research*, 23(4), 476-503.
- Gupta, Sunil and Donald R. Lehmann (2003), "Customers as assets," *Journal of Interactive Marketing*, 17(1), 9-24.
- Haslam, Nick (2006), "Dehumanization: An integrative review," *Personality and Social Psychology Review*, 10, 252-264.
- Heider, Fritz and Marianne Simmel (1944), "An experimental study of apparent behavior," *The American Journal of Psychology*, 57, 243-259.
- Helmreich, Robert, Eliot Aronson and Jim LeFan (1970), "To err is humanizing sometimes: Effects of self-esteem, competence, and a pratfall on interpersonal attraction" *Journal of Personality and Social Psychology*, 16, 259.
- Hoffmann, Laura, Nicole C. Krämer, Anh D. Lam-chi, and Stefan Kopp (2009), "Media Equation revisited: Do users show polite reactions towards an embodied agent?" in

- Proceedings of the 9th international conference on Intelligent Virtual Agents IVA 2009*, Amsterdam, The Netherlands: Springer, 159–165.
- Huang, Karen, Michael Yeomans, Alison W. Brooks, Julia Minson and Francesca Gino (2017), “It doesn’t hurt to ask: Question-asking increases liking,” *Journal of personality and social psychology*, 113 (3), 430.
- Kemp, Simon (2020), “Report: Most important data on digital audiences during coronavirus,” *Growth Quarters—The Next Web*. (April 24), <https://thenextweb.com/growth-quarters/2020/04/24/report-most-important-data-on-digital-audiences-during-coronavirus/>.
- Kiesler, Sara, Jane Siegel and Timothy W. McGuire (1984), “Social psychological aspects of computer-mediated communication,” *American Psychologist*, 39, 1123.
- Kniffin, Kevin M., Jayanth Narayanan, Frederik Anseel, John Antonakis, Susan P. Ashford, Arnold B. Bakker and Mark van Vugt, (2021), “COVID-19 and the workplace: Implications, issues, and insights for future research and action,” *American Psychologist*, 76 (1), 63.
- Kim, Seo Y., Bernd H. Schmitt and Nadia M. Thalmann (2019), “Eliza in the uncanny valley: anthropomorphizing consumer robots increases their perceived warmth but decreases liking,” *Marketing Letters*, 1-12.
- Kreiner, David S., Summer D. Schnakenberg, Angela G. Green, Michael J. Costello and Anis F. McClain (2002), “Effects of spelling errors on the perception of writers,” *The Journal of General Psychology*, 129, 5-17.
- Kruger, Justin, Nicolas Epley, Jason Parker and Zhi-Wen Ng (2005), “Egocentrism over e-mail: Can we communicate as well as we think?” *Journal of personality and social psychology*, 89 (6), 925.
- Kuchenbrandt, Dieta, Friederike Eyssel, Simon Bobinger and Maria Neufeld (2013), “When a robot’s group membership matters: Anthropomorphization of robots as a function of social categorization,” *International Journal of Social Robotics*, 5, 409-417.
- Larivière, B., Bowen, D., Andreassen, T.W., Kunz, W., Sirianni, N.J., Voss, C., Wunderlich, N.V. and De Keyser, A. (2017), “Service Encounter 2.0”: An investigation into the roles of technology, employees and customers. *Journal of Business Research*, 79, 238-246.

- Lea, Martin and Russell Spears (1992), "Paralanguage and social perception in computer-mediated communication," *Journal of Organizational Computing and Electronic Commerce*, 2, 321-341.
- Lee, Howard (2014), "Paging Dr. Watson: IBM's Watson supercomputer now being used in healthcare," *Journal of AHIMA*, 85 (5), 44-47.
- Lee, Min K. (2018), "Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management," *Big Data & Society*, 5 (1), DOI: 10.1177/2053951718756684.
- Lin, Mariana (2019), "How we can bring AI personalities to life," *TEDSummit*,  
[https://www.ted.com/talks/mariana\\_lin\\_how\\_we\\_can\\_bring\\_ai\\_personalities\\_to\\_life](https://www.ted.com/talks/mariana_lin_how_we_can_bring_ai_personalities_to_life).
- LivePerson (2017), "How consumer view bots in customer care,"  
<https://www.liveperson.com/resources-reports/bots-in-customer-care>
- LivePerson (2019), "How consumers view conversational commerce in 2019,"  
<https://www.liveperson.com/resources-reports/consumer-preferences-2019>.
- Logg, Jennifer M., Julia A. Minson and Don A. Moore (2019), "Algorithm appreciation: People prefer algorithmic to human judgment," *Organizational Behavior and Human Decision Processes*, 151, 90-103.
- Looser, Christine E. and Thalia P. Wheatley (2010), "The tipping point of animacy: How, when, and where we perceive life in a face," *Psychological Science*, 21, 1854-1862.
- Lortie, Catherine L. and Matthieu J. Guitton (2011), "Judgment of the humanness of an interlocutor is in the eye of the beholder," *PLOS One*, 6 (9), e0025085.
- Luangrath, Andrea W., Joann Peck and Victor A. Barger (2017), "Textual paralanguage and its implications for marketing communications," *Journal of Consumer Psychology*, 27, 98-107.
- Lucas, Gale M., Albert Rizzo, Jonathan Gratch, Stefan Scherer, Giota Stratou, Jill Boberg and Louis-Philippe Morency (2017), "Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers," *Frontiers in Robotics and AI*, 4, 51.
- Malhotra, Naresh K., Sung S. Kim and James Agarwal (2004), "Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model," *Information Systems Research*, 15 (4), 336-355.

- Malle, Bertram F., Kerstin Fischer, James E. Young, Ajung Moon and Emily C. Collins (2020), "Trust and the discrepancy between expectations and actual capabilities of social robots," in *Human-robot interaction: Control, analysis, and design*, D. Zhang and B. Wei, eds. New York, NY: Cambridge Scholars Publishing.
- McCoy, John P. and Tomer D. Ullman (2018), "A minimal Turing Test," *Journal of Experimental Social Psychology*, 79, 1-8.
- Memon, Mumtaz A., Cheah J. Hwa, T. Ramayah, Hiram Ting and Francis Chuah (2018), "Mediation Analysis: Issues and Recommendations," *Journal of Applied Structural Equation Modeling*, 2 (1), i–ix.
- Metz, Cade and Keith Collins (2018), "To give A.I. the gift of gab, Silicon Valley needs to offend you," *The New York Times* (February 21), <https://www.nytimes.com/interactive/2018/02/21/technology/conversational-bots.html>.
- Mende, Martin, Maura L. Scott, Jenny van Doorn, Dhruv Grewal, and Ilana Shanks (2019), "Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses." *Journal of Marketing Research* 56 (4), 535-556.
- Mirnig, Nicole, Gerald Stollnberger, Markus Miksch, Susanne Stadler, Manuel Giuliani, and Manfred Tscheligi (2017), "To err is robot: How humans assess and act toward an erroneous social robot," *Frontiers in Robotics and AI*, 4, 21.
- Morewedge, Carey K., Jesse Preston and Daniel M. Wegner (2007), "Timescale bias in the attribution of mind," *Journal of Personality and Social Psychology*, 93, 1–11.
- Moussawi, Sara, Koufaris, Marios and Benbunan-Fich, Raquel (2021), "How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents," *Electron Markets*, 31, 343–364.
- Nass, Clifford and Scott Brave (2005), "*Wired for speech: How voice activates and advances the human-computer relationship*," Cambridge, MA: MIT Press.
- Nass, Clifford and Youngme Moon (2000), "Machines and mindlessness: Social responses to computers," *Journal of Social Issues*, 56, 81-103.
- Norman, Donald A. (2004), "Everyday Things". Basic books.
- Oliveira, Raquel, Patrícia Arriaga, Fernando P. Santos, Samuel Mascarenhas, and Ana Paiva (2020), "Towards prosocial design: A scoping review of the use of robots and virtual agents to trigger prosocial behaviour." *Computers in Human Behavior*, 106547.

- Pierce, Jason R., Gavin J. Kilduff, Adam D. Galinsky and Niro Sivanathan (2013), "From glue to gasoline: How competition turns perspective takers unethical," *Psychological Science*, 24, 1986-1994.
- Press, Gil (November 25 2019). "AI statsnews: Chatbots increase sales by 67% but 87% of consumers prefer humans". Forbes. <https://www.forbes.com/sites/gilpress/2019/11/25/ai-stats-news-chatbots-increase-sales-by-67-but-87-of-consumers-prefer-humans/?sh=19efe3cf48a3>
- Pinker, Steven and Paul Bloom (1990), "Natural language and natural selection," *Behavioral & Brain Sciences*, 13, 707-784.
- Reddy, Trips. (2017), "How chatbots can help reduce customer service costs by 30%," *IBM*. <https://www.ibm.com/blogs/watson/2017/10/how-chatbots-reduce-customer-service-costs-by-30-percent/>
- Reichheld, Frederick F. and W. Earl Sasser (1990). "Zero Defections: Quality Comes to Services," *Harvard Business Review*, September–October, 105–111.
- Schanke, Scott, Gordon Burtch, and Gautam Ray (2021). "Estimating the impact of "humanizing" customer service chatbots." *Information Systems Research*, 32(3), 736-751.
- Schegloff, Emanuel A., Gail Jefferson and Harvey Sacks (1977), "The preference for self-correction in the organization of repair in conversation," *Language*, 53, 361-382.
- Schroeder, Juliana and Nicholas Epley (2015), "The sound of intellect: Speech reveals a thoughtful mind, increasing a job candidate's appeal," *Psychological Science*, 26, 877-891.
- Schroeder, Juliana and Nicholas Epley (2016), "Mistaking minds and machines: How speech affects dehumanization and anthropomorphism," *Journal of Experimental Psychology: General*, 145, 1427-1437.
- Schroeder, Juliana, Michael Kardas and Nicholas Epley (2017), "The humanizing voice: Speech reveals, and text conceals, a more thoughtful mind in the midst of disagreement," *Psychological Science*, 28, 1745-1762.
- Schroll, Roland, Benedikt Schnurr and Dhruv Grewal (2018), "Humanizing products with handwritten typefaces," *Journal of Consumer Research*, 45(3), 648-672.



- Sommer, Barbara A., and Robert Sommer (2003), "A virtual lab in research methods." *Teaching of Psychology* 30 (2), 171-173.
- Tam, Kim-Pong (2015), "Are anthropomorphic persuasive appeals effective? The role of the recipient's motivations," *British Journal of Social Psychology* 54(1), 187-200.
- Tian, Leimin, and Sharon Oviatt (2021), "A taxonomy of social errors in human-robot interaction," *ACM Transactions on Human-Robot Interaction (THRI)* 10 (2), 1-32.
- Touré-Tillery, Maferima, and Ann L. McGill (2015) "Who or what to believe: Trust and the differential persuasiveness of human and anthropomorphized messengers," *Journal of Marketing* 79(4), 94-110.
- Turing, Alan (1950), "Computing machinery and intelligence," *Mind*, 59, 433-460.
- Vignovic, Jane A. and Lori F. Thompson (2010), "Computer-mediated cross-cultural collaboration: Attributing communication errors to the person versus the situation," *Journal of Applied Psychology*, 95, 265.
- Waytz, Adam and Michael I. Norton (2014), "Botsourcing and outsourcing: Robot, British, Chinese, and German workers are for thinking—not feeling—jobs," *Emotion*, 14 (2), 434.
- Weisz, Justin D., Mohit Jain, Narendra Nath Joshi, James Johnson, and Ingrid Lange (2019), "BigBlueBot: teaching strategies for successful human-agent interactions." In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 448-459.
- Westerman, David, Aaron C. Cross and Peter G. Lindmark (2018), "I believe in a thing called bot: Perceptions of the humanness of 'chatbots'," *Communication Studies*, 1-18.
- Wheatley, Thalia, Olivia Kang, Carolyn Parkinson and Christine E. Looser (2012), "From mind perception to mental connection: Synchrony as a mechanism for social understanding," *Social and Personality Psychology Compass*, 6 (8), 589-606.
- Wilson, Timothy D., David B. Centerbar, Deborah A. Kermer and Daniel T. Gilbert (2005), "The pleasures of uncertainty: Prolonging positive moods in ways people do not anticipate," *Journal of Personality and Social Psychology*, 88, 5–21.
- Zhao, Xinshu, John G. Lynch Jr, and Qimei Chen (2010), "Reconsidering Baron and Kenny: Myths and truths about mediation analysis." *Journal of Consumer Research* 37(2), 197-206.

Zhao, Xuan, Elizabeth Phillips and Bertram F. Malle (2019), “How people infer a humanlike mind from a robot body,” forthcoming, <https://doi.org/10.31234/osf.io/w6r24>

Złotowski, Jakub, Diane Proudfoot, Kumar Yogeeswaran and Christoph Bartneck (2015), “Anthropomorphism: opportunities and challenges in human–robot interaction,” *International Journal of Social Robotics*, 7 (3), 347-360.