Journal of Business Research 115 (2020) 14–24

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

Customer service chatbots: Anthropomorphism and adoption

Ben Sheehan*, Hyun Seung Jin, Udo Gottlieb

School of Advertising, Marketing and Public Relations, QUT Business School, Queensland University of Technology, 2 George Street, Brisbane, QLD 4001, Australia

ARTICLE INFO

Keywords:
Chatbots
Artificial conversational entities
Self-service technology
Anthropomorphism
Perceived humanness
Need for human interaction

ABSTRACT

Firms are deploying chatbots to automate customer service. However, miscommunication is a frequent occurrence in human-chatbot interaction. This study investigates the relationship between miscommunication and adoption for customer service chatbots. Anthropomorphism is tested as an account for the relationship. Two experiments compare the perceived humanness and adoption scores for (a) an error-free chatbot, (b) a chatbot seeking clarification regarding a consumer input and (c) a chatbot which fails to discern context. The results suggest that unresolved errors are sufficient to reduce anthropomorphism and adoption intent. However, there is no perceptual difference between an error-free chatbot and one which seeks clarification. The ability to resolve miscommunication (clarification) appears as effective as avoiding it (error-free). Furthermore, the higher a consumer's need for human interaction, the stronger the anthropomorphism - adoption relationship. Thus, anthropomorphic chatbots may satisfy the social desires of consumers high in need for human interaction.

1. Introduction

Chatbots are computer programs with natural language capabilities, which can be configured to converse with human users (Maudlin, 1994). Tintarev, O'Donovan, and Felfernig (2016) conceptualize chatbots as automated advice givers in that they can facilitate decision making. The chatbot ecosystem includes voice-driven digital assistants (Siri, Cortana, Alexa and Google Home) as well as text-based systems deployed to instant messaging platforms. Projections suggest that by 2020, one quarter of customer service processes will integrate chatbot technology (Moore, 2018) and that the average person will have more conversations per day with a chatbot than with their partner (Gartner, 2016). When chatbots facilitate customer service, independent of a human service agent, they can be conceptualized as a self-service technology (SST) (Doorn et al., 2017).

Despite the proliferation of these chatbots, they often fall short of consumers' expectations because they fail to understand user input. For example, Facebook's Project M (a text based virtual assistant) is thought to have failed in over 70% of interactions (Griffith & Simonite, 2018), requiring a human service agent to intervene. Miscommunication errors may damage the public perception of chatbots, such as when the press reported on two mental health chatbots designed for children which failed to recognize sexual abuse (White, 2018). Even the world's best chatbots miscommunicate. A review of transcripts from one of the world's preeminent chatbot competitions, the Loebner Prize, shows that miscommunication in human-chatbot interaction is very common (Martin, 2017, 2018).

Therefore, our two studies compare human-chatbot interactions where miscommunication occurs to interactions in which it does not. The first study had participants view an animation of a human-chatbot interaction in order to rate the chatbot's performance. The second study enhanced ecological validity, asking participants to converse with a chatbot directly and rate the chatbot with which they interacted. We aim to shed light on three interconnected research questions. First, how might we bridge the gap between the ideal chatbot with perfect comprehension and today's commercial chatbots, which are prone to miscommunication? To achieve this, two experiments compare a hypothetically perfect chatbot (henceforth: error-free) to a chatbot which struggles to infer meaning and thus seeks clarification (clarification) to a third chatbot which produces an error in comprehension (error). We do this in order to examine the effect of a chatbot seeking clarification. Improving chatbot adoption is valuable, given text-based chatbots operating inside instant messaging platforms can reach of over 2.5 billion people (Clement, 2019).

Second, our studies investigate potential explanations for the relationship between chatbot (mis)communication and adoption intent. This research tests the mediating role of anthropomorphism, described as inductive inference in which the perceiver attributes humanlike characteristics, motivations, intentions or underlying mental states to a non-human entity (Waytz et al., 2010). Anthropomorphism has been a key variable in chatbot development for decades. Nearly 70 years ago, Turing (1950) provided instructions for what he called an "imitation

* Corresponding author.
E-mail address: b2.sheehan@qut.edu.au (B. Sheehan).

https://doi.org/10.1016/j.jbusres.2020.04.030
Received 13 February 2019; Received in revised form 8 April 2020; Accepted 11 April 2020
Available online 29 April 2020
0148-2963/

game", designed to test a machine's capacity for intelligent behavior. The test, now commonly known as the Turing test, asks human judges to detect whether they are in conversation with a computer or a human confederate. Turing's focus was machine intelligence, but the mechanism by which he proposed to explore it was communication. Modern versions of the Turing test have produced some of the world's best chatbots, including ALICE (Artificial Language Internet Computer Entity) and Mitsuku. According to the Turing test, the perfect chatbot would be indistinguishable from a human interlocutor. Thus, perceived humanness or anthropomorphism has guided advances in chatbot technology since the 1950's. Hence, anthropomorphism was selected as an appropriate mediator for our study. A chatbot's use of language is theorized to produce attributions of human-likeness via the psychological process known as anthropomorphism (Kiesler, Powers, Fussell, & Torrey, 2008; Nass, 2004).

Third, the present studies ask, for whom is the anthropomorphism of a customer service chatbot most important? The moderating role of a consumer's need for human interaction (NFHI) was considered. A number of studies have cited consumers' preference for interpersonal interaction as the basis for rejecting a new self-service technology (SST) (Collier & Kimes, 2012; Meuter, Ostrom, Bitner, & Roundtree, 2003). Individuals are thought to have varying levels of preference for social interaction. Many consumers consider interaction with a firm's employees as a positive dimension of service delivery. Lee and Lyu (2016) demonstrate that those with a high need for human interaction reported unfavourable attitudes towards using SSTs. Therefore, our studies investigate whether anthropomorphic chatbots might interact with a consumer's need for human interaction, in subsequent adoption appraisals.

These studies contribute to the literature given there is little prior research into anthropomorphism or automated sociality as predictors of SST adoption. The findings advance our understanding of miscommunication and anthropomorphism within the natural language domain, whilst providing managerial insights in preparation for the rollout of next-generation interfaces.

2. Conceptual background

2.1. Self-Service technology adoption intent

Technology has rapidly changed the nature of service delivery. Many high-touch and low-tech customer service operations have been overhauled so that technology either supports or supplants the human employee (Wang, Harris, & Patterson, 2013). From a firm's perspective, deploying SSTs offers a number of benefits. First, SSTs can increase efficiency and customer satisfaction (Huang & Rust, 2013; Lee, 2014). Second, an SST can standardize service delivery (Selnes & Hansen, 2001). Third, because an SST may supplement or act as a substitute for the human employee, empirical research has linked investment in SSTs with a firm's positive financial performance (Hung, Yen, & Ou, 2012) and an increase in stock price (Yang & Klassen, 2008). Therefore, Meuter, Bitner, Ostrom, and Brown (2005) talk of the tremendous lure of automating service delivery. However, it is not the act of deploying SSTs that delivers benefits to the firm. Rather, firms enjoy the benefits once consumers try the SST and commit to future use (adoption). Thus, adoption is often the focus of SST studies and is treated as the dependent variable in a range of theoretical models.

2.2. Hypothesis development

2.2.1. Chatbot behavior and adoption: Error-free versus error versus clarification

There is a high degree of variance in chatbot quality. Our studies compare a hypothetically perfect chatbot (error-free) to a chatbot which struggles to infer meaning and thus seeks clarification (clarification) to a third chatbot which produces an error in comprehension (error). Predicting that the error-free chatbot will outperform the error-producing chatbot is straightforward. However, we predict that the chatbot seeking clarification will receive similar adoption scores as the error-free chatbot. This is because users of a clarification seeking chatbot seems more humanlike and users still receive an appropriate response from the chatbot, despite the additional effort required to respond to the clarification request. In a customer service context, the consumer and firm are ultimately negotiating the specifics of a future transaction, so clarification by the chatbot may be seen as due care and attention to the needs of the customer.

2.2.2. Chatbot behavior, anthropomorphism & adoption

The error-free chatbot offers no indication that it is anything but human. It correctly interprets all human utterances and responds with relevant and precise humanlike utterances of its own. Commercial examples of error-free chatbots do not presently exist.

The second chatbot is referred to as the error chatbot. Chatbots can produce errors in any number of ways. This study focused on chatbot errors in communication (i.e. deducing meaning) as opposed to technical errors. Scheutz, Schermerhorn, and Cantrell (2011) identify the failure to maintain contextual awareness as a very common source of chatbot communication error. This is because the development of conversational software which can logically infer meaning from context is difficult. Note that context in this application refers to preceding utterances within a conversation as opposed to psychological, relational or cultural context. A chatbot producing this type of error would violate the Gricean maxim of relevance, which states that a "partner's contribution should be appropriate to immediate needs at each stage of the transaction" (Grice, 1975, p. 47). Adherence to Gricean maxims is critical for successful communication as communication requires a shared set of assumptions between parties (Grice, 1975). These assumptions revolve around each party's ability to produce and deduce meaning. When a chatbot violates these assumptions, it should be perceived as presenting less humanlike cues.

Again, the idea that an error-free chatbot will be perceived as more humanlike and more readily adopted than an error producing chatbot is intuitive. Our research is designed to examine the role of clarification as a means to bridge the gap between the ideal and current chatbot states.

The clarification chatbot does not have sufficient intelligence to correctly interpret all human utterances on the first parse. However, the chatbot is intelligent enough to identify the source of the miscommunication, known as a trouble source (Schegloff, 1992) and seek clarification. This clarification is similar to Fromkin (1971) concept of conversation repair. Either a message sender or receiver can seek clarification (Hutchby & Wooffitt, 2008). A message sender, sensing the potential for miscommunication may rephrase their previous statement, following an utterance such as "What I mean is…". Alternatively, the message receiver may seek clarification of a particular message using "huh?", "what?" or an apology-based format such as "I'm sorry. What do you mean?" (Robinson, 2006).

Seeking clarification has been linked to social coordination in that it demonstrates one's ability to use synchronized interaction strategies, such as turn-taking and role-switching (Corti & Gillespie, 2016; Kaplan & Hafner, 2006). Corti and Gillespie (2016) discuss clarification as being a fundamental component of intersubjective effort, where intersubjectivity is defined as shared meaning, co-created and co-existing within two or more conscious minds (Stahl, 2015). Thus, a chatbot which seeks clarification is perhaps as humanlike as an error-free chatbot, given that it can identify a trouble source and demonstrates intersubjective effort. A chatbot which seeks clarification may be as readily adopted as an error-free chatbot, given clarification seeking is a natural part of interpersonal communication. As such, we propose:

H1a. (Adoption intent): The clarification chatbot will receive the same adoption scores as the error-free chatbot.

H1b. (Adoption intent): The clarification chatbot and the error-free

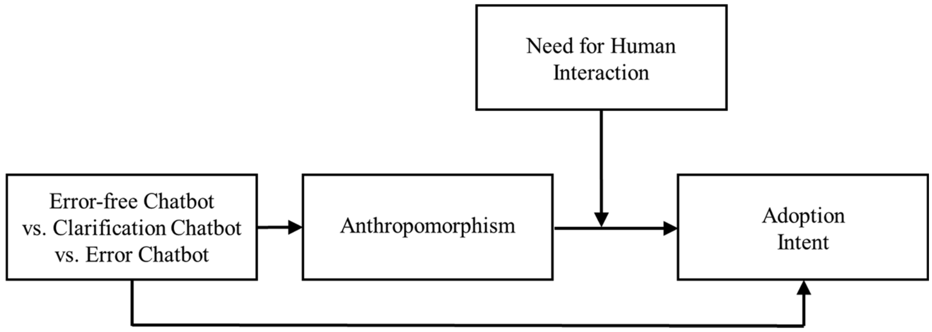

Fig. 1. Model of the relationships explored (moderated mediation).

chatbot will score higher for adoption than the error chatbot.

H2a. (Anthropomorphism): The clarification chatbot will receive the same anthropomorphism scores as the error-free chatbot.

H2b. (Anthropomorphism): The clarification chatbot and the error-free chatbot will score higher for anthropomorphism than the error chatbot.

2.2.3. The mediating role of anthropomorphism

The anthropomorphism of products and brands is a popular marketing strategy (Aggarwal & McGill, 2012). The effects of anthropomorphism in a commercial context have received considerable interest from academics in recent years (MacInnis & Folkes, 2017). Previous research has linked anthropomorphism with product design preferences (Landwehr, McGill, & Herrmann, 2011), the consequences of product failure (Puzakova, Kwak, & Rocereto, 2013) and trust in promotional messages (Touré-Tillery & McGill, 2015).

Humans anthropomorphize non-human entities from a young age (Barrett, Richert, & Driesenga, 2001; Lane, Wellman, & Evans, 2010). This is because anthropomorphism ties into a number of motivations that are central to the human experience. Epley, Waytz, and Cacioppo (2007) categorize these motivations as sociality motivation and effectance motivation. Sociality motivation refers to the need to establish social connections, which promotes co-operation. People who are more socially connected are said to have a lower level of sociality motivation. In support of this idea, chronically lonely individuals are more likely to anthropomorphize technology (Epley, Akalis, Waytz, & Cacioppo, 2008). Effectance motivation can be conceptualized as the desire to understand and master one's environment. Epley et al. (2007) present effectance as a strategy to reduce uncertainty. By anthropomorphising the non-human, a person can anticipate the entity's behavior, increasing the odds of a favourable interaction.

Both sociality and effectance motivation reside inside the human mind. That is, they are qualities of the person who is anthropomorphizing. However, the present studies focus on Epley et al. (2007) third antecedent of anthropomorphism, known as elicited agent knowledge (EAK). EAK refers to the actual humanlike cues projected by the non-human agent. A person interacting with a non-human entity will examine the entity's features and behavior to check for perceived similarity. As Eyssel, Kuchenbrandt, Hegel, and De Ruiter (2012) explain, access to EAK is triggered by similarity between the anthropomorphic object and the anthropocentric knowledge structures within the perceiver. Waytz et al. (2010) summarize in saying that anthropomorphism appears to originate within both the perceiver and the perceived. According to this theory, manipulation of a chatbot's behavior will affect a consumer's access to EAK and therefore anthropomorphic perceptions.

We propose that anthropomorphism contributes to explaining the relationship between chatbot behavior and adoption intent. A great deal of previous research has linked technological cues to anthropomorphic perceptions (Kiesler et al., 2008; Nass, 2004), supporting the chatbot behavior – anthropomorphism relationship. Recent research has linked chatbot anthropomorphism to positive consumer evaluations, where chatbots producing more human-like cues generated stronger emotional connections with the firm (Araujo, 2018). This latest research alludes to the anthropomorphism – adoption relationship. We propose;

H3. (Anthropomorphism as mediator): Chatbot type (error-free, clarification and error) will have an indirect effect on adoption intent through anthropomorphism.

2.2.4. The moderating role of need for human interaction

Several studies have examined consumers' need for human interaction (NFHI) as a predictor of SST attitudes and adoption intent (Collier & Kimes, 2012; Dabholkar & Bagozzi, 2002; Lee & Lyu, 2016). The construct is defined as the desire for human contact by the consumer during a service experience (Dabholkar, 1996).

Those with a high NFHI are less satisfied by SST's (Meuter, Ostrom, Roundtree, & Bitner, 2000; Meuter et al., 2003). For example, consumers who avoid scanning their own groceries cite the desire to interact with store employees as a major consideration (Dabholkar, Bobbitt, & Lee, 2003). Lee and Lyu (2016) connect this need for human interaction to hedonic attitudes, suggesting these consumers derive a social benefit from interpersonal relations in service delivery. As previously discussed, a highly anthropomorphic chatbot may be perceived as human enough to trigger consumer's perceptions of a human actor. As such, a humanlike chatbot might be humanlike enough to satisfy a consumer's need for human interaction. The following hypothesis is proposed:

H4. (NFHI as moderator): As a consumer's need for human interaction increases, the strength of the relationship between anthropomorphic perceptions and adoption intent also increases.

The conceptual model developed for this study is shown below in Fig. 1.

3. Study one methodology

3.1. Sample

A sample of 190 Americans (53% male, M = 37.2, SD = 11.6) from Amazon's Mechanical Turk (MTurk) website elected to participate in the study. MTurk is an empirically sound sampling technique (Chandler & Shapiro, 2016). Power analysis for Analysis of Variance (ANOVA) was considered, given an alpha of 0.05, a power of 0.80 and a medium effect size (f^2 = 0.15) (Faul, Erdfelder, Buchner, & Lang, 2009). Based on these assumptions, the sample size was determined to be sufficient. One case was removed due to incomplete data.

3.2. Stimulus materials

The stimulus consisted of a pre-recorded animation of a human-chatbot interaction, contextualized around the automated booking of hotel accommodation. Three animations were prepared: the error-free condition in which no miscommunication occurs, the animation in which the chatbot seeks clarification to confirm the meaning of a user's input and the error animation. In all three animations, the human user asks the chatbot if the hotel room includes a washer and dryer. The error free chatbot responds in the affirmative and advises the user that an onsite laundry service is also available. In this way, the error-free chatbot correctly interprets the term "*dryer*" as related to laundry when used in combination with the term "*washer*". The clarification chatbot does not make this connection in the first parse. Instead the clarification chatbot asks if the user is referring to a "*clothes dryer*" or a "*hair dryer*". Finally, the error chatbot fails to interpret the user's utterance entirely. All three animations conclude in the same way – with the user making a reservation. A full copy of the script used for each condition is provided in Appendix A.

The animations used in this study were designed to look and feel like an interaction occurring on Facebook Messenger, given there are over 100,000 text-based chatbots operating on Messenger today (Constine, 2017; Johnson, 2017). The chatbot was given a fictitious name (Beachside Hotel) with a custom logo. The animation was set inside a wireframe image of a white iPhone 6 to add to realism. The three animations were identical, with the exception of the experimental manipulation. Care was taken to ensure the three animations were consistent in all other respects. For example, the pause time between the human input and the chatbot response was the same across all conditions. Total animation duration was 1 min and 31 s. A screenshot of the animation is presented in the Web Appendix.

3.3. Procedure

Participants were randomly assigned to one of the three conditions and began by providing demographic data and responding to the questions measuring their need for human interaction. Next, participants were provided with a basic description of what a chatbot is and does. They were expressly told that they would be watching an animation depicting a human-chatbot interaction – thus participants were never under any illusion that both interlocutors were human.

Participants were provided with a diagram as shown in Fig. 2, such that they could interpret the animation – identifying the human actor and the chatbot. Finally, participants watched the animation and provided responses to the survey items measuring adoption intent and anthropomorphic perceptions.

3.4. Measures

Adoption: Participants were asked to think about the human-chatbot interaction and indicate their agreement with the following statement: "I would use a chatbot like this to book hotel accommodation". This wording was modified from similar single-item measures of SST adoption (Lee, Castellanos, & Choi, 2012).

Anthropomorphism: The instrument used was a modified version of the Godspeed Questionnaire (Bartneck, Kulic, Croft, & Zoghbi, 2009), which provides a set of semantic differential items to measure anthropomorphism of social robots. The instrument includes 5 items (α = 0.91): (a) fake – natural, (b) machinelike – humanlike, (c) artificial – lifelike, (d) unconscious – conscious and (e) communicates inelegantly – communicates elegantly.

Need for Human Interaction (NFHI): This variable was measured using Dabholkar (1996) four-item scale (α = 0.89). Items such as "I like interacting with the person who provides the service" focus on the social aspect of interacting with a human service employee.

All three measures used 7-point response options. All items were monotone, avoiding extreme or suggestive language. Demographic data regarding age, gender and education was also collected. Survey items are provided in the Web Appendix.

3.5. Results

3.5.1. Preliminary analysis

In order to examine whether the sample characteristics were similar across three conditions. ANOVA was used for age and education and a chi-square test was run for gender. No significant group differences were found across the three conditions: Age (F (2,185) = 0.80, p = .41), education (F (2,185) = 1.65, p = .20), and gender (χ^2 = 0.97, p = .62). In addition, multiple regression showed that the demographic variables had no relationships with the dependent measures: anthropomorphism (-1.1 < t_s < 0.65; p_s > 0.27) and adoption

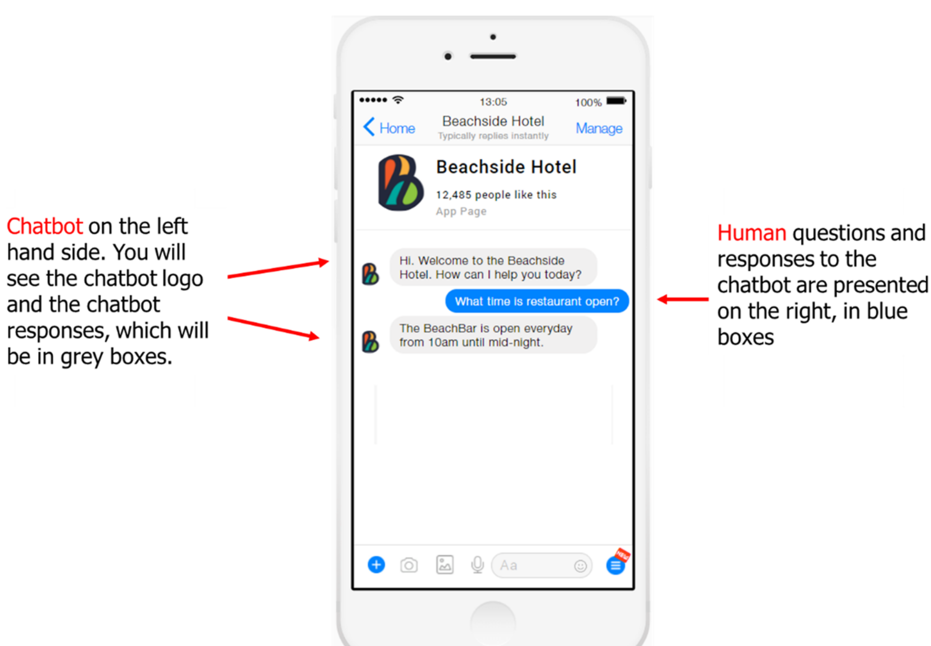

Fig. 2. Image provided to participants as part of the instructions.

Table 1
Correlation matrix.

Variable	Mean (SD)	1	2	3	4
1. Adoption intent	5.39 (1.04)	1			
2. Anthropomorphism	4.74 (1.40)	0.516**	1		
3. NFHI	4.07 (1.47)	−0.256**	−0.030	1	
4. Age	37.21 (11.60)	−0.034	−0.033	0.180*	1

Note: Correlations are based on entire sample, not split by condition. * $p < .05$, ** $p < .001$. Correlation tables split by condition are provided in the web appendix.

intent (-0.26 $< t_s <$ 0.14; $p_s >$ 0.75). Hence, no further analyses included the demographic variables. A correlation matrix is provided below in Table 1, while correlation tables, split by condition are available in the Web Appendix.

Before testing the hypotheses, we examined the normality of adoption and anthropomorphism. The values for skewness and kurtosis were in the acceptable range (between −2 and + 2). Variance inflation factors (VIF) were assessed for multicollinearity and were acceptable. Skewness, kurtosis and VIF values are presented in the Web Appendix.

3.5.2. Hypothesis testing

3.5.2.1. Analysis of variance to test H_1 and H_2. A one-way between groups analysis of variance (ANOVA) was used to investigate the impact that chatbot behavior (error-free, clarification and error) had on adoption intent (H_1) and anthropomorphism (H_2). The Leven's test showed that the assumption of homogeneity of variance was met for the adoption variable (p = .12). However, the anthropomorphism variable violated the assumption, suggesting that the variances in three conditions were not homogeneous (p = .011). Therefore we used Welch's ANOVA for anthropomorphism.

The results showed that there was a statistically significant difference in both instances: adoption intent (F (2,186) = 10.79, $p < .001$) and anthropomorphism (F (2,119.91) = 9.38, $p < .001$). Table 2 presents the means and standard deviations in each condition. Planned contrasts revealed that the error chatbot (M = 4.73) produced significantly lower adoption scores than the error-free chatbot (M = 5.63, t = -3.77, $p < .001$) and the clarification chatbot (M = 5.75, t = -4.24, $p < .001$). The same pattern was observed with regards to anthropomorphism. The error chatbot (M = 3.99) produced significantly lower anthropomorphism scores than the error-free chatbot (M = 5.09, t = -4.91, $p < .001$) and clarification chatbot (M = 5.11, t = -4.79, $p < .001$). However, there were no significant differences between scores for the error-free and clarification chatbots: anthropomorphism (t = -0.11, p = .91) and adoption intent (t = -0.46, p = .64).

We hypothesized that the clarification chatbot would receive the same adoption and anthropomorphism scores as the error-free chatbot because it would be perceived as demonstrating intersubjective effort in the identification of a trouble source. As per the analyses, we did not find a statistically significant difference between the error-free and clarification conditions. Thus, both H_1a and H_2a were supported. Both the error-free and clarification chatbots outperformed the error chatbot, thus H_1b and H_2b were supported.

3.5.2.2. Mediation analysis to test H_3. Hayes PROCESS Model 4 with 5000 bootstrapped samples (Hayes, 2017) was used to test anthropomorphism as a mediator between chatbot behavior (IV) and adoption (DV). Traditional mediation literature focuses on models with dichotomous or continuous independent variables, however, Hayes and Preacher (2013) provide a tutorial for testing mediation and moderation with a multi-categorical independent variable as was the case here with three experimental conditions. The PROCESS Macro includes an option to specify the independent variable as multi-categorical, which automatically re-coded the three experimental conditions into two dummy coded variables, X_1 and X_2, such that the error-condition became the baseline (X_1 error = 0, error-free = 1 and X_2 error = 0, clarification = 1). Please note, all coefficients in this and subsequent analyses are unstandardized.

Chatbot behavior significantly predicts anthropomorphism, X_1 (b = 1.07, SE = 0.23, $p < .001$), X_2 (b = 1.10, SE = 0.23, $p < .001$). Anthropomorphism significantly predicts adoption intent (b = 0.46, SE = 0.22, $p < .001$). The indirect effects (X_1 / X_2 - > anthropomorphism - > adoption) were also significant, X_1 (b = 0.49, $BootSE$ = 0.14, 95% CI = 0.25, 0.81), X_2 (b = 0.501, $BootSE$ = 0.16, 95% CI = 0.23, 0.85). The data supports anthropomorphism as mediator (H_3) as the 95% confidence intervals do not span zero.

3.5.2.3. Moderated mediation to test H_4. Moderated mediation analysis was performed to assess (a) anthropomorphism as a mechanism for the relationship between chatbot behavior and adoption intent and (b) NFHI as a boundary condition for the relationship between anthropomorphism and adoption. PROCESS Model 14 with 5000 bootstrapped samples (Hayes, 2017) and mean centered variables were used.

As shown in Table 3, the chatbots behavior significantly predicted anthropomorphism; X_1 (b = 1.10, SE = 0.17, $p < .001$), X_2 (b = 1.13, SE = 0.23, $p < .001$). The participants need for human interaction was significantly related to adoption intent; (b = -0.22, SE = 0.05, $p < .001$) while the interaction term (anthropomorphism × need for human interaction) was also significant; (b = 0.10, SE = 0.04, $p < .01$).

The index of moderated mediation (Hayes, 2017), which is the omnibus test of whether the indirect effect varies across levels of the moderator was tested. The 95% confidence intervals did not span 0 for either X_1 ($Index$ = 0.11, $BootSE$ = 0.05, 95% CI = 0.02, 0.21) or X_2 ($Index$ = 0.11, $BootSE$ = 0.05, 95% CI = 0.03, 0.22), thus significant moderation occurred, providing support for H_3 and H_4. Anthropomorphism mediates the chatbot condition - > adoption relationship, while NFHI moderates the anthropomorphism - > adoption relationship. The interaction is illustrated in Fig. 3.

As per Fig. 3, as a consumer's need for human interaction increases, the effect of anthropomorphism on adoption intent also increases. This suggests that consumers who derive pleasure from interacting with a human service agent also derive pleasure from interacting with a chatbot, providing it is perceived as humanlike.

Study one appears to support the theoretical model presented. The next step was to test the model using genuine human-chatbot interaction. This study had participants evaluate the human-chatbot interaction of a hypothetical third party. Study two was designed so that participants could report their perceptions of a chatbot that they themselves had interacted with. Furthermore, we wanted to measure perceived usefulness and ease of use as covariates, in order to further establish anthropomorphism as a mediator between chatbot behavior

Table 2
Mean (SD) comparison of anthropomorphism and adoption scores across the three conditions.

Variable	Condition 1 (Error-free, n = 63)	Condition 2 (Clarification, n = 63)	Condition 3 (Error, n = 63)
Adoption Intent	5.63 (1.14)[a]	5.75 (1.35)[b]	4.73 (1.51)[ab]
Anthropomorphism	5.09 (1.03)[a]	5.11 (1.42)[b]	3.99 (1.43)[ab]

Note: The same letter across the three conditions indicates a significant ($p < .001$) mean difference in each dependent variable.

Table 3
Results of moderated mediation analysis.

Predictor	b	p	95% CI	
	Outcome: Anthropomorphism			
X_1 (error "0" vs. error-free "1")	1.10	< 0.01	0.64	1.56
X_2 (error "0" vs. clarification "1")	1.13	< 0.01	0.67	1.60
	Outcome: Adoption Intent			
X_1 (error "0" vs. error-free "1")	0.46	0.03	0.04	0.88
X_2 (error "0" vs. clarification "1")	0.45	0.03	0.03	0.88
Anthropomorphism	0.45	< 0.01	0.33	0.58
NFHI	-0.22	< 0.01	-0.33	-0.11
Anthro × NFHI	0.10	< 0.01	0.03	0.17
Index of Moderated Mediation	Index	BootSE	95% CI	
X_1 (error "0" vs. error-free "1")	0.11	0.05	0.02	0.21
X_2 (error "0" vs. clarification "1")	0.11	0.05	0.03	0.22

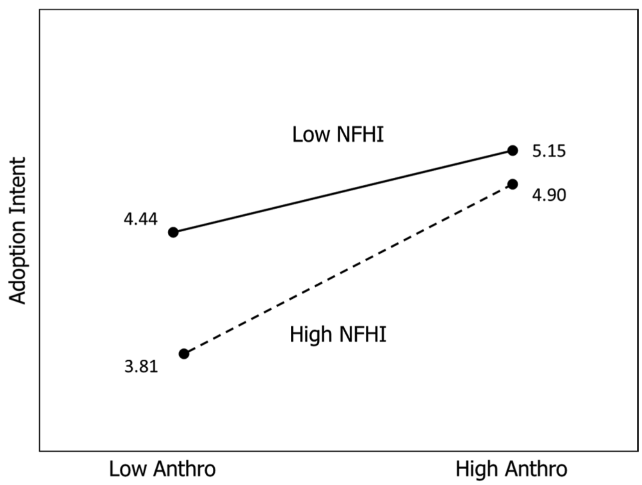

Fig. 3. Pattern of interaction.

and adoption intent.

4. Study two methodology

4.1. Sample

A sample of 200 Americans (54% male, $M = 37.2$, $SD = 9.44$) from MTurk elected to participate in the study. Three cases were removed due to incomplete data.

4.2. Stimulus material

Participants were required to hold a scripted conversation with one of three purpose built chatbots (error-free, clarification and error versions). The chatbots were designed to assist users in enquiring about overseas train travel in Europe. They represented the fictitious Euro-Rail Network and were given the name Eve. The chatbots were built using an online platform, FlowXO which doesn't require coding experience, making replication straightforward.

The chatbots were designed to greet the human user and respond to user input. As per study one, the three chatbots were identical, with the exception of the experimental manipulation. For example, when the human user asks the chatbot "*how much baggage can I bring*?"; the error-free chatbot explains there are no baggage restrictions, the clarification chatbot asks if the user is referring to "baggage restrictions" and then explains there are no restrictions, while the chatbot designed to make a mistake apologizes for failing to understand the user's input. Prior to running the study, we had 10 laypeople (students) assess the conversation script for realism.

In order to eliminate extraneous variables, participants were asked to stick to the scripted questions they were provided. Transcripts of participant interactions with the chatbots were reviewed on the FlowXO platform to ensure that this occurred and exposure to the stimulus was consistent. A full copy of the script used for each condition is provided in Appendix B, while a screenshot of the chatbot is available in the Web Appendix. Total chatbot interaction time was approximately 90 s.

4.3. Procedure

Participants were randomly assigned to one of the three conditions and began by providing demographic data and responding to the questions measuring their need for human interaction. Next, participants were given an explanation of what a chatbot is and some task related information. They were told that they were planning a holiday to the United Kingdom and were hoping to explore via the rail network. Based on this premise, they were to imagine they had engaged this chatbot in order to learn more about train travel options.

Participants were provided with the script they were to use during their chatbot conversation. They clinked a link to access the chatbot, hosted on the FlowXO servers and asked to enter their scripted line, observe the chatbots response and enter the next scripted line until the chatbot advised them that the conversation was over. Following the chatbot interaction, participants provided responses to the survey items measuring adoption intent and a range of mediating variables as discussed in the next section.

4.4. Measures

Adoption, anthropomorphism ($\alpha = 0.94$) and the need for human interaction ($\alpha = 0.88$) were measured as described in study one. Additional variables were measured as follows:

Perceived Usefulness and Perceived Ease of Use: These variables were to be measured because previous meta-analysis has identified them as key mediators in SST acceptance (Blut, Wang, & Schoefer, 2016). By adding them to this second study, they could be used as covariates in the analysis to determine if anthropomorphism uniquely contributes to adoption intent. Both variables were measured with three items each (useful $\alpha = 0.93$, ease of use $\alpha = 0.90$). The items were taken from Curran and Meuter's technology adoption studies (Curran & Meuter, 2005), which themselves were derived from Davis (1989). All measures are provided in the Web Appendix and were taken on a 7-point Likert scale.

4.5. Results

As per study one, ANOVA was used to confirm that average age and gender distribution was similar across the three conditions. In addition, regression analysis showed that the demographic variables had no relationship with the other variables of interest. Skewness and kurtosis was assessed as per study one and deemed acceptable. A correlation matrix is provided in Table 4. In order to assess the variables for

Table 4
Correlation matrix.

Variable	Mean (SD)	1	2	3	4	5	6
1. Adoption	5.30 (1.5)	1					
2. Anthro	4.28 (1.5)	0.681^{**}	1				
3. NFHI	4.13 (1.5)	0.262^{**}	0.256^{**}	1			
4. Usefulness	5.24 (1.4)	0.834^{**}	0.706^{**}	0.205^{**}	1		
5. Ease of use	6.01 (1.0)	0.583^{**}	0.395^{**}	0.059	0.562^{**}	1	
6. Age	37.28 (9.2)	0.149*	0.164*	0.164*	0.055	0.089	1

Note: Correlations are based on entire sample, not split by condition. * $p < .05$, ** $p < .001$. Correlation tables split by condition are provided in the web appendix.

Table 5
Mean (*SD*) comparison of scores across the three conditions.

Variable	Condition 1 (Error-free, n = 65)	Condition 2 (Clarification, n = 66)	Condition 3 (Error, n = 66)
Adoption Intent	6.05 (0.99) [a]	5.91 (1.11) [b]	3.94 (1.74) [ab]
Anthropomorphism	4.90 (1.46) [a]	4.71 (1.36) [b]	3.20 (1.34) [ab]
Usefulness	5.75 (1.12) [a]	5.74 (1.00) [b]	4.21 (1.47) [ab]
Ease of Use	6.32 (0.77) [a]	6.20 (0.89) [b]	5.52 (1.25) [ab]

Note: The same letter across the three conditions indicates a significant ($p < .001$) mean difference in each dependent variable.

multicollinearity, VIF values were calculated and a confirmatory factor analysis was run to assess convergent and discriminant validity. All values were within tolerances and are presented in the Web Appendix. However, due to the high correlations between perceived usefulness and anthropomorphism, perceived usefulness was removed from the mediation analysis to test H_3.

4.5.1. Hypothesis testing

4.5.1.1. Analysis of variance to test H_1 & H_2. ANOVA was used to investigate the impact that chatbot behavior (error-free, clarification and error) had on adoption intent and the potential mediators (anthropomorphism, usefulness and ease of use).

The results showed that there were statistically significant differences in all instances: adoption (F (2194) = 51.81, $p < .001$), anthropomorphism (F (2194) = 29.27, $p < .001$), usefulness (F (2194) = 35.08, $p < .001$) and ease of use (F (2194) = 12.31, $p < .001$). Table 5 presents the means and standard deviations in each condition. Planned contrasts found the same pattern as shown in study one, i.e. the error chatbot produced significantly lower scores for all variables than the error-free chatbot and the clarification chatbot, however there were no significant differences between scores for the error-free and clarification chatbots. Thus, as in study one, both H_1a & H_1b and H_2a & H_2b were supported by study two.

4.5.1.2. Mediation analysis to test H_3. Hayes PROCESS Model 4 with 5000 bootstrapped samples (Hayes, 2017) was used to test anthropomorphism as a mediator between chatbot behavior (IV) and adoption (DV). The three experimental conditions were dummy coded into two variables as per study one (X_1 error = 0, error-free = 1, X_2 error = 0, clarification = 1). Chatbot behavior significantly predicts anthropomorphism, X_1 (b = 1.70, SE = 0.24, $p < .001$), X_2 (b = 1.51, SE = 0.24, $p < .001$). Anthropomorphism predicts adoption intent (b = 0.53, SE = 0.06, p < .001). Finally, the indirect effects support mediation, X_1 - > Anthro - > Adoption (b = 0.91, SE = 0.18, 95% CI = 0.58, 1.28), X_2 - > Anthro - > Adoption (b = 0.81, SE = 0.16, 95% CI = 0.50, 1.14).

Extending upon study 1, the model was re-run, using perceived ease of use (EU) as a covariate. Perceived usefulness was dropped due to potential multicollinearity concerns. All paths remained significant, X_1 - > Anthro - > Adoption (b = 0.60, SE = 0.16, 95% CI = 0.32, 0.94), X_2 - > Anthro - > Adoption (b = 0.54, SE = 0.14, 95% CI = 0.29, 0.83). A diagram of the mediation model and full statistical analysis is available in the Web Appendix.

Therefore, study two suggests that anthropomorphism is an appropriate mediator between chatbot behavior and adoption. Furthermore, the results suggest that anthropomorphism explains unique variance in adoption scores, beyond ease of use from the extant literature.

Finally, in order to understand why participants prefer anthropomorphic chatbots, a serial mediation model was tested. The model tested whether the chatbot behavior - > adoption relationship was mediated by anthropomorphism and ease of use in sequence. Hayes PROCESS Model 6 was used. All paths were significant, suggesting participants prefer anthropomorphic chatbots because they are easier to use, increasing adoption intent. Total indirect effects were as follows; X_1 - > Anthro - > EU - > Adoption (b = 0.17, SE = 0.05, 95% CI = 0.07, 0.29), X_2 - > Anthro - > EU - > Adoption (b = 0.15, SE = 0.05, 95% CI = 0.06, 0.26). Values for all paths in the serial mediation model are included in the Web Appendix.

Table 6
Results of moderated mediation analysis.

Predictor	b	p	95% CI	
	Outcome: Anthropomorphism			
X_1 (error "0" vs. error-free "1")	1.70	< 0.01	1.22	2.18
X_2 (error "0" vs. clarification "1")	1.51	< 0.01	1.01	2.00
	Outcome: Adoption Intent			
X_1 (error "0" vs. error-free "1")	1.18	< 0.01	0.76	1.59
X_2 (error "0" vs. clarification "1")	1.14	< 0.01	0.73	1.54
Anthropomorphism	0.52	< 0.01	0.40	0.63
NFHI	-0.11	0.04	-0.21	-0.01
Anthro × NFHI	0.09	< 0.01	0.02	0.15
Index of Moderated Mediation Index BootSE 95% CI				
X_1 (error "0" vs. error-free "1")	0.15	0.06	0.04	0.26
X_2 (error "0" vs. clarification "1")	0.13	0.05	0.03	0.25

4.5.1.3. Moderated mediation analysis to test H_4. Hayes PROCESS Model 14 with 5000 bootstrapped samples (Hayes, 2017) was used to test the effect of chatbot behavior on adoption intent, mediated by anthropomorphism with an individuals need for human interaction moderating the relationship between anthropomorphism and adoption. The analysis used mean centred variables.

As shown in Table 6, the chatbots behavior significantly predicted anthropomorphism: X_1 (b = 1.70, SE = 0.24, $p < .001$), X_2 (b = 0.1.51, SE = 0.24, $p < .001$). The participants need for human interaction was significantly related to adoption intent; (b = -0.11, SE = 0.05, p = .04) while the interaction term (anthropomorphism × need for human interaction) was also significant; (b = 0.09, SE = 0.03, p < .01). The index of moderated mediation (Hayes, 2017), which is the omnibus test of whether the indirect effect varies across levels of the moderator was tested. The 95% confidence intervals did not span 0 for either X_1 (*Index* = 0.15, *BootSE* = 0.06, 95% CI = 0.04, 0.26) or X_2 (*Index* = 0.13, *BootSE* = 0.05, 95% CI = 0.03, 0.25), thus significant moderation occurred, providing additional support for H_4. The interaction is illustrated in Fig. 4.

The interaction presented in Fig. 4 is similar to the data collected in study one. Individual's low in the need for human interaction are more likely to adopt a customer service chatbot. Furthermore, for those high in NFHI, anthropomorphism increases the likelihood of chatbot adoption.

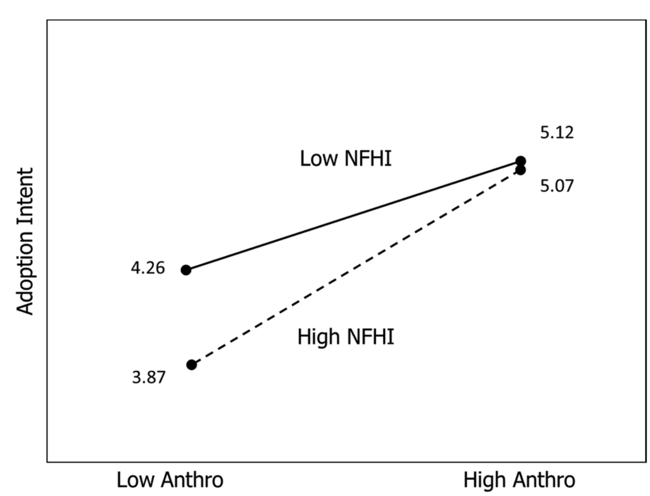

Fig. 4. Pattern of interaction.

5. Discussion

These studies were designed to investigate a potential relationship between chatbot (mis)communication and adoption intent. Effects were theorized to occur as a result of a participant's access to elicited agent knowledge (EAK), a known component of anthropomorphic perceptions.

Across both studies, exposure to the error-free and clarification conditions produced similar anthropomorphism and adoption scores. This suggests that a seeking clarification in order to repair or avoid miscommunication is a means of bridging the gap between ideal (error-free) chatbots and current (error prone) chatbots. Furthermore, the reduction in adoption scores for the error chatbot appears to be the result of a decrease in anthropomorphism. Finally, the relationship between anthropomorphism and adoption intent for the error chatbot is moderated by a consumer's need for human interaction. Anthropomorphism is more positively related to adoption when one's need for human interaction is high. Anthropomorphic chatbots may be more readily adopted because they mimic human service agents and are thus perceived as easier to use.

5.1. Theoretical contribution

These results suggest that consumers are unlikely to reject a customer service chatbot for simply seeking clarification – providing that the clarification repairs any potential miscommunication. This occurs despite the additional effort expended by the user in order to respond to the clarification request. Thus, a chatbot which is humanlike enough to identify potential miscommunication (as opposed to completely error-free) appears to be sufficient in customer service scenarios. It appears that human-chatbot miscommunication is not something to be avoided at all costs. This makes sense conceptually, given miscommunication between two human interlocutors is a frequent occurrence. The ability to resolve miscommunication appears to be as effective as avoiding it. In a customer service scenario, some consumers may prefer a chatbot which seeks clarification, given that the parties are essentially negotiating goods and services to be supplied under specific terms.

Conversely, unresolved errors reduce anthropomorphism and adoption intent. Consistent with Gricean maxims (1975), the type of error tested (failure to maintain contextual awareness) appears central to perceptions of humanness. Gricean maxims are said to generate implicature, where implicature is what is suggested as opposed to what is expressly stately (Blackburn, 1996). In failing to maintain contextual awareness, the chatbot may have provided the implicature that it has non-human cognition, given that relevance comes naturally to human interlocutors.

In summary, a chatbot seeking clarification appears to provide (a) as much access to elicited agent knowledge as a perfect or error-free chatbot and (b) more access to elicited agent knowledge than a chatbot which fails to maintain contextual awareness. As discussed in Section 2.2, this is because the basis of EAK is anthropocentric knowledge structures. Seeking clarification is normal between two humans – in fact it is a very humanlike behavior as it demonstrates social coordination (Kaplan & Hafner, 2006) and intersubjective effort (Corti & Gillespie, 2016). Thus, when a chatbot seeks clarification it triggers schemata related to humanness. Conversely, a chatbot which fails to maintain contextual awareness does not trigger access to EAK. This is likely because humans are innately skilled in crafting and interpreting context, which requires little effort (Flowerdew, 2014).

These results also represent a contribution to the SST literature. A substantial number of empirical studies have examined the adoption of SSTs; however, systematic literature reviews by Blut et al. (2016) and Hoehle, Scornavacca, and Huff (2012) do not mention anthropomorphism or any other variables related to automated social presence. It seems likely that anthropomorphism will play an ever-increasing role in SST adoption, as natural language systems attempt to replicate interpersonal service delivery.

5.2. Managerial implications

Previous research has linked anthropomorphism of a firm's products and branding to positive attitudinal responses from consumers (MacInnis & Folkes, 2017). This research contributes by connecting anthropomorphism to behavioural intentions (adoption) with regards to self-service technologies. Increasing anthropomorphism in this context appears to be a risk-free strategy. Consumers low in the need for human interaction (NFHI) do not appear to be negatively affected by a chatbot high in humanlike cues. This is interesting given previous research posited that those low in NFHI (high in the need for independence) often prefer to avoid interpersonal interaction in consumption contexts. Conversely, those consumers high in the NFHI reported increased adoption scores when anthropomorphism via access to EAK was high. This interaction effect suggests that anthropomorphic chatbots can be human enough to satisfy the social and hedonic desires of those high in NFHI. However, practitioners should note that anthropomorphism may result in unintended consequences. Previous studies have linked the anthropomorphism of a technology to a user's expectations of its capabilities (Knijnenburg & Willemsen, 2016; Nowak & Biocca, 2003). A very humanlike chatbot may be expected to have very humanlike cognition. This may result in consumers overestimating a chatbot's abilities and subsequently being disappointed or frustrated when those expectations are violated. In a customer service context, this overestimation is important, as the gap between expectations and experience is a major driver of customer dissatisfaction (Hill & Alexander, 2006).

5.3. Limitations and future research

Our work has limitations that can seed future inquiry. First, participants in study one were only exposed to an animation of a human-chatbot interaction. This was rectified in study two, however, additional studies featuring genuine human-chatbot interaction would improve generalizability. Second, while we had laypeople assess the realism of the stimulus used in study two, our stimulus material was developed to incorporate the experimental manipulation. It may not represent a typical human-chatbot interaction. Participants in study two were told to stick to the script, in order to control for extraneous variables. This increased internal validity, however future researchers may wish to maximize ecological validity and give users the latitude to phrase their questions and input as they see fit. Finally, participants in study two were told they were interacting with a chatbot, which may have either increased miscommunication salience or raised expectations. Researchers may wish to explicitly measure communication quality - in scenarios where participants are unaware that their communication partner is non-human.

Other researchers may wish to examine the role of anthropomorphism in SSTs designed to assist consumers in complex service scenarios. The service contexts used in these studies (hotel accommodation and train travel) are considered low in credence qualities (Mazaheri, Richard, & Laroche, 2012). Future research may wish to test the relationships identified against service scenarios high in credence qualities, such as medical or legal advice. This is important given chatbots are being tested in the provision of triage medical advice (Burgess, 2017) and legal advice to asylum seekers (Cresci, 2017). Both medical and legal advice are considered high in credence qualities (Mitra, Reiss, & Capella, 1999).

Finally, miscommunication is but one determinant of a dissatisfying chatbot interaction. Future research may wish to consider the role of additional independent variables, such as the lack of a tangible interface or trust in automated systems.

Appendix A:. Transcript of study 1 stimulus used, illustrating experimental manipulation.

	A: Error-free	C: Clarification	B: Error
CB	Hi. Welcome to the Beachside Hotel. How can I help you today?	Hi. Welcome to the Beachside Hotel. How can I help you today?	Hi. Welcome to the Beachside Hotel. How can I help you today?
H	Any rooms left for Independence Day?	Any rooms left for Independence Day?	Any rooms left for Independence Day?
CB	–	Do you mean 4th of July 2018?	–
H	–	Yes	–
CB	We have some rooms still available. How many guests in your party?	We have some rooms still available. How many guests in your party?	We have some rooms still available. How many guests in your party?
H	4 adults	4 adults	4 adults
CB	Ok great. We have the Oceanview room for $209 per night. It sleeps 4 adults in 2 separate rooms, each with a private bathroom.(image of room)	Ok great. We have the Oceanview room for $209 per night. It sleeps 4 adults in 2 separate rooms, each with a private bathroom.(image of room)	Ok great. We have the Oceanview room for $209 per night. It sleeps 4 adults in 2 separate rooms, each with a private bathroom.(image of room)
H	Internet included?	Internet included?	Internet included?
CB	–	Do you mean "wi-fi"?	–
H	–	yes	–
CB	Yes. Rooms include unlimited wi-fi, air-conditioning, cable TV, a full kitchen, daily house-keeping and all the other extra comforts afforded to our valued Beachside guests.	Yes. Rooms include unlimited wi-fi, air-conditioning, cable TV, a full kitchen, daily house-keeping and all the other extra comforts afforded to our valued Beachside guests.	Yes. Rooms include unlimited wi-fi, air-conditioning, cable TV, a full kitchen, daily house-keeping and all the other extra comforts afforded to our valued Beachside guests.
H	Room includes a washer and dryer?	Room includes a washer and dryer?	Room includes a washer and dryer?
CB	–	Do you mean "clothes dryer" or "hair dryer"?	Sorry. I thought we were talking about 'room reservation' on the '4th of July 2018'.
H	–	Clothes dryer	Does the room include a clothes dryer?
CB	Absolutely. We also offer an onsite laundry service upon request.	Absolutely. We also offer an onsite laundry service upon request.	Absolutely. We also offer an onsite laundry service upon request.
H	Reserve the room for me.	Reserve the room for me.	Reserve the room for me.
CB	Done. Click buttons below for copies of invoice and booking confirmation. We look forward to welcoming you and your guests at the Beachside on the 4th of July 2018.	Done. Click buttons below for copies of invoice and booking confirmation. We look forward to welcoming you and your guests at the Beachside on the 4th of July 2018.	Done. Click buttons below for copies of invoice and booking confirmation. We look forward to welcoming you and your guests at the Beachside on the 4th of July 2018.

CB = Chatbot, H = Human.

Appendix B:. Transcript of study 2 stimulus used, illustrating experimental manipulation.

	A: Error-free	C: Clarification	B: Error
CB	Hi. I'm Eve, the Euro-Rail chatbot. How can I be of assistance?	Hi. I'm Eve, the Euro-Rail chatbot. How can I be of assistance?	Hi. I'm Eve, the Euro-Rail chatbot. How can I be of assistance?
H	I want prices for unlimited travel passes	I want prices for unlimited travel passes	I want prices for unlimited travel passes
CB	Ok. Unlimited travel inside the entire Euro-Rail network or inside a specific country?	Ok. Unlimited travel inside the entire Euro-Rail network or inside a specific country?	Ok. Unlimited travel inside the entire Euro-Rail network or inside a specific country?
H	UK. But does that include Wales?	UK. But does that include Wales?	UK. But does that include Wales?
CB	–	Do you mean "United Kingdom"?	Sorry. I don't understand.
H	–	Yes	What? You don't get it. I want unlimited travel for the UK.
CB	Yes, the BritRail pass includes travel on the England, Scotland and Wales networksWho will be travelling?	Yes, the BritRail pass includes travel on the England, Scotland and Wales networksWho will be travelling?	Yes, the BritRail pass includes travel on the England, Scotland and Wales networksWho will be travelling?
H	1 adult	1 adult	1 adult
CB	Great. We have unlimited pass options ranging from 2 days to 1 month. Prices vary depending on whether you want to travel on First or Second class	Great. We have unlimited pass options ranging from 2 days to 1 month. Prices vary depending on whether you want to travel on First or Second class	Great. We have unlimited pass options ranging from 2 days to 1 month. Prices vary depending on whether you want to travel on First or Second class
H	What's the difference?	What's the difference?	What's the difference?
CB	–	Do you mean "First class vs. Second class"?	
H	–	Yes	
CB	First class offers (i) large, comfortable reclining seats, (ii) a gourmet meal delivered to your seat and (iii) an open bar. Second class offers (i) standard seats and (ii) access to the bar-buffet cart	First class offers (i) large, comfortable reclining seats, (ii) a gourmet meal delivered to your seat and (iii) an open bar. Second class offers (i) standard seats and (ii) access to the bar-buffet cart	First class offers (i) large, comfortable reclining seats, (ii) a gourmet meal delivered to your seat and (iii) an open bar. Second class offers (i) standard seats and (ii) access to the bar-buffet cart
H	Second class is fine	Second class is fine	Second class is fine
CB	Excellent. Please find pricing options for 2nd class BritRail passes3 days unlimited / US$17615 days unlimited / US$4701 month unlimited / US$695Are you ready to book your pass or do you have further questions?	Excellent. Please find pricing options for 2nd class BritRail passes3 days unlimited / US$17615 days unlimited / US$4701 month unlimited / US$695Are you ready to book your pass or do you have further questions?	Excellent. Please find pricing options for 2nd class BritRail passes3 days unlimited / US$17615 days unlimited / US$4701 month unlimited / US$695Are you ready to book your pass or do you have further questions?
H	How much luggage can I take with me?	How much luggage can I take with me?	How much luggage can I take with me?
CB	–	Are you enquiring about "baggage restrictions"?	Sorry. I thought we were talking about "unlimited BritRail passes". Do you have any further questions?
H	–	Yes	How many bags can I bring?
CB	There are no restrictions on the amount (number of items / weight) of the luggage that may be brought on an overnight train, but there is limited space in the compartment to store the luggage	There are no restrictions on the amount (number of items / weight) of the luggage that may be brought on an overnight train, but there is limited space in the compartment to store the luggage	Sorry. I don't understand.
H	That's all. Bye	That's all. Bye	That's all. Bye

CB	On behalf of Euro-Rail, thanks for reaching out. Regards, Eve.	On behalf of Euro-Rail, thanks for reaching out. Regards, Eve.	On behalf of Euro-Rail, thanks for reaching out. Regards, Eve.

CB = Chatbot, H = Human.

Appendix C. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbusres.2020.04.030.

References

Aggarwal, P., & McGill, A. (2012). When brands seem human, do humans act like brands? Automatic behavioral priming effects of brand anthropomorphism. *Journal of Consumer Research, 39*(2), 307–323.

Araujo, T. (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.

Barrett, J., Richert, R., & Driesenga, A. (2001). God's beliefs versus mother's: The development of nonhuman agent concepts. *Child Development, 72*(1), 50–65.

Bartneck, C., Kulic, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics, 1*(1), 71–81.

Blackburn, S. (1996). Implicature. In The Oxford Dictionary of Philosophy (pp. 188–189). Oxford University Press.

Blut, M., Wang, C., & Schoefer, K. (2016). Factors influencing the acceptance of self-service technologies: A meta-analysis. *Journal of Service Research, 19*(4), 396–416.

Burgess, M. (2017). The NHS is trialling an AI chatbot to answer your medical questions. Retrieved October 8, 2017, from http://www.wired.co.uk/article/babylon-nhs-chatbot-app.

Chandler, J., & Shapiro, D. (2016). Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology, 12*(1), 53–81.

Clement, J. (2019). Global number of mobile messaging users 2018–2022. Statista. https://www.statista.com/statistics/483255/number-of-mobile-messaging-users-worldwide/ (Accessed on 20th Jan, 2020).

Collier, J., & Kimes, S. (2012). Only if it is convenient: Understanding how convenience influences self-service technology evaluation. *Journal of Service Research, 16*(1), 39–51.

Constine, J. (2017). Facebook will launch group chatbots at F8. Retrieved March 28, 2018, from https://techcrunch.com/2017/03/29/facebook-group-bots/.

Corti, K., & Gillespie, A. (2016). Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human. *Computers in Human Behavior, 58*, 431–442.

Cresci, E. (2017). Chatbot that overturned 160,000 parking fines now helping refugees claim asylum. Retrieved October 8, 2017, from https://www.theguardian.com/technology/2017/mar/06/chatbot-donotpay-refugees-claim-asylum-legal-aid.

Curran, J., & Meuter, M. (2005). Self-service technology adoption: Comparing three technologies. *Journal of Services Marketing, 19*(2), 103–113.

Dabholkar, P. (1996). Consumer evaluations of new technology-based self-service options: An investigation of alternative models of service quality. *International Journal of Research in Marketing, 13*, 29–51.

Dabholkar, P., & Bagozzi, R. (2002). An attitudinal model of technology-based self-service: Moderating effects of consumer traits and situational factors. *Journal of the Academy of Marketing Science, 30*(3), 184–201.

Dabholkar, P., Bobbitt, M., & Lee, E. (2003). Understanding consumer motivation and behaviour related to self-scanning in retailing: Implications for strategy and research on technology-based self service. *International Journal of Service Industry Management, 14*(1), 59–95.

Doorn, J. Van, Mende, M., Noble, S., Hulland, J., Ostrom, A., Grewal, D., & Petersen, J. A. (2017). Domo Arigato Mr. Roboto : Emergence of automated social presence in organizational frontlines and customers' service experiences, 20(1), 43–58.

Epley, N., Akalis, S., Waytz, A., & Cacioppo, J. (2008). Creating social connection through inferential reproduction: Loneliness and perceived agency in gadgets, gods, and greyhounds. *Psychological Science, 19*, 114–120.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886.

Eyssel, F., Kuchenbrandt, D., Hegel, F., & De Ruiter, L. (2012). Activating elicited agent knowledge: How robot and user features shape the perception of social robots. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication,* 851–857.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analyses using GPower 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160.

Flowerdew, J. (2014). *Discourse in context.* London: Bloomsbury Academic.

Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*(1), 27–52.

Gartner. (2016). Top Strategic Predictions for 2017 and Beyond: Surviving the Storm Winds of Digital Disruption. Retrieved from https://www.gartner.com/binaries/content/assets/events/keywords/cio/ciode5/top_strategic_predictions_fo_315910.pdf.

Griffith, E., & Simonite, T. (2018). Facebook's virtual assistant M is dead. So are chatbots. Retrieved from https://www.wired.com/story/facebooks-virtual-assistant-m-is-dead-so-are-chatbots/.

Grice, H. (1975). *Logic and conversation. Syntax and semantics 3: Speech acts*. New York: Academic Press41–58.

Hayes, A. (2017). *Introduction to Mediation, Moderation and Conditional Process Analysis* (2nd ed.). New York, NY: Guilford Press.

Hayes, A., & Preacher, J. (2013). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology, 67*, 451–470.

Hill, N., & Alexander, J. (2006). *The Handbook of Customer Satisfaction and Loyalty Measurement* (3rd ed.). London, UK: Routledge.

Hoehle, H., Scornavacca, E., & Huff, S. (2012). Three decades of research on consumer adoption and utilization of electronic banking channels: A literature analysis. *Decision Support Systems, 54*(1), 122–132.

Huang, M., & Rust, R. (2013). IT-related service: A multidisciplinary perspective. *Journal of Service Research, 16*(3), 251–258.

Hung, C., Yen, D., & Ou, C. (2012). An empirical study of the relationship between a self-service technology investment and firm financial performance. *Journal of Engineering & Technology Management, 29*(1), 62–70.

Hutchby, I., & Wooffitt, R. (2008). *Conversation Analysis* (2nd ed.). Boston, MA: Polity.

Johnson, K. (2017). Facebook Messenger hits 100,000 bots. Retrieved November 22, 2017, from https://venturebeat.com/2017/04/18/facebook-messenger-hits-100000-bots/.

Kaplan, F., & Hafner, V. (2006). The challenges of joint attention. *Interaction Studies, 7*(2), 135–169.

Kiesler, S., Powers, A., Fussell, S., & Torrey, C. (2008). Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition, 26*(2), 169–181.

Knijnenburg, B., & Willemsen, M. (2016). Inferring capabilities of intelligent agents from their external traits. *ACM Transactions on Interactive Intelligent Systems, 6*(4), 1–25.

Landwehr, J., McGill, A., & Herrmann, A. (2011). It's got the look: The effect of friendly and aggressive "facial" expressions on product liking and sales. *Journal of Marketing, 75*(3), 132–146.

Lane, J., Wellman, H., & Evans, E. (2010). Children's understanding of ordinary and extraordinary minds. *Child Development, 81*(5), 1475–1489.

Lee, H.-J. (2014). Consumer-to-store employee and consumer-to-self-service technology (SST) interactions in a retail setting. *International Journal of Retail and Distribution Management, 43*(8), 676–692.

Lee, H. J., & Lyu, J. (2016). Personal values as determinants of intentions to use self-service technology in retailing. *Computers in Human Behavior, 60*, 322–332.

Lee, W., Castellanos, C., & Chris Choi, H. (2012). The effect of technology readiness on customers' attitudes toward self-service technology and its adoption; The empirical study of U.S. airline self-service check-in kiosks. *Journal of Travel & Tourism Marketing, 29*(8), 731–743.

MacInnis, D., & Folkes, V. (2017). Humanizing brands: When brands seem to be like me, part of me, and in a relationship with me. *Journal of Consumer Psychology, 27*(3), 355–374.

Martin, A. (2017). AISB Loebner Prize 2017 Finalist Selection Transcripts. Retrieved November 15, 2017, from http://www.aomartin.co.uk/uploads/loebner_2017_finalist_selection_transcripts.pdf.

Martin, A. (2018). AISB Loebner Prize 2018 Finalist Selection Transcripts. Retrieved from https://www.aisb.org.uk/media/files/LoebnerPrize2018/Transcripts_2018.pdf.

Maudlin, M. (1994). ChatterBots, TinyMuds, and the Turing Test: Entering the Loebner Prize competition. In Proceedings of the Eleventh National Conference on Artificial Intelligence. AAAI Press.

Mazaheri, E., Richard, M., & Laroche, M. (2012). The role of emotions in online consumer behavior: A comparison of search, experience, and credence services. *Journal of Services Marketing, 26*(7), 535–550.

Meuter, M., Bitner, M., Ostrom, A., & Brown, S. (2005). Choosing among alternative service delivery modes: An investigation of customer trial of self-service technologies. *Journal of Marketing, 69*(2), 61–83.

Meuter, M., Ostrom, A., Bitner, M., & Roundtree, R. (2003). The influence of technology anxiety on consumer use and experiences with self-service technologies. *Journal of Business Research, 56*(11), 899–906.

Meuter, M., Ostrom, A., Roundtree, Robert, … Jo. (2000). Self-service technologies: Understanding customer satisfaction with technology-based service encounters. Journal of Marketing, 64(3).

Mitra, K., Reiss, M., & Capella, L. (1999). An examination of perceived risk, information search and behavioral intentions in search, experience and credence services. *Journal of Services Marketing, 13*(3), 208–228.

Moore, S. (2018). Gartner Says 25 Percent of Customer Service Operations Will Use Virtual Customer Assistants by 2020. Retrieved February 20, 2018, from https://www.gartner.com/newsroom/id/3858564.

Nass, C. (2004). Etiquette equality: Exhibitions and expectations of computer politeness. *Communications in Computer and Information Science, 47*(4), 35–37.

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments.

Presence-teleoperators and Virtual Environments, 12(5), 481–494.
Puzakova, M., Kwak, H., & Rocereto, J. (2013). When humanizing brands goes wrong: The detrimental effect of brand anthropomorphization amid product wrongdoings. *Journal of Marketing, 77*, 81–100.
Robinson, J. (2006). Managing trouble responsibility and relationships during conversational repair. *Communication Monographs, 73*(2), 137–161.
Schegloff, E. (1992). Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology, 97*(5), 1295–1345.
Scheutz, M., Schermerhorn, P., & Cantrell, R. (2011). Toward human-like task-based dialogue processing for HRI. *AI Magazine, 32*(4), 77–84.
Selnes, F., & Hansen, H. (2001). The potential hazard of self-service in developing customer loyalty. *Journal of Service Research, 4*(2), 79–90.
Stahl, G. (2015). Conceptualizing the intersubjective group. *International Journal of Computer-Supported Collaborative Learning, 10*(3), 209–217.
Tintarev, N., O'Donovan, J., & Felfernig, A. (2016). Introduction to the special issue on human interaction with artificial advice givers. *ACM Transactions on Interactive Intelligent Systems (TiiS), 6*(4), 26.
Touré-Tillery, M., & McGill, A. (2015). Who or what to believe: Trust and the differential persuasiveness of human and anthropomorphized messengers. *Journal of Marketing, 79*(4), 94–110.
Turing, M. (1950). Turing. Computing machinery and intelligence. *Mind, 49*, 433–460.
Wang, C., Harris, J., & Patterson, P. (2013). The roles of habit, self-efficacy, and satisfaction in driving continued use of self-service technologies: A longitudinal study. *Journal of Service Research, 16*(3), 400–414.
Waytz, A., Morewedge, C., Epley, N., Monteleone, G., Gao, J.-H., & Cacioppo, J. T. (2010). Making sense by making sentient: Effectance motivation increases anthropomorphism. *Journal of Personality and Social Psychology, 99*(3), 410–435.
White, G. (2018). Child advice chatbots fail to spot sexual abuse. Retrieved from https://www.bbc.com/news/technology-46507900.
Yang, J., & Klassen, K. (2008). How financial markets reflect the benefits of self-service technologies. *Journal of Enterprise Information Management, 21*(5), 448–467.

Ben Sheehan is a sessional academic and PhD candidate at the QUT Business School, Queensland University of Technology, Australia, having previously completed an M.Phil. at QUT and an MBA at the University of Queensland, Australia. Prior to joining academia, Ben held senior Marketing positions with a number of Australian and American firms.

Dr Hyun Seung Jin (PhD, University of North Carolina at Chapel Hill) is a Senior Lecturer in the School of Advertising, Marketing and Public Relations at the QUT Business School. Prior to joining QUT, Dr Jin worked at Samsung Co. in Seoul, Korea and taught advertising at Kansas State University and marketing at the University of Missouri at Kansas City in the United States. Dr Jin has published articles in Journal of Advertising, Journal of Advertising Research, International Journal of Advertising, European Journal of Marketing, Psychology & Marketing, Journal of Business Research, Journal of Consumer Affairs, Tourism Management, Journal of Strategic Marketing, Personality and Individual Differences, Health Marketing Quarterly, and others.

Udo Gottlieb, PhD is a Lecturer in the QUT Business School at Queensland University of Technology, Australia. Udo draws on over a decade's worth of international consulting experience across 4 continents. His current research provides organisations with strategies to improve their effectiveness when interacting with clients. His work has been published in scholarly journals including the European Journal of Marketing, Journal of Marketing for Higher Education and Electronic Commerce Research and Applications, among others.