Title

Imperfectly Human: The Humanizing Potential of (Corrected) Errors in Text-Based Communication

Authors

Bluvstein, Shirley

Zhao, Xuan

Barasch, Alixandra

et al.

Publication Date

2024

DOI

10.1086/728412

Imperfectly Human:

The Humanizing Potential of (Corrected) Errors in Text-Based Communication

ABSTRACT

Today more than ever before, online text-based interactions have become a common means of communication between consumers and companies. But with the advent of AI-powered chatbots, customers sometimes struggle to ascertain the humanness of their online interaction partners (e.g., customer service agents). The current research investigates the humanizing potential of one common feature in text communication—typographical errors ("typos"). Across five experiments, two supplemental experiments, and pilot data (N = 3,399), participants perceived customer service agents who made and subsequently corrected a typo to be more human—and more helpful—than agents who made no typos or made but did not correct a typo. These findings provide novel insights into how conversational features influence customers' perceptions of online agents. In an era where AI frequently surpasses human performance in a variety of domains, consumers may perceive the act of making (and correcting) errors to be a hallmark of humanness.

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

Imperfectly Human:

The Humanizing Potential of (Corrected) Errors in Text-Based Communication

Over the past few decades, two technological inventions have fundamentally reshaped how consumers connect with companies: The Internet has enabled people to communicate with other humans via text on screens instead of by looking directly into each other's eyes, and humanlike chatbots and conversational Artificial Intelligence (AI) technologies have enabled people to communicate with machine algorithms instead of only with other humans. Due to these two inventions, many companies are now able to provide online chat options to customers who want to obtain information or assistance in real time. Although automating consumer service via chatbots promises efficiency and scalability, potentially yielding faster service for customers and financial returns for companies (Reddy 2017), consumers may prefer to receive service from human agents because they assume that a human can better empathize with them, understand their issues, and ultimately solve their problems (LivePerson 2019).

While *being* human is a biological fact, *perceiving* humanness is a psychological process. Research has examined how the aesthetics or content of text can influence perceived humanness. However, little is known on how the *process* of writing — that is, the way in which written communication dynamically unfolds in real time — can impact perception. This paper seeks to fill this gap and demonstrate a novel heuristic for perceiving humanness that arises from the act of writing: making, and subsequently correcting, typographical errors ("typos"). Our work thus highlights an historical inflection point in the relationship between machines and humans: For decades, the frontier of developing more humanlike machine intelligence was to make them

"smarter" and *less* prone to errors; but now, with increasingly sophisticated AI, it may be that the *more* error-prone machines can seem more human in certain contexts.

Perceiving Humanness

Because online conversations are often written, they inherently lack the "richness" of other types of human interactions, meaning that consumers have reduced ability to assess their chat partners' human (or non-human) identities and mental capacities (Kiesler, Siegel, and McGuire 1984; Pinker and Bloom 1990; Schroeder and Epley 2015; Schroeder, Kardas, and Epley 2017). Prior research has examined humanizing cues in social interaction such as appearance, speed of movement, voice, and identity (Looser and Wheatley 2010; Morewedge, Preston, and Wegner 2007; Schroeder and Epley 2016; Zhao, Phillips, and Malle 2019), all of which are largely absent from text.

Within the scope of text-based communication, prior work has focused primarily on two sets of humanizing cues. First, the semantic content of what a person writes—such as the words they choose to use—can provide signals of humanness. This includes both the extent to which a communicator expresses coherent, sophisticated, and relevant thought (Christian 2012; Lortie and Guitton 2011), as well as the linguistic features and social desirability of what they express (McCoy and Ullman 2018). For instance, when guessing whether content was generated by humans or AI, people tend to associate first-person pronouns, spontaneous self-expression, or family topics with humans (Jakesch et al. 2023). Second, people infer humanness by considering the style in which content is communicated. For instance, textual paralanguage—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)—has

been shown to humanize communicators or products (Candello, Pinhanez, and Figueiredo 2017; Luangrath, Peck, and Barger 2017; Schroll, Schnurr, and Grewal 2018).

Notably, the majority of this prior work has focused on how static text — the final product of a communication effort—influences perception of humanness toward the communicator. What is missing in this research is an understanding of how the very process of generating text communication can humanize a communicator. 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. This opens a new and important research space of investigating how the *act* of writing can reveal (or conceal) humanness.

To Err is Human? To Correct, Divine

Making errors is not always seen as a positive, humanizing behavior. Although common wisdom suggests that "to err is human," mistakes are often associated with "stupid" and ignorant AI systems not seen as measuring up to human standards. 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). Moreover, consumers

sometimes prefer products that were made by mistake to products made intentionally (Reich, Kupor, and Smith 2017). Akin to how human (vs. machine) labor can seem more attractive because it is perceived to contain sentimental value (Fuchs et al. 2015) and to be unique (Granulo et al. 2021), errors may be viewed fondly as humanizing pratfalls in certain contexts. For instance, hitting a wrong button on a keyboard seems to be a common human experience to which most people can relate (Norman 2004). Capitalizing on this intuition, some programmers inserted typos in their programs in an attempt to make their chatbots appear more humanlike in the annual Turing Test (i.e., the Loebner Prize; Christian 2012).

However, despite its intuitive appeal, the potentially humanizing effect of typos lacks empirical support. In fact, other research has found that typos consistently 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 spelling and grammatical errors in emails or essays as less competent, less trustworthy, and as having weaker mental capacities (Figueredo and Varnhagen 2005; Kreiner et al. 2002; Lea and Spears 1992; Vignovic and Thompson 2010).

In light of these opposing findings, we contend that an error is neither inherently humanizing nor dehumanizing, but what happens *after* the error is key. 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 in written text can signal humanness 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 the mind is engaged and intentional. Second, correcting one's mistake also signals that one cares about how one is being perceived. 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 how they will be perceived by others (Boden 2018). Drawing on the above reasoning, we hypothesize that 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, and compared to an agent who makes but does not correct their errors.

This novel hypothesized process for perceiving humanness suggests a modern "reversal" of the Turing Test (Turing 1950), such that today's computer programs are becoming so sophisticated, they may now appear more human by simulating humanlike fallibility rather than by being perfect. However, our intent in conducting this research is *not* to provide companies with a tool to manipulate customers, but instead to facilitate new scientific insight about the heuristics consumers use to assess humanness. We hope such understanding can be valuable for

consumers and policymakers in an era where company interactions increasingly occur in a virtual space (see General Discussion).

OVERVIEW OF STUDIES

We test our predictions across five experiments and two supplemental experiments with a variety of research paradigms. Studies 1A – 1C each test our primary hypothesis that corrected typos, when observed in text-based online chats that unfold over time, will humanize a customer service agent (compared to no typo). To test the critical role of correcting one's error in humanizing the agent, Study 2 includes an additional, typo-only condition in which the chat agent's typos were left unaddressed. Finally, Study 3 investigates the moderating role of agent identity: by either keeping the agent's identity ambiguous or declaring it as a chatbot or a human, this study reveals that the humanizing effect of corrected typos is stronger when the agent's identity is ambiguous, but still persists when its identity is known. Critically, this finding suggests that it is possible for firms to implement these humanizing cues while still being transparent to consumers.

For all studies, we report all data exclusions, manipulations, and measures. Due to space constraints, we report some of our preregistered analyses in the Web Appendix (but still list all measured variables within our first study). All studies except for Supplemental Study 2 were preregistered. Stimuli, measures, data, and preregistration forms are available on the Open Science Framework (https://osf.io/6s8wm).

PILOT DATA: LAY BELEIFS ABOUT TYPOS

Before testing our primary hypotheses, we wondered whether people would intuitively recognize the value of typos in customer service interactions. Given that typos can be seen as negative or undesirable features in human-to-human interactions, we predicted that people may hold the *expectation* that typos will harm customer service interactions and fail to recognize their potential humanizing value. A pilot study supported this prediction (see Web Appendix for full description of procedure and results), showing that participants (N = 403) indeed assumed that customers would be less interested in engaging with customer service agents who made and corrected typos (M = 3.80, SD = 1.53), than those who did not make typos (M = 5.76, SD = 1.03; $F(1, 393) = 233.89$, $p < .001$, $\eta^2 = .373$). Moreover, only 18% of the participants reported that they could think of a situation where typos might be beneficial, which was significantly lower than the chance level of 50%, $\chi^2 = p < .001$.

STUDY 1A: CAN TYPOS HUMANIZE A CHATBOT?

In Study 1A, participants interacted with a pre-programmed chatbot during a live conversation. We created this dynamic chatbot paradigm to enable a prolonged, real-time engagement between participants and the chat agent, ensuring the psychological realism of our method and the ecological validity of our findings.

Method

Participants and design. This experiment employed a between-subjects design with two conditions (corrected-typo, uncorrected-typo). Four hundred participants on Prolific completed the study for $2.00 each. According to our pre-registered exclusion criteria, we excluded 14 participants who either provided nonsensible and irrelevant responses or inadvertently asked questions that revealed the program's technological limitations, and one participant due to

suspicious spam activity, resulting in 385 participants in the final analysis (M_{age} = 33.90, 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 in the role of a research assistant named Angela (selected because it was a common female name), who interviewed the participant with a list of personal questions. The program detected keywords in participants' responses and provided minimal, pre-determined responses to convey listening and understanding (see example chat in Figure 1; full script and computer codes available in our OSF folder).[1]

Procedure. We informed participants that this study was part of a nationwide project on participant well-being 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. In the typo condition, the agent made and corrected a typo ("helo /"*help"; see Figure 1), 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 included a second typo and correction in Angela's first question ("talking" / "*taking"; see Figure 1). 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 (e.g., "What's one big

goal in your life right now?"). Finally, Angela concluded the chat by indicating that the time was up, thanked the participant, and instructed them to proceed to the next page.

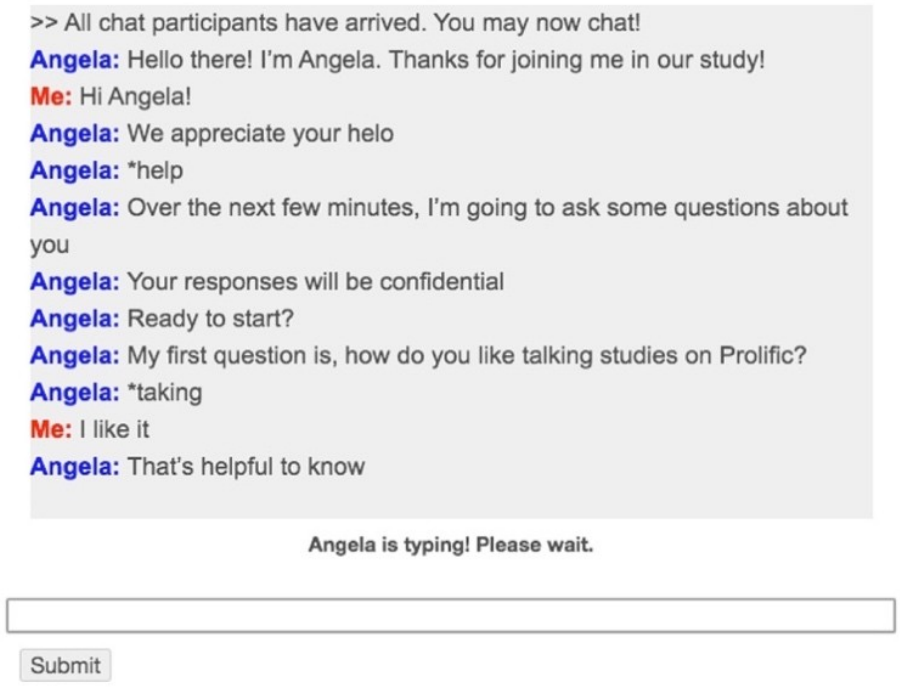

Figure 1. Example of the automated chatbot in Study 1A.

Measures. Following the chat, participants filled out a survey measuring their evaluation of the agent and the conversation. 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 (Adams 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*; we linearly transformed the scale ratings from 1 to 7 for our subsequent analyses). The

next two items were developed to capture people's lay belief that AI systems 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 (α = .80) 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 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; answered on Likert scales from 1 – *not at all* to 7 – *very much*).[2]

Finally, participants elaborated on their humanness judgment in an open-ended question, indicated what typo the agent made (if any) in a manipulation check question, 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. As predicted, participants engaging 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, η^2 =.052).

Perceived helpfulness. Participants in both conditions believed that the agent would be somewhat able to understand and help them solve their issues, with directionally but non-significantly greater perceptions of helpfulness in the typo condition (M = 4.91, SD = 1.62) than in the no-typo condition (M = 4.70, SD = 1.46; F(1, 383) = 1.86, p = .17, η^2 = .005). While there was no direct effect of condition on helpfulness, there was an indirect effect through perceived humanness: Perceived humanness mediated the effect of typo on the agent's perceived helpfulness (b = .51, SE = .11, 95% CI = [.28, .74]).

Discussion

Study 1A employed an interactive paradigm where participants engaged in a live interaction with an ambiguous communicator powered by our pre-programmed chatbot. We observed that making and correcting typos led the chat agent to be perceived as more human. Contrary to our predictions, we did not observe an effect of condition on perceptions of the agent's helpfulness. We suspect that this might have at least partly resulted from the fact that the reflective conversation questions were subjective and open-ended in nature, so participants may not readily see the chatbot as "helpful" given they were asking questions and not providing solutions. In future studies, we examine chatbots in more typical service contexts, where helpfulness is an expected outcome of the conversation.

SUMMARY OF STUDIES 1B AND 1C

Study 1B examined the humanizing effect of typos when customers actually interacted with another human service agent (full description of procedure and results can be found in the

Web Appendix). A team of research assistants interacted with 200 participants (M_{age} = 33.60, 45% female) via a private online chatroom. Research assistants in the typo condition made and corrected two typos during the conversation (as in Study 1A), whereas those in the control condition made no typos. As predicted, 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, η^2 = .053).

In Study 1C, we sought to test how corrected typos impacted perceived humanness of a chat agent across multiple customer service contexts. A total of 301 participants (M_{age} = 34.19, 47.1% female) were randomly assigned to imagine 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. Participants then saw a screenshot displaying a customer service agent's greeting message where the agent either misspelled "else" as "esle" and corrected it in a subsequent message (or did not make a typo). As predicted, when the chat agent corrected a typo, participants perceived the agent to be more human (M = 5.15, SD = 1.78) compared to when the agent did not make a typo (M = 2.73, SD = 1.56; F(1, 294) = 153.25, p < .001, η^2 =.345). Furthermore, 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; F(1, 295) = 55.06, p <.001, η^2 =.157).[3] As in Study 1A, 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]).

To examine whether corrected typos humanize communicators more than other potentially humanizing features, we conducted Supplemental Study 1 (N = 815; see Web

Appendix for details). In this study, we compared the perceived humanness from a corrected typo with three other cues often associated with humans in text-based communication—i.e., the agent's name, gender, and a human photo. Correcting a typo had the strongest effect on perceived humanness ($F(1, 798) = 227.33$, $p < .001$, $\eta^2 = .22$) compared to the other three tested cues.

STUDY 2: CORRECTED VERSUS UNCORRECTED TYPO

To investigate whether it is correcting a typo, rather than just making a typo, that humanizes the chat agent, Study 2 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.

Method

Participants and design. This experiment employed a between-subjects design with three conditions: corrected-typo, uncorrected-typo, and no-typo. A total of 603 participants completed the study for $2.00 each. According to our pre-registered exclusion criteria, we excluded participants whose responses to the questions were unintelligible or lacked minimal effort (N = 28), whose chat responses exposed the mechanics of the chatbot (N = 51[4]), and who experienced technical errors during the chat (N = 6), resulting in a final sample of 518 participants for final analysis (M_{age} = 36.94; 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 the virtual chatroom, participants were greeted by a chat agent. In

both the corrected-typo and the uncorrected-typo conditions, the agent made two typos during the conversation (i.e., "*helo*" / "**help*", "*talking*"/ "**taking*"). However, in the uncorrected-typo condition, the agent continued the conversation without addressing the typos. During the conversation, the agent asked a total of ten personal questions (e.g., "What is your date of birth?", "What is your zip code?"), and participants were allowed to skip questions by simply pressing a skip button on their interface.

Measures. Participants reported their *perceptions of the agent's humanness* ($\alpha = .92$) using the same items as in previous studies. Participants also indicated what typo the agent made, if any, in a manipulation check question.

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 ($M = 3.45$, $SD = 1.67$; $t(340) = 3.32$, $p = .001$, $d = 0.36$) or when the agent made but did not correct the typos ($M = 3.50$, $SD = 1.85$; $t(331) = 2.83$, $p = .005$, $d = 0.31$). In fact, when the agent did not make any typos or did not correct the typos it did make, participants perceived similar levels of humanness in the agent ($t(359) = 0.30$, $p = .764$, $d = 0.02$).

Discussion

Using a live chat paradigm, Study 2 revealed that it is correcting a typo, rather than merely making a typo, that leads to greater perceptions of humanness. 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. We replicated this pattern in Supplemental Study 2 (N = 391) using a different set of typos ("*meat*" → "*meet*"; "*plajiarisim*" → "*plagiarism*") with a button-based chatbot interface (see Web Appendix for details).

STUDY 3: AMBIGUOUS VERSUS KNOWN CHAT AGENT IDENTITY

Across earlier studies, the chat agent's identity was kept ambiguous and left up to participants' inferences. However, in many customer service contexts, consumers *know* that they are talking to a chatbot or a human (e.g., because the company discloses this information). Thus, an important remaining question is whether corrected typos also increase subjective perceptions of humanness even when the identity of the agent is unambiguous. In Study 3, we tested this by directly manipulating the identity of the agent (ambiguous, chatbot, or human) and examining whether the humanizing effect of corrected typos would be moderated by knowledge of the agent's identity. We predicted that corrected typos would have a stronger humanizing effect when the agent's identity was ambiguous because people would be more sensitive to cues that can potentially reveal an agent's identity when that identity isn't already known. However, in light of people's tendency to anthropomorphize objects that merely resemble human appearances or actions (e.g., Aggarwal and McGill 2007; Zhao et al. 2019), we also predicted that correcting typos would still humanize a chat agent that is known to be a chatbot. Finally, given the

dehumanizing nature of written communication for human conversations, we were curious to examine whether correcting typos could even further humanize a known human agent.

Method

Participants and design. This experiment employed a 2 (error: typo, no-typo) × 3 (agent identity: ambiguous, chatbot, human) between-subjects design. A total of 386 participants were recruited via Prolific and completed the study for $0.50 each ($M_{age}$ = 34.47; 54.9% female).

Chat interface. We created a dynamic online platform, showing a chat agent sending one message at a time, and asked participants to evaluate the agent. We employed this paradigm—instead of a live chat—to keep the conversation identical and to only manipulate participants' knowledge of the agent's identity across conditions.

Procedure. Participants learned that they would view an online chat conducted on a cellular company's customer service platform. Depending on their condition assignment, participants were either told that this company was training "their new AI-powered chatbot that they call Angela" (chatbot condition) or "their new employee, named Angela" (human condition) to "use an interactive chat platform to answer customer service questions," or that the cellular company was "developing an interactive chat platform to answer customer service questions" (ambiguous condition). Participants in all conditions were then asked to watch and evaluate a dynamic online chat conducted on this platform. On the next page, participants saw the chat agent send greeting messages line-by-line. The agent in the typo condition misspelled "else" as "esle" and corrected it in a subsequent message, whereas the agent in the no-typo condition spelled everything correctly.

Measures. After seeing the chat page, participants responded to three *perceived humanness* questions identical to those in previous studies (α = .79; we removed "How likely is

it that the customer service agent is a bot versus a human?" because the agent's identity was known in some conditions). Participants also reported their *perceptions of the agent's helpfulness* (r = .85). We counterbalanced whether perceived humanness was measured before or after perceived helpfulness to ensure that the main effects and mediation results were robust to measurement order. Finally, participants completed manipulation checks regarding the agent's identity ("a human employee," "an AI-powered chatbot," or "it wasn't mentioned") and what typo the agent made (if any), before reporting their gender and age.

Results

Manipulation checks. For the typo manipulation check, 95.8% of the participants correctly identified their condition. For the identity manipulation check, 127 out of 127 (100%) participants in the chatbot condition and 104 out of 129 in the human condition (80.6%) correctly reported the agent's identity; by contrast, only 86 out of 130 in the ambiguous condition (66.2%) correctly indicated that the agent's identity was not mentioned, whereas others believed it was an AI-powered chatbot (18.5%; 24 out of 130) or a human employee (15.4%; 20 out of 130). Including or excluding those who failed the identity manipulation check led to largely similar results; below, we report results including all participants (results excluding participants who failed the agent identity manipulation are reported in the Web Appendix).

Perceived humanness. As predicted, a two-way ANOVA on perceived humanness revealed a significant main effect of the corrected-typo manipulation ($F(1, 380) = 102.11$, $p < .001$, $\eta^2_p = .21$), such that participants perceived the agent to be more human when it made a corrected typo (M = 4.49, SD = 1.56) than when it did not (M = 3.15, SD = 1.43). It also revealed a significant main effect of agent identity ($F(2, 380) = 53.06$, $p < .001$, $\eta^2_p = .22$) such that participants perceived a known human agent to be more humanlike (M = 4.57, SD = 1.57)

than an ambiguous agent (M = 3.97, SD = 1.63; t(257) = 3.01, p = .001, d = .37), which was more human than a known chatbot agent (M = 2.90, SD = 1.24; t(255) = 5.86, p < .001, d = .73). Finally, there was a significant interaction effect (F(2, 380) = 6.01, p = .003, η^2_p = .031; see Figure 2). Examining the effect of the typo manipulation on each agent identity condition separately, we found that the humanizing effect of a corrected typo was larger when the agent's identity was ambiguous (corrected-typo vs. no typo: Ms = 4.97 vs. 3.14, SDs = 1.34, 1.38; t(128) = 7.62, p < .001, d = 1.34) or was known to be a human (Ms = 5.30 vs. 3.78, SDs = 1.31, 1.45; t(127) = 6.25, p < .001, d = 1.10) than when it was known to be a chatbot (Ms = 3.25 vs. 2.53, SDs = 1.18, 1.22; t(125) = 3.40, p < .001, d = 0.60).

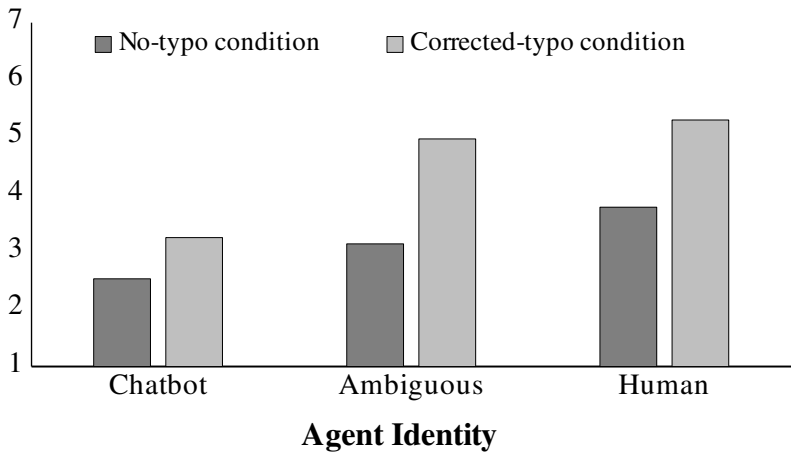

Figure 2. Results on perceived humanness in Study 3. The dotted horizontal line represents the neutral rating of 4. Error bars represent ±1 standard error around the means.

Perceived helpfulness. A two-way ANOVA revealed a main effect of the typo manipulation (F(1, 380) = 11.83, p < .001, η^2_p = .030), such that participants perceived the agent to be more helpful when it made a corrected typo (M = 4.59, SD = 1.49) than when it did not (M

= 4.10, SD = 1.60).[5] It also revealed a main effect of agent identity ($F(2, 380) = 25.15$, $p < .001$, $\eta^2_p = .12$), such that participants believed that a human employee (M = 4.85, SD = 1.43) would be marginally more helpful than an ambiguous agent (M = 4.54, SD = 1.43; $t(257) = 1.75$, $p = .08$, d = .22), and an ambiguous agent would be more helpful (M = 4.54, SD = 1.43) than a chatbot (M = 3.61, SD = 1.58; $t(255) = 4.95$, $p < .001$, d = .62). There was no interaction effect ($F(2, 380) = .528$, $p = .59$, $\eta^2_p = .003$). Finally, perceived humanness of the agent mediated the difference between the no-typo condition and the corrected-typo condition (b = .84, SE = .11; 95% CI = [.63, 1.06]).

Discussion

Consistent with our prediction, knowledge about a chat agent's identity moderated the humanizing effect of typos. In particular, the humanizing effect of the typo was larger when the agent's identity was ambiguous than when the agent was known to be a chatbot. But importantly, correcting a typo still humanized the communicator even when people knew that the chat agent was a chatbot or a human. Furthermore, seeing an agent correct a typo again led people to expect the agent to be more helpful than seeing an agent that made no typo.

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 such conversations are increasingly pre-programmed AI chatbots (rather than humans), it is important to understand how consumers perceive the humanness of chat agents and the consequences of these perceptions for consumer judgments and decisions. Across five experiments and two supplemental experiments using a variety of stimuli and paradigms—from static or dynamic

message displays, to live chats with a programmed chatbot, to live chat with a human agent—our research consistently demonstrates that making and correcting one's typos in a conversation can humanize a service agent (see summary results in Figure 3). Furthermore, correcting typos can elicit favorable perceptions from customers, such as making those agents appear more helpful. Crucially, we found that it is the correcting of one's typo, rather than merely making a typo, that humanizes an agent.

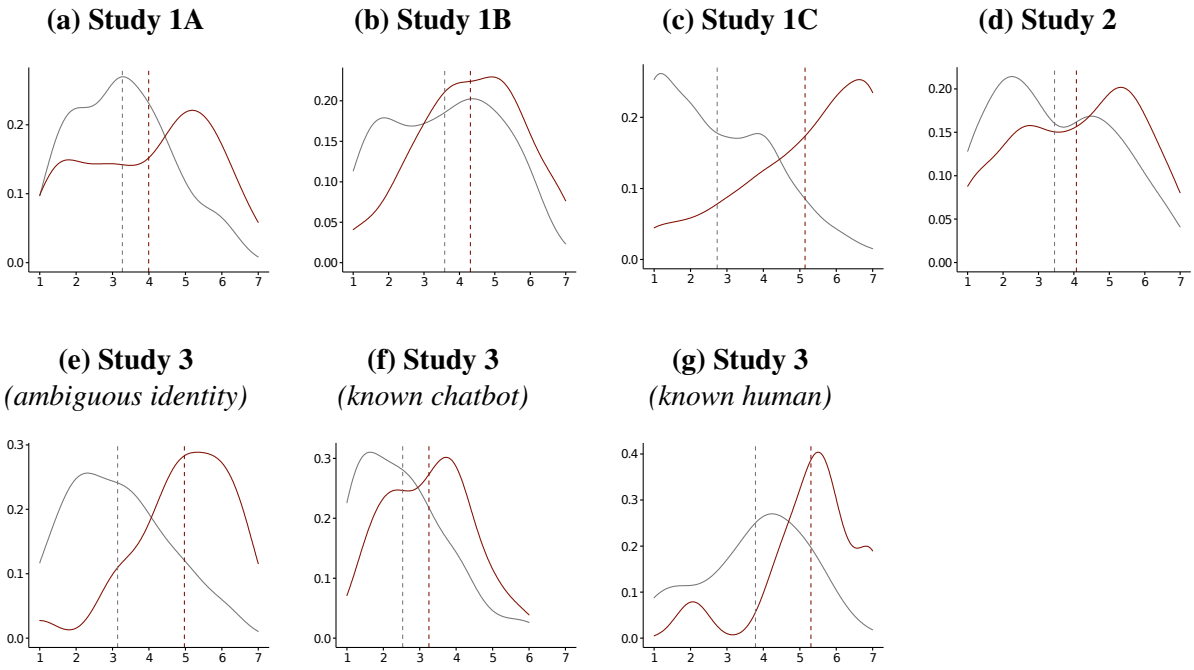

Figure 3. Distribution of perceived humanness in the no typo (in grey) versus corrected typo (in red) conditions across all studies.

Theoretical Contributions

By emphasizing the dynamic process of how text is generated, our research fills an important gap in the literature on 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. 2023; Luangrath et al. 2017; Schroll et al. 2018) or verbal and nonverbal cues of the communicator (e.g., speed of movement, voice; Morewedge et al. 2007; Schroeder and Epley 2016), our research highlights how the very process of writing the text 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) 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.

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 might also tempt companies to trick users into believing they are interacting with a human when it is really a bot. Concerns regarding harm to consumers illustrate why policymakers should provide more guidance to, and constraints on, companies seeking to deploy increasingly humanlike chatbots. Indeed, recent policy efforts in the United States include

some states requiring bots to disclose their identities when interacting with consumers or voters and proposed solutions like watermarking AI-generated content.

As attempts to deceive consumers within text-based conversations will likely result in negative consequences for the agent and the company, we believe companies should focus their efforts on authentic and transparent humanization of customer service agents to maximize potential benefits (e.g., appearing helpful and warm) and minimize potential backlash (e.g., appearing manipulative and untrustworthy). As we found in Study 3, even when explicitly telling people that an agent is a human or chatbot, corrected typos *still* had an additional humanizing impact, suggesting that companies can be honest while simultaneously enjoying the benefits of enhanced humanization of their agents. 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. Relatedly, our research provides the reassuring news that, contrary to the intuitions of many people, making a few typos when interacting with consumers is unlikely to create permanent damage in customer relationships (and may even have positive effects on a variety of other downstream consequences, such as perceptions of an agent's warmth, willingness to reward an agent or take their recommendation, and interest in interacting with an agent again in the future).

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 sometimes 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.

Limitations and Future Directions

Our results elicit several questions regarding generalizability. First, the current investigation used typos that involved adjacent keys on the keyboard (e.g., mistakenly typing "p" as "o" in "help"), which may be uniquely humanizing because they imply "motoric failures" caused by physical embodiment that only humans have. It would be interesting to examine whether the humanizing effect of corrected typos would be stronger in situations where the motoric failure is more obvious, such as when it involves a categorical change (e.g., typing "p" as "["), or weaker for typos that seem unrelated to motoric failures (e.g., typing "p" as "h").

Second, 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 general competence (see Web Appendix for these results). It is unclear, however, whether the type of errors (e.g., grammatical errors, incorrect information), the frequency of errors (i.e., number of errors the communicator makes), or the customer's *goal* within the interaction (e.g., seeking therapy vs. scientific facts) might moderate how the errors affect consumer perceptions. Moreover, relationships among the perception of humanness, helpfulness, and general competence may be nuanced and context-dependent— for instance, in domains where machines are known to outperform humans (e.g., information retrieval), consumers might perceive a more machine-like agent to be more helpful.

Third, even though we limited our current investigation to correcting typos as a humanizing cue, our theoretical argument suggests that other changes in text that unfold over time can signify a thoughtful, engaged, and intentional mind. From pressing backspace to delete one's writing, to changing one's ideas as conversation unfolds, future research could examine how these traces of a fallible mind can be uniquely humanizing. Moreover, given that other characteristics of written communication (e.g., emojis and informal language) have also been shown to humanize a writer, future research can compare the humanizing power of the dynamic process of writing versus the final product of writing.

Fourth, it is worth investigating what other downstream consequences may result from observing typos or errors more broadly. Our Web Appendix reports the results of the corrected-typo manipulation on several other possible consequences, some of which showed relatively consistent effects (e.g., such as perceptions of an agent's warmth, willingness to reward an agent or take their recommendation, and interest in interacting with an agent again in the future), but others which were inconsistent (e.g., customers' willingness to disclose their personal information to agents). 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. 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.

REFERENCES

Adams, Tim (2018), "The charge of the chatbots: How do you tell who's human online?" *The Guardian* (November 18).

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.

Aronson, Elliot, Ben Willerman, and Joanne Floyd (1966), "The effect of a pratfall on increasing interpersonal attractiveness," *Psychonomic Science*, 4, 227-228.

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.

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.

Christian, Brian (2012), *The Most Human Human: What artificial intelligence teaches us about being alive*, New York, NY: Anchor Books.

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.

Fuchs, Christoph, Martin Schreier, and Stijn MJ Van Osselaer. "The handmade effect: What's love got to do with it?" *Journal of marketing* 79, no. 2 (2015): 98-110.

Gray, Heather M., Kurt Gray and Daniel M. Wegner (2007), "Dimensions of mind perception," *Science*, 315, 619.

Granulo, Armin, Christoph Fuchs, and Stefano Puntoni. "Preference for human (vs. robotic) labor is stronger in symbolic consumption contexts." *Journal of Consumer Psychology* 31, no. 1 (2021): 72-80.

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.

Jakesch, Maurice, Jeffrey T. Hancock, and Mor Naaman (2023). "Human heuristics for AI-generated language are flawed," *Proceedings of the National Academy of Sciences,* 120(11).

Kiesler, Sara, Jane Siegel and Timothy W. McGuire (1984), “Social psychological aspects of computer-mediated communication,” *American Psychologist*, 39, 1123.

Kreiner, David S., Summer D. Schnakenberg, Angela G. Green, Michael J. Costello and Anis F. McClin (2002), “Effects of spelling errors on the perception of writers,” *The Journal of General Psychology*, 129, 5-17.

Lea, Martin and Russell Spears (1992), “Paralanguage and social perception in computer-mediated communication,” *Journal of Organizational Computing and Electronic Commerce*, 2, 321-341.

LivePerson (2019), “How consumers view conversational commerce in 2019,” https://www.liveperson.com/resources-reports/consumer-preferences-2019.

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.

McCoy, John P. and Tomer D. Ullman (2018), “A minimal Turing Test,” *Journal of Experimental Social Psychology*, 79, 1-8.

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.

Norman, Donald A. (2004), “Everyday Things'. Basic books.

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*.

Reich, Taly. Daniella Kupor and Rosanna Smith (2017) ,"Made By Mistake: When Mistakes Increase Product Preference ", in NA - Advances in Consumer Research Volume 45, eds. Ayelet Gneezy, Vladas Griskevicius, and Patti Williams, Duluth, MN : *Association for Consumer Research*, Pages: 834-838.

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.

Turing, Alan Mathison (1950), "Mind." *Mind,* 59(236), 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.

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.

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

End Notes

[1] We developed the chat program using JavaScript, CSS, and HTML codes and embedded the interactive chat interface in a Qualtrics survey. 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. Interested readers are invited to try the chatbot at: https://tinyurl.com/humantypo. Those who wish to deploy similar chatbots in their research may contact the correspondence author. As reported in the main text, this program detected keywords in participants' responses and provided minimal, pre-determined responses to convey listening and understanding. 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 multiple pilot studies to ensure that participants on average perceived the agent to be moderately human.

[2] Note that for the sake of brevity, we are only able to report participants' perceptions of humanness and helpfulness in the main text. We also measured several other downstream consequences of perceived humanness across our studies, including perceptions of the agent's a) warmth and b) competence (Study 1A, Study 1B, Study 3, Supplemental Study 2), c) perceived interpersonal closeness to the agent and d) satisfaction with the conversation (Study 1A, Study 1B, Study 2), e) willingness to reward the agent and f) impression of the company using the agent (Study 1A, Study 3, Supplemental Study 2), g) consideration of the agent's recommendation and h) interest in future interaction with the agent (Study 1C), and i) whether participants disclose personal information to the agent (Study 1A, Study 1B, Study 2, Study 3, Supplemental Study 2). Many of these secondary measures also have positive effects from corrected typos (e.g., perceived warmth, willingness to reward the agent, consideration of the agent's recommendations), though other measures have less consistent effects (e.g., disclosure of personal information). We describe and fully report all of these measures in the Web Appendix.

[3] This study also manipulated whether perceived humanness was measured before or after perceived helpfulness; regardless of the order, participants always perceived the agent in the typo condition to be more helpful, ts > 3.93, ps < .001, ruling out the possibility that any effect of condition on perceived helpfulness resulted only from participants considering the humanness first.

[4] 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.

[5] Similar to Study 1C, measuring perceived humanness before or after perceived helpfulness did not affect the results. An ANOVA revealed no interaction effects between measurement order and the other factors on perceived humanness (ps > .11) or perceived helpfulness (ps > .30).