Xiong Ji Jill (Orcid ID: 0000-0001-9345-3794)

Shen Hao (Orcid ID: 0000-0003-3361-6058)

The rise of chatbots: The effect of using chatbot agents on consumers' responses to request rejection

Shubin YU [a], Ji (Jill) XIONG [b, *], and Hao SHEN [c]

[a] BI Norwegian Business School

[b] Southwestern University of Finance and Economics

[c] The Chinese University of Hong Kong,

* Corresponding author

[a] Shubin YU is assistant professor at the Department of Communication and Culture, BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway. Email: Shubin.Yu@pku.edu.cn.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/jcpy.1330

[b, *] Ji (Jill) XIONG is associate professor of Marketing at School of Business Administration, Southwestern University of Finance and Economics, Room A331, Tongbo Building, 555, Liutai Avenue, Wenjiang District, Chengdu, Sichuan, P. R. China, 611130. Email: xiongji@swufe.edu.cn

[c] Hao SHEN is Professor of Marketing at The Chinese University of Hong Kong, Room 1116, 11/F, Cheng Yu Tung Building, 12 Chak Cheung Street, Shatin, N.T., Hong Kong, +852 3943 7820. Email: shenhao@baf.cuhk.edu.hk.

The rise of chatbots: The effect of using chatbot agents on consumers' responses to request rejection

To improve service management, reduce labor costs, and provide standardized services quickly and consistently, firms have introduced chatbots to handle customer requests, and these are now widely used to replace human service agents (Araujo, 2018; Komiak, Wang, & Benbasat, 2004). Chatbots, developed through artificial intelligence (AI), can be found in various industries, including banks, shopping malls, and entertainment centers (Chung et al., 2020; Morgan, 2017). Companies use robots not only to interact with consumers more efficiently but also to collect consumer data. In 2020, the chatbot market was valued at $17.17 billion and is expected to reach $102.29 billion by 2026, with an average annual growth rate of 34.75% between 2021 and 2026 (Research and Markets, 2021). One of the key factors driving the market growth is the increasing demand for AI customer support services, as the use of chatbots can significantly reduce the workforce and help companies cut costs by 30% on handling service requests (Brain, 2018).

However, whether using chatbots or human service agents, it is inevitable that companies sometimes have to reject a customer's service request. For example, consumers may be told that they cannot make an appointment for a product repair service because the service provider is unavailable at that time, that they cannot buy a product because it is out of stock, or that they cannot reserve a table because the restaurant is already full. The question is, when a request is rejected, will consumers react more or less negatively to a chatbot agent than to a human agent? This is the topic we investigate in the present research.

Chatbots are designed to handle consumer requests based on a preprogrammed set of procedures and algorithms (Broeck, Zarouali, & Poels, 2019; Kim & Duhachek, 2020). Therefore, they operate in a more standardized fashion (Longoni, Bonezzi, & Morewedge, 2019) and are perceived by consumers as less flexible and less innovative (Ding & Keh, 2016).

Because of this, it might be intuitive to infer that consumers would be more satisfied with the services provided by human service agents than by chatbots. Indeed, consumers often have an aversive attitude toward services provided by robots (Choi et al., 2020; Dana & Thomas, 2006; Dawes, Faust, & Meehl, 1989; Dietvorst, Simmons, & Massey, 2015; Leung, Paolacci, & Puntoni, 2018; Mende et al., 2019) because they perceive that robots lack uniqueness (Longoni et al., 2019) and empathy (Luo et al., 2019).

In this research, however, we propose that interactions with robots may change consumers' expectations about the services they will receive. To elaborate, because robots complete tasks based on a preprogrammed set of procedures (Broeck, et al., 2019; Kim & Duhachek, 2020), their service is rule-bound (Paluch, Wirtz, & Kunz, 2022; Wirtz et al., 2018). As a result, consumers tend to have low expectations about the level of flexibility in service from robots (Malle et al., 2015; Wirtz et al., 2018). In the context where service agents fail to satisfy consumer needs and preferences, consumers may believe that their request rejection is due to the robot's response constraint rather than its unwillingness to help them. Consequently, consumers might be more accepting of an unresolved service request if it is handled by a chatbot than by a human. We further propose that if consumers successfully settle their request, they might react to the service from a human more favorably than that from a chatbot, because they might be more likely to attribute the positive result to a human's willingness to tailor the service to their request.

We conducted six studies to test these conjectures and, in doing so, we make several contributions. First, we are among the first to examine consumers' different responses to negative service encounters incurred by robot agents versus human agents. We found that consumers evaluate the services provided by robots (vs. humans) more positively because they have lower expectations that the robots will be able to provide flexible services to them. Consequently, their dissatisfaction with the request rejection is lower when the service is handled by robots. We also found a set of boundary conditions for the aforementioned effect. Therefore, our work provides a deeper understanding of consumers' different perceptions of robots and humans. Second, although the rule-bound limits of service robots are often criticized for lacking flexibility, our work demonstrates how and when the inflexibility of rule-bound services may also benefit service evaluation. Thus, our research provides new insights into the role of flexibility in driving service satisfaction. Third, our research adds to the literature on attribution research in general by showing that the type of agent can shape consumers' expectations about the services and further influence their attribution of the service problem to the agent.

Perceptions of Robot Agents Versus Humans

Research has shown that people have stereotypical associations with computers and machines, such that they believe one key distinction between humans and machines is that humans have emotions and sensations, whereas machines do not (Gray & Wegner, 2012; Loughnan & Haslam, 2007). To elaborate, because machines are programmed to follow a certain set of rules to complete tasks, people tend to believe that machines do not have emotions or motivations (Monroe, Dillon, & Malle, 2014; Yogeeswaran et al., 2016). In addition, people might feel negatively about the concept of mechanistic dehumanization because it likens humans to unfeeling machines (Haslam & Loughnan, 2014). Conversely, endowing machines with emotions might make people uncomfortable because doing so threatens human distinctiveness (Gray & Wegner, 2012; Loughnan & Haslam, 2007). Therefore, people tend to draw a clear line between robots and human beings. As a result, they may perceive machines/robots as less affect-driven and without the capability to complete subjective tasks (Castelo, Bos, & Lehmann, 2019).

This distinction leads consumers to avoid robots in service interactions (Fryer et al., 2017; Mende et al., 2019; Önkal et al., 2009; Promberger & Baron, 2006). This could due to several reasons. First, robots follow rules to complete tasks and do not have control over their decisions. Therefore, it is more difficult to shift responsibility for consequential decisions to computers than to humans. This leads consumers to trust medical recommendations from an algorithm less than those from a human doctor (Promberger & Baron, 2006) and rely less on advice from an algorithm than from a human to forecast stock prices (Önkal et al., 2009). Second, the decision process of an algorithm is perceived to fit more with objective tasks that require logical, rule-based analysis than subjective tasks that require intuition and "gut instincts" (Inbar, Cone, & Gilovich, 2010). Therefore, people have been found to be more likely to rely on algorithms than on humans to solve numerical tasks with an objectively correct answer (Logg, Minson, & Moore, 2019), whereas they rely less on an algorithm than on humans to predict joke funniness (Yeomans et al., 2019). Similarly, because utilitarian value assessment typically uses criteria based on rational and logical dimensions, whereas hedonic value assessment relies on criteria based on experiential and emotional dimensions, consumers are more likely to choose AI recommenders than human ones when they have a utilitarian goal but are less likely to do so when they have a hedonic goal (Longoni & Cian, 2022). Third, because people draw a clear line between robots and human beings, they react to a humanoid robot negatively. For example, Mende et al. (2019) found that when consumers encounter a humanoid service robot, they may experience discomfort because they perceive a mismatch between the robot's anticipated human qualities and its actual imperfect, nonhuman qualities.

The Present Conceptualization

In a service context, technology can easily perform frequent and routine interactions with consumers (Frey, Osborne, & Holmes, 2016). Therefore, unlike analytical AI, which typically focuses on diagnosing problems or making recommendations, robots are frequently used in service contexts to maximize efficiency and to maintain service quality (Huang & Rust, 2017, 2018). They have been widely used in food ordering and delivery, self-service, and customer service for routine issues (Frey et al., 2016; Paluch et al., 2022).

Because they are programmed to follow rules and standards to deliver services, robot agents are also expected to provide more standardized services, in which each step is laid out in order and the outcomes are low in divergence (Shostack, 1987). The "McDonaldization phenomenon," in which consumers expect consistent and predictable standards of service and products, now applies to consumers' expectations in their interactions with robot service agents (Curran, Meuter, & Surprenant, 2003). Thus, unlike human service agents, who can provide more flexible and personalized services to adapt to individual consumers' needs, robot agents are expected to implement more standardized solutions and thus are less able to cater to consumers' personal preferences (Shostack, 1987).

Although little research has directly examined the flexibility of service from robots, some studies have found results consistent with our assumption. For example, Longoni, Bonezzi, and Morewedge (2019) showed that consumers are more reluctant to use healthcare services delivered by AI providers than by human providers because they are concerned that AI may ignore their unique characteristics. Similarly, another study showed that because consumers associate robotic production with consistency but associate human production with uniqueness, they prefer products created by humans if they seek uniqueness when purchasing symbolic products (Granulo, Fuchs, & Puntoni, 2020). Given these findings, it is reasonable to propose that consumers perceive the standardized services provided by robot agents as less flexible (Haslam, 2006).

In situations when the service agent rejects the consumer's request, one major contributor to consumer dissatisfaction is consumers' perception that the service agent does not care or is not making efforts to satisfy their needs (Bitner, 1990; Bitner, Booms, & Tetreault, 1990). As noted previously, because a robot agent follows a set of rules to complete a task, it is expected to provide standardized services with low variation (Haslam, 2006; Montagu, & Matson, 1983; Nissenbaum & Walker, 1998). Consequently, consumers expect a robot agent to be less flexible. In contrast, consumers expect a human service agent to be more flexible to adapt to their individual needs (Lovelock, 1983; Shostack, 1987). For example, when consumers are informed that they cannot reserve a table because the restaurant is already full, they may expect the server to find a way to solve this problem, such as by arranging a new table. If the server does not solve the problem, consumers may perceive that the server is not trying their best to help them. Consequently, they would experience a higher level of dissatisfaction due to their expectation that human agents should be flexible and thus evaluate the service more negatively. However, they may not expect robot agents to provide flexible services. Their service evaluation for the same service handled by robot agents, therefore, should also be less negative.

Given these considerations, we propose that a rejection of a service request will be less likely to lower consumers' service evaluation when the request is handled by a robot agent rather than a human. More formally, we hypothesize that,

H1a: When consumers receive a rejection of their service request, they evaluate the service less negatively if the service is handled by a chatbot agent (vs. a human agent).

H1b: The effect of chatbots (vs. human agents) on service evaluation is driven by consumers' perception that robot agents are less flexible.

Boundary Conditions

Our theorizing enables us to delineate a set of boundary conditions that test our proposed mechanism. Our conceptualization hinges on the premise that consumers tend to attribute a robot service agent's decision to the fact that it is following a rule rather than its unwillingness to help. It is important to define the context in which we may observe our proposed effect and understand when this effect may disappear or be reversed.

Service outcome. Our research is mainly interested in the context of service request rejection. However, our theorizing provides implications for other contexts as well. First, according to our conceptualization, consumers perceive the service from robots to be less flexible than that from humans. If so, when consumers have not yet experienced the service, they might predict the service from a human to be better than that from a robot. This assumption is consistent with algorithm aversion, which has been demonstrated in previous research (Dietvorst et al., 2015; Leung et al., 2018; Longoni et al., 2019; Mende et al., 2019). Second, the service agent may accept consumers' service request. In this case, will consumers still perceive the services from robots and humans differently? Our conceptualization indicates that consumers tend to attribute a robot agent's decision to the rule it is following rather than its willingness to help. Thus, consumers might appreciate a robot less because they think robots are simply following a rule to process their request, but they might appreciate a human agent more because they believe the human has tailored the service to their request. Therefore, we may observe that consumers react to a human (vs. robot) agent more favorably if their request is accepted.

H2a: Consumers predict the service from a human to be more flexible and therefore better than that from a robot if they have not experienced it.

H2b: Consumers react to the service from a human more favorably than that from a robot if their request is accepted.

Emotional expression. The service provider typically says sorry to consumers when it fails to satisfy their request (Bolkan & Daly, 2009). Although most of time the service provider may simply say sorry to consumers, emotions are sometimes expressed in the apology (Fehr & Gelfand, 2010; Ohbuchi, Kameda, & Agarie, 1989). Being programmed to follow rules to deliver services may affect consumers' perception of not only the service flexibility but also the sincerity of emotions expressed in the messages during the service delivery (Wirtz et al., 2018). For example, when the service agent fails to fulfill the consumer's request, they may express empathy to demonstrate concern for consumers' suffering (Brown et al., 2010; Davis, 1983) and to express guilt and remorse to restore consumers' loss of esteem for the firm (Bramel, Taub, & Blum, 1968; Howell, Turowski, & Buro, 2012; Walster et al., 1973). However, consumers may think that a robot simply follows a set of standardized procedures to handle service tasks without the ability to express real emotions (Haslam, 2006; Malle et al., 2015; Montagu & Matson, 1983). Consumers may feel uncomfortable with the idea of endowing machines with emotions and may perceive emotions expressed by machines as insincere (Gray & Wegner, 2012; Loughnan & Haslam, 2007). Affect-laden services provided by robot agents, as a result, should also be evaluated as less sincere and more negatively than those provided by human service agents. Drawing on these considerations, we propose that when a service agent apologizes for a service failure and expresses emotion, consumers may react to the service less favorably if the agent is a robot versus a human. Thus,

H3: When an apology message for a failed service delivery does not involve emotions, consumers might respond to the service more favorably if it is handled by a chatbot agent versus a human agent. However, this effect is reversed if the service agent conveys emotions to apologize for the failed service delivery.

Overview of Studies

We conducted six studies to investigate our hypotheses. Studies 1–3 tested H1 and provided evidence that consumers might respond to a robot agent less negatively than to a human agent when their service request is rejected. Study 1 showed that after a rejection, participants evaluated the service less negatively when the rejection was served by a chatbot agent than by a human because they expected the robot agent to be less flexible in tailoring the service to satisfy their request. Studies 2 and 3 further tested our effect in a real-life scenario. Studies 4A and 4B tested H2a and H2b, respectively. Specifically, Study 4A showed that when consumers have not experienced the service, they expect the service from a human to be more flexible and therefore predict human service to be better than the service from a robot. Study 4B demonstrated that when consumers' service request was accepted, they also reacted to the service more favorably when it was handled by a human than by a robot. Study 5 tested H3 and examined the effect of agent type when the service agent apologized to consumers for a failed service delivery. This study showed that although participants responded to a robot agent more favorably than to a human agent when the agent simply said sorry for a failed service delivery, they responded to a human agent more favorably than to a robot agent when the service agent expressed emotions in the apology. In summary, our studies provide consistent evidence for the effect of different types of service agents on consumers' reaction to the rejection of their service request, and we examine factors that moderate this effect.

We targeted at least 40 observations per cell in offline studies and 80 per cell in online studies. Therefore, Studies 2 and 3 (offline studies) employed 95 and 202 participants, respectively, and Studies 1, 4A, 4B, and 5 (online studies) employed 204, 194, 701, and 506 participants, respectively. Studies 1, 2, 3, and 4A comprised two conditions, whereas Studies 4B, and 5 comprised four conditions. Results of sensitivity analyses show that these sample sizes provide 80% power (α = .05) to detect small effects in most studies (Study 1: η_p^2 = .04; Study 2: η_p^2 = .08; Study 3: η_p^2 = .04; Study 4A: η_p^2 = .04; Study 4B: η_p^2 = .011; Study 5: η_p^2 = .015). Across studies, we used different scenarios to ensure the robustness of the effects we proposed. (For detailed stimuli, see Web Appendix.)

Study 1

Study 1 aimed to provide initial evidence for our assumption. That is, when consumers' request is rejected by a service agent, they may evaluate the service less negatively if the service is handled by chatbots than by human service agents. This effect occurs because consumers expect chatbots' service to be less flexible than that of humans.

Method

We recruited 204 respondents from Prolific and compensated each of them 0.60 GBP for participating (69% female; M_{age} = 36.38 years, SD = 12.06). This study was designed as a one-factor, two-level (service agent: chatbot vs. human) between-subjects study.

Participants were told that we are interested in people's perception of a service experience. On this pretense, they were exposed to a scenario in which they were asked to imagine that they needed to cancel their iPhone order through a service agent using Apple's official Twitter account. In the conversation screenshot, the service agent was described as either a human or a chatbot, depending on the condition. Participants were informed that the service agent failed to cancel the order because the package had already been sent out (for the screenshot, see Web Appendix). After participants read the scenario and the conversation, we asked them to indicate whether the service agent was a chatbot or a human (1 = chatbot, 2 = human). Then, they were asked to evaluate the service on two questions ("*The service agent provides a good service*," "*I am satisfied with the agent's service*"; 1 = *strongly disagree*, 7 = *strongly agree*; $r = .91$). Participants also indicated their expected flexibility of the agent on two questions ("*To what extent did you expect that the service agent could modify the rule to satisfy your needs/could do more to satisfy your needs?*") along scales from 1 (*not at all*) to 7 (*very much*) ($r = .43$).

Results

Manipulation check. Participants in the chatbot condition were more likely to identify the service agent as AI-powered (100%) compared with participants in the human service agent condition (22%; $\chi^2(1) = 141.90$, $p < .001$), suggesting that our manipulation was successful.

Service evaluation. Consistent with our expectation, we observed that when participants experienced a service request rejection, they considered the service more acceptable if it was rejected by the AI-powered agent ($M_{\text{chatbot}} = 2.77$, $SD = 1.48$) than by the human agent ($M_{\text{human}} = 2.04$, $SD = 1.29$; $F(1, 202) = 13.96$, $p < .001$, $\eta_p^2 = .07$).

The expected flexibility of the service. As expected, we also observed a significant effect of service agent type on the expected flexibility of the agent ($F(1, 202) = 15.06$, $p < .001$, $\eta_p^2 = .07$). In particular, participants expected a chatbot agent to be less flexible ($M_{\text{chatbot}} = 3.33$, $SD = 1.50$) than a human agent in providing services ($M_{\text{human}} = 4.15$, $SD = 1.51$; see Table 1).

Mediation analysis. Based on our prediction, participants who experienced a service rejection might be less likely to attribute the request rejection to the agent's unwillingness to help if the agent is a robot than a human, because the robot agent is expected to be less flexible than the human agent. We tested the mediating role of the expected flexibility in the effect of agent type on service evaluation. As reported previously, the effect of agent type on both flexibility and service evaluation was significant. In addition, the expected flexibility had a negative effect on service evaluation ($B = -.15$, $t(201) = -2.37$, $p = .018$). Moreover, bootstrapping indicated that the indirect effect of the expected flexibility on the effect of agent type on service evaluation was significant (point estimate $= -.12$, 95% CI = [−.28, −.01]).

Discussion

Study 1 showed that when participants' request was rejected by the service agent, they evaluated the service less negatively if it was provided by a chatbot than by a human. They also expected the service to be less flexible if it was provided by a chatbot than by a human. Therefore, the results of this study provided evidence for our conceptualization.

Study 2

To enhance the realism of the study and the external validity of the results, we conducted Study 2 using a real-world setting.

Method

Study 2 was designed as a field study with a one-factor, two-level (service agent type: chatbot vs. human) between-subjects design, with service evaluation measured as the dependent variable.

This study was conducted in Chinese. Our research assistants sent out gift vouchers to 200 university students in south China under the name of a fictitious company. The default gift was a bag of sweet donuts, which is an undesirable gift for most people in the local population. Students were told that they could get the gift one week later from our research assistant. However, if they did not like the default gift, they could redeem it for other gifts of equal value, such as a bar of chocolate or a Pikachu doll (for the voucher, see Web Appendix). To do so, they had to contact the service agent via an instant messenger. Participants were told that we have two types of service agents: an AI-powered service agent and a human customer service representative. According to the instructions on the voucher, participants were assigned to either a chatbot or a human and were asked to contact a particular account via instant messenger. However, the actual conversations in both conditions were conducted by our research assistants. They responded to participants' requests strictly based on a prepared script. One hundred thirteen students (M_{age} = 22.24 years; 73% female) contacted the agent to ask to redeem other gifts instead of the default one.

When participants contacted the agent, the service agent introduced themselves as being either AI-powered or human. Afterward, the service agent asked for the participant's name, email, and address to send the gift. Participants then sent in their request to redeem a different gift, specifying the gift they wanted. However, regardless of which gift participants requested, all participants were told that no other gift of equal value was currently available. As a result, all of the participants failed to receive the gift they wanted and ended up getting a bag of sweet donuts. They were told that the gift would be mailed to them later.

Then, participants received a short questionnaire from the agent. They were asked to evaluate the service by indicating their agreement with three statements: (1) "*I think the company* *provided good service*," (2) "*I felt satisfied with the service provided by the company*," and (3) "*I felt uncomfortable about the service*" (reverse-coded). All three questions were answered along scales from 1 (*disagree*) to 7 (*agree*) and were averaged to serve as an index for service evaluation ($\alpha = .84$). Finally, as an attention check question, we asked participants to recall whether they received service from an AI-powered service agent or a human service agent. Eighteen participants failed the attention check and were excluded from analysis, leaving 95 data points.

Results

Service evaluation. Similar to Study 1, we observed that when participants experienced service request rejection, they considered the service more acceptable if it was provided by the AI-powered agent ($M_{\text{chatbot}} = 4.57$, $SD = 1.37$) than by the human agent ($M_{\text{human}} = 3.95$, $SD = 1.51$; $F(1, 93) = 4.35$, $p = .040$, $\eta_p^2 = .05$; see Table 2).

Discussion

The findings of Study 2 again support our hypothesis that consumers are more accepting of a service rejection if the service was provided by a chatbot than by a human agent. Importantly, Study 2 was conducted in the field with actual consequences and consumer–agent interaction; thus, it advances the external validity of our conceptualization.

Study 3

Study 3 again examined our proposed effect in a real-world setting and further showed that the effect we proposed can be observed not only in consumers' service evaluation but also in their attitude toward the company.

Method

Study 3 had a single-factor, two-level (service agent: chatbot vs. human) between-subjects design. This study was conducted in Chinese. Only female participants were approached on the campus of a university in South China and invited to participate in our study. An existing company, Focus Gene, unfamiliar to most students, was chosen to prevent any pre-existing knowledge or attitude toward the company. Participants were informed that we were helping Focus Gene company to test its service system, and participants who agreed to help could receive either a L'OCCITANE hand cream (worth \$15 dollars, 95 RMB) or a MIYAS hand cream (worth \$2 dollars, 15 RMB). We targeted female participants because they had more product knowledge about hand cream.

After agreeing to participate, participants were asked to contact Focus Gene's service assistant using WeCom, a communication platform for enterprises. They were given a QR code to add the assistant's WeCom account. In one condition, they were told that this account is linked to a human service assistant. In the other condition, they were told that this account is linked to a chatbot service assistant. However, the actual conversations in both conditions were conducted by our research assistant, who replied to participants according to a script.

After participants contacted the service assistant, they were asked to indicate which of the brands of hand cream they wanted to receive. Because the brand L'OCCITANE is more well-known and its products are more valuable than MIYAS, most participants chose the former. Those who chose MIYAS were stopped from continuing the rest of the study. Afterward, participants were required to provide their demographic information, including gender and age.

Upon submitting their answers, participants were informed that the assistant would like to check the availability of the hand cream they requested. After 10 seconds, the service assistant replied that because there was no L'OCCITANE hand cream currently available, participants would receive a Chamomile Moisturizing Hand Cream from MIYAS.

After the conversation, participants were given a link to a survey regarding the service provided by Focus Gene. They were asked to evaluate the service along a star rating scale from one to seven stars. Then, they also were asked to indicate how willing they were to recommend Focus Gene to others on a scale from 1 (*very unwilling*) to 7 (*very willing*) and their attitude toward the company from 1 (*bad/unfavorable/dislikable/disagreeable*) to 7 (*good/favorable/likable/agreeable*) (α = .95).

Finally, as an attention check question, we asked participants to recall whether they received service from an AI-powered service agent or a human service agent. After they finished the questionnaire, participants were told the purpose of the study and received a monetary compensation of $2.50 (20 RMB) instead of the hand cream.

Two hundred forty-three people participated in this study. Twenty-nine participants failed the attention check and were excluded from analysis. In addition, 12 participants (seven participants in human condition and five in AI condition) were dropped because they did not choose L'OCCITANE hand cream, leaving 202 participants who completed all questions (100% female; M_{age} = 23.92 years, SD = 2.24).

Results

We expected that participants whose request was rejected might respond to the service less negatively if it were provided by a chatbot assistant than a human assistant. This was the case. We observed the effect in multiple dependent variables. That is, we observed a main effect of agent type on service evaluation (M_{chatbot} = 5.16 vs. M_{human} = 4.46; $F(1, 200) = 13.87$, $p < .001$,

η_p^2 = .07), recommendation intention ($M_{chatbot}$ = 4.34 vs. M_{human} = 3.83; $F(1, 200) = 6.71$, p = .010, η_p^2 = .03) and attitude toward the company ($M_{chatbot}$ = 4.61 vs. M_{human} = 4.14; $F(1, 200)$ = 7.54, p = .007, η_p^2 = .04; see Table 3).

Study 4

Studies 1–3 provided consistent evidence that when participants' service request is rejected, they react less negatively if the service is provided by a chatbot agent than a human agent. Studies 4A and 4B aimed to show whether service outcome would moderate this effect.

In particular, Study 4A showed that when consumers have not experienced the service, they might expect the service from a human to be more flexible and therefore predict their reaction to be better than the service from a robot. Study 4B showed that the effect we observed in Studies 1–3 holds only when consumers' request was rejected by the service agent; the effect was reversed when their request was accepted.

Study 4A

Method. We designed Study 4A as a one-factor, two-level (service agent: chatbot vs. human) between-subjects study. We recruited 194 respondents from Prolific and compensated each of them 0.50 GBP for participating (30.9% female; M_{age} = 24.61 years, SD = 8.05). Participants were exposed to a scenario in which they were asked to imagine that they needed to cancel their iPhone order through a service agent using Apple's official Twitter account, similar to what we did for Study 1. Next, we told them that they were assigned to either a chatbot assistant or a human assistant to handle their service. After they read the scenario, they were first asked to indicate whether the service agent would be an AI-powered agent or a human agent.

Then, they were asked to predict the service by indicating their agreements with the two statements: (1) "*The service agent will provide a good service*," and (2) "*I will be satisfied with the agent's service*." Both questions were answered along a seven-point scale (1 = *strongly disagree*, 7 = *strongly agree*; $r = .89$). Afterward, they were asked to predict the flexibility of the service. Because participants have not experienced the service, we used questions that are different from those used in Study 1 to measure the predicted flexibility by asking participants to indicate the extent to which they predict that the service agent (1) will have the flexibility to adapt to their requests; (2) will have the ability to adapt to their needs. Both items were measured along a seven-point scale (1 = *definitely no*, 7 = *definitely yes*; $r = .82$).

Results. In our manipulation check, participants in the chatbot condition were more likely to choose the service agent as an AI-powered service agent (95.7%) compared with participants in the human service agent condition (19%; $\chi^2(1) = 114.0$, $p < .001$), suggesting that our manipulation was successful.

More importantly, we observed that when participants were assigned to a chatbot service agent, they predicted the service to be worse ($M_{\text{chatbot}} = 4.03$, $SD = 1.39$) compared with participants who had a human agent ($M_{\text{human}} = 4.91$, $SD = 1.26$; $F(1, 192) = 21.76$, $p < .001$, $\eta^2_{\text{p}} = .10$). We also found that they predicted the service from a chatbot service agent to be less flexible than that from a human agent ($M_{\text{chatbot}} = 3.78$, $SD = 1.51$ vs. $M_{\text{human}} = 4.95$, $SD = 1.37$; $F(1, 192) = 32.20$, $p < .001$, $\eta^2_{\text{p}} = .14$; see Table 4).

Mediation analysis showed that before participants experienced the service, flexibility had a positive effect on service evaluation (B = .64, $t(192) = 11.68$, $p < .001$). In addition, bootstrapping showed that the indirect effect of flexibility on the effect of agent type on service evaluation was significant (point estimate = .64, 95% CI = [.38, .95]). These results suggest that when consumers have not experienced the service, the perceived inflexibility of chatbots leads consumers to predict that service from a robot will be worse than that from a human.

Study 4B

Study 4B further examined whether the effect we observed in previous studies depends on the service outcome. If consumers are more likely to attribute the service that they receive from a robot (vs. a human) to a rule rather than to the agent's willingness to help, they should also appreciate the service less when their service request is accepted. Study 4B tested this possibility.

Method. We recruited participants from Prolific. They were assigned randomly to one of four cells of a 2 (service agent: chatbot vs. human) × 2 (request handling status: failure vs. success) between-subjects design. This study was pre-registered on AsPredicted.org (https://aspredicted.org/4ik7t.pdf). As pre-registered, we tried to recruit 700 participants. We finally got seven hundred one participants (M_{age} = 38.0 years, *SD* = 12.6; 67.0% female).

Participants were told to imagine that they have purchased a cotton T-shirt from an online store called FREEMEIGE (fictitious). The T-shirt was too big, so they contacted the service agent to return it. Next, participants were shown a screenshot of the conversation with the service agent. To manipulate the type of service agent, we told participants that the agent was either a chatbot or a human. To manipulate the handling status, we told participants in the request-handling-success condition that they were able to return the T-shirt and get a full refund. However, participants in the service failure condition were told that their service request could not be satisfied because they have confirmed the receipt of the product (see Web Appendix for the stimuli).

After participants read the screenshot, they were asked to evaluate the service by indicating their agreements with the two statements: (1) "*I think this store provided a good service*," and (2) "*I am satisfied with the service from this store*." Both questions were answered along a seven-point scale (1 = *strongly disagree*, 7 = *strongly agree*; r = .97). In addition, we assessed the overall attitude toward the online store using a four-item bipolar scale (*negative/positive*, *bad/good*, *unfavorable/favorable*, *dislike it/like it*) (α = .99). At the end of the study, participants were asked to indicate their agreement with the statement that the service agent solved their service request on a seven-point scale (1 = *strongly disagree*, 7 = *strongly agree*). They were also asked to recall whether they interacted with a chatbot or a human service agent (1 = AI-powered service agent, 2 = Human service representative). Both questions were measured as manipulation checks.

Results. In the manipulation check, participants in the chatbot condition were more likely to choose the service agent as a chatbot (98.6%) compared with participants in the human service agent condition (11.4%, $\chi^2(1) = 537.98$, $p < .001$), suggesting that our manipulation of service agent type was successful. In addition, an ANOVA with service request handling status as the independent variable and participants' rating of whether the agent solved their service request as the dependent variable revealed that participants in the service success condition agreed more with the statement that the service agent solved their service request than those in service failure condition ($M_{\text{success}} = 6.22$ vs. $M_{\text{failure}} = 1.37$; $F(1, 699) = 3986.85$, $p < .001$, $\eta_p^2 = .85$).

We expected that participants whose service request was not being met would evaluate the service provided by the chatbot as less negative than the service provided by the human agent. We also predicted that this effect would be reversed if their service request was accepted. The results were consistent with our predictions. We observed a main effect of service status (p

< .001), which was qualified by a significant interaction between service agent type and service status ($F(1, 697) = 23.55$, $p < .001$, $\eta_p^2 = .03$). As Table 5 shows, when the service request was rejected, we observed a significant effect of agent type such that participants reacted to the service less negatively if it was provided by a chatbot than by a human agent ($M_{\text{chatbot}} = 2.04$ vs. $M_{\text{human}} = 1.62$; $F(1, 697) = 10.31$, $p = .001$, $\eta_p^2 = .02$). However, they reacted to the service from a chatbot (vs. a human agent) less positively when the request was handled successfully ($M_{\text{chatbot}} = 5.63$ vs. $M_{\text{human}} = 6.11$; $F(1, 697) = 13.32$, $p < .001$, $\eta_p^2 = .02$; for details, see Table 5).

We obtained similar results for the attitude toward the online store. That is, in addition to the main effects of service status ($p < .001$), we observed a significant effect of the interaction between service agent type and service status on the attitude ($F(1, 697) = 18.88$, $p < .001$, $\eta_p^2 = .03$). In particular, when the service request was rejected, participants showed a less negative attitude toward the online store when they interacted with a chatbot agent than with a human agent ($M_{\text{chatbot}} = 2.00$ vs. $M_{\text{human}} = 1.59$; $F(1, 697) = 10.09$, $p = .002$, $\eta_p^2 = .01$). However, they showed a less positive attitude toward the online store in chatbot condition than in human condition when the request was handled successfully ($M_{\text{chatbot}} = 5.66$ vs. $M_{\text{human}} = 6.04$; $F(1, 697) = 8.82$, $p = .003$, $\eta_p^2 = .01$).

Discussion. Study 4B further demonstrated that service outcome moderated the effect we observed. That is, when participants' request was rejected, they reacted to the service less negatively when the rejection was delivered by a chatbot than by a human. However, when their request was accepted, they reacted to the service less positively in the former case than in the latter.

Study 5

Study 5 aimed to investigate another possibility: Will the perception that a service agent is simply following a set of rules to provide services hurt consumers' service evaluation when their request is rejected?

To elaborate, service agents typically apologize to consumers after a service failure (Bolkan & Daly, 2009). Sometimes instead of simply saying sorry, the service agent may express emotions to consumers in an apology. For example, previous research has shown that the service agent may use emotional messages to enhance the efficacy of an apology (Coombs, 2014; Moon & Rhee, 2012). An apology may involve expression of emotions such as empathy or guilt to acknowledge consumers' suffering and to take the responsibility for the failure (Howell et al., 2012).

However, the fact that a robot agent simply follows a set of standardized procedures to handle service tasks may lead consumers to think it is not able to express emotions (Haslam, 2006; Malle et al., 2015; Montagu & Matson, 1983). Therefore, when an apology conveys emotional messages to consumers, they might perceive it as less sincere when it comes from a robot agent versus a human agent. If so, they might change their service evaluation accordingly. Study 5 aimed to test these assumptions.

Methods

Five hundred eighty-one MTurk workers were compensated .85 USD for participating (39% female; M_{age} = 37.83 years, SD = 11.54). This experiment was a 2 (service agent: chatbot vs. human) × 2 (apology: emotional vs. non-emotional) between-subjects design. Seventy-four participants failed an instructional manipulation check (i.e., participants were asked to select "*strongly disagree*" in an attention check question unrelated to our scenario) in the middle of the study and were stopped from completing other questions, leaving 506 participants who completed all questions.

We told participants that we are interested in people's perception of different service scenarios. On this pretense, they were exposed to a scenario in which they were asked to imagine that they needed to cancel their iPhone order through a service agent using Apple's official Twitter account. In the conversation screenshot, the service agent was described as either a human or a chatbot, depending on the condition. Participants were informed that the service agent failed to cancel the order because the package had already been sent out (see Web Appendix for the screenshot). In the emotional apology condition, participants received an emotional message that conveyed an apology from the service agent for failing to solve the problem. Participants read, "I am sorry. You cannot cancel the order because it is already 'in transit'. I know it is our responsibility to satisfy your request and on behalf of our company, I extend our sincerest apologies for being unable to satisfy your request this time. I am truly sorry for any inconvenience that you may experience. I hope it won't prevent you from seeking my help in the future." The non-emotional apology condition only included the first two sentences ("I am sorry. You cannot cancel the order because it is already 'in transit'.") and did not include an emotional message (see Web Appendix for the screenshot). We conducted a separate test with another group of participants to ensure that the message shown in the emotional apology condition was indeed perceived as an emotional apology. One hundred forty-one participants completed the post-test (38 female; M_{age} = 34.09 years, SD = 8.28). After reading the message, participants indicated the extent to which they thought the message was an emotional apology on two items ("*To what extent do you think the apology offered by the service agent is emotional*?"

“*To what extent do you think the apology offered by the service agent contains emotion*?”; r =.84; 1 = *not at all*, 7 = *to a great extent*). The result confirmed that the message in the emotional apology condition was perceived as an emotional apology to a greater extent than the one in the non-emotional apology condition ($M_{\text{emotional}}$ = 5.01, SD = 1.53 vs. $M_{\text{non-emotional}}$ = 4.13, SD = 2.07; $t(138) = -2.85$, p = .005, η_p^2 = .06), demonstrating that our manipulation of emotional apology in the main study was successful.

After participants read the scenario and the conversation, they were asked to evaluate the service by indicating whether the agent provided reliable service along a scale from 1 (*not at all*) to 7 (*very much*). Participants also indicated whether they would contact the service agent if they have problems in the future (1= *definitely no*, 7 = *definitely no*). Then, participants were asked to indicate the extent to which the apology was made sincerely (1 = *not at all*, 7 = *very much*). As a manipulation check, they also answered whether the service agent was an AI-powered service agent or a human (1 = AI-powered service agent, 2 = human service agent).

Results

Manipulation check. More participants in the chatbot condition perceived the service agent as an AI-powered service agent (89%) compared with participants in the human service agent condition (50%, $\chi^2(1)$ = 92.55, p < .001), suggesting that our manipulation of agent type was successful.

Service evaluation. We expected that participants would evaluate the service from a robot more favorably than that from a human when their request is rejected and they receive an emotionless apology from the agent. However, when the agent's apology contains emotion, we predicted that participants would evaluate a robot less positively than a human because they do not perceive the emotions expressed by a robot to be sincere. This was the case. We observed a main effect of apology on service evaluation ($F(1, 502) = 9.87$, $p = .002$, $\eta_p^2 = .02$). More importantly, this effect was qualified by a significant two-way interaction between agent type and apology ($F(1, 502) = 16.35$, $p < .001$, $\eta_p^2 = .03$). In particular, participants in the non-emotional condition evaluated the service more favorably if it was provided by the AI-powered agent ($M_{chatbot} = 4.75$, $SD = 1.43$) than by the human agent ($M_{human} = 4.31$, $SD = 1.65$; $F(1, 502) = 4.61$, $p = .032$, $\eta_p^2 = .01$). In contrast, those in the emotional condition evaluated the service less favorably if it was provided by the chatbot ($M_{chatbot} = 4.62$, $SD = 1.76$) than by the human ($M_{human} = 5.32$, $SD = 1.44$; $F(1, 502) = 12.95$, $p < .001$, $\eta_p^2 = .03$).

Intention to contact the service agent in the future. A two-way ANOVA with participants' intention to use the service in the future as the dependent variable revealed similar results as the service evaluation. Specifically, a main effect of apology ($F(1, 502) = 29.64$, $p < .001$, $\eta_p^2 = .06$) was qualified by a significant agent type × apology interaction ($F(1, 502) = 8.53$, $p = .004$, $\eta_p^2 = .02$). After receiving a non-emotional apology, participants' intention to contact the AI-powered service agent in the future was higher than those who had the human agent ($M_{chatbot} = 4.59$, $SD = 1.65$ vs. $M_{human} = 4.00$, $SD = 1.89$; $F(1, 502) = 7.74$, $p = .006$, $\eta_p^2 = .02$). However, after receiving an emotional apology, their intention to contact the AI-powered service agent in the future was directionally lower than those who had the human agent, though the difference was not statistically significant ($M_{chatbot} = 4.95$, $SD = 1.68$ vs. $M_{human} = 5.22$, $SD = 1.32$; $F(1, 502) = 1.75$, $p = .19$, $\eta_p^2 = .003$).

The perception of apology. We expected that when the apology is not emotional, participants would react to it more positively if it were made by a robot than by a human.

However, we believed the reverse would be true when the apology is emotional, because participants might not believe that a robot could express emotions. This was also the case. We observed a main effect of apology on the agent's perceived sincerity ($F(1, 502) = 28.98$, $p < .001$, $\eta_p^2 = 0.06$). More importantly, this effect was qualified by a significant two-way interaction between agent type and apology ($F(1, 502) = 6.22$, $p = .013$, $\eta_p^2 = .01$). In particular, when the apology was not emotional, participants perceived it as non-significantly more sincere when it was provided by the AI-powered agent than the human agent ($M_{chatbot} = 4.44$, $SD = 1.83$ vs. $M_{human} = 4.09$, $SD = 1.97$; $F(1, 502) = 2.36$, $p = .12$, $\eta_p^2 = .01$). In contrast, the reverse was significantly true when the apology was emotional ($M_{chatbot} = 4.89$, $SD = 1.80$ vs. $M_{human} = 5.33$, $SD = 1.45$; $F(1, 502) = 3.99$, $p = .046$, $\eta_p^2 = .01$; for details, see Table 6).

Discussion

Study 5 replicated the results of Studies 1–4 by showing that when experiencing a service request rejection, participants evaluated the chatbot's service less negatively. More importantly, Study 5 also shows that the effect was reversed when the service agent expressed emotions when apologizing for the service rejection.

Some limitations of this study are worth noting. Although we tried to manipulate the emotionality of the apology, such a manipulation may cause other differences as well. For example, participants may feel that the agent wants to take responsibility or hopes to have a future interaction with the customer when they express emotions in an apology. In addition, the emotional apology also contains more words than a non-emotional apology. Therefore, we also acknowledge the possibility that other factors may contribute to this effect.

General Discussion

Companies are increasingly adopting service robots to interact with consumers. In doing so, they can improve operational efficiency and save costs. However, little is known about consumers' perception of robots that replace human service agents to handle service requests. Our research sheds light on this question by examining consumers' perceptions and evaluations of robot service agents compared with human service agents, specifically in the context of service agents accepting or rejecting service requests.

Four experimental studies and two field studies provided consistent evidence that when the service request was rejected, consumers perceived the service agent as less flexible when the agent was a robot rather than a human being and thus evaluated the service less negatively (Studies 1–3). Study 4A further confirmed that the expected flexibility could also be the source of people's aversion to algorithms. When consumers have not experienced a service, they expect the service from a robot to be less flexible than that from a human, and therefore they tended to predict the robot's service to be worse. Study 4B demonstrated that service outcome moderated the effect we observed. When consumers' request was resolved, they evaluated the service more positively when it was handled by a human (vs. robot) agent. Study 5 further investigated the moderating role of emotions and revealed that the effect was reversed when the service agent expressed emotions when apologizing for the service rejection.

Theoretical Implications

This research contributes to several research streams. First, with the growing interest in using robots to replace human labor in various industries, researchers have also begun to investigate the different perceptions of robots and human service agents and consumers' preferences for robots or humans as the service agent (Granulo et al., 2020; Yeomans et al., 2019). The majority of the literature has demonstrated algorithm aversion, in which consumers prefer humans over robots for providing services (Dietvorst et al., 2015; Leung et al., 2018; Mende et al., 2019), because consumers perceive that algorithms lack human abilities to consider individual uniqueness or provide empathy (Longoni et al., 2019; Luo et al., 2019). Only a few works have shown situations in which consumers prefer algorithms to humans (Castelo et al., 2019; Dietvorst et al., 2016; Logg et al., 2019). These works focused on contexts in which consumers had not experienced the service yet and were asked to choose between algorithms and humans for recommendations or advice. Our research contributes to this emerging literature by showing a positive effect of robots versus humans in the service context. We replicated prior research on algorithm aversion when consumers have not experienced the services, as they expected robot agents to be less flexible in dealing with their service requests. More importantly, we also observed that after consumers' service requests were rejected, they evaluated the service less negatively when the service was handled by a chatbot agent versus a human agent.

Our findings also advance the understanding of consumers' perceptions of robot versus human service agents and the influences on service evaluation. Previous literature has discussed robots' rule-bound limits (Ding & Keh, 2016; Paluch et al., 2022; Wirtz et al., 2018). We build on this line of research by suggesting that consumers perceive robots as less flexible than human service agents in handling service requests. We further expand the literature by showing that whether this perception increases or decreases service evaluation depends largely on the context. When consumers have not experienced the service yet, or when their service request is accepted, the service agent's perceived inflexibility negatively influences the service evaluation. However, when consumers' service request is rejected by the service agent, the service agent's perceived inflexibility serves as a buffer and increases the service evaluation.

Third, our research also contributes to the literature on attribution research. Attribution theory suggests that consumers make attributions for the causes of service problems along three dimensions: locus of causality, stability, and control (Weiner, 2000). Although most of research suggests that the locus of attribution depends on the provider's objective responsibility (Oliver & DeSarbo, 1988), we show that attribution may also be affected by consumers' subjective perception of the provider based on whether the provider is a robot or human. That is, our research shows that the type of agent (i.e., robot or human) may affect consumers' expectations about whether the agent can provide flexible services and may further influence their attribution of the service problem to the rule that the agent is following versus its willingness to help. Therefore, our research provides insights into how consumers make attributions in a service context.

Forth, this research also makes a novel contribution to the understanding of the role of apologies in service failure recovery. Apologies are considered an effective service recovery attribute (Basso & Pizzutti, 2016; Wirtz & Mattila, 2004) and have been identified as a beneficial factor that increases consumers service evaluation and improves customer relationships (Basso & Pizzutti, 2016; Goodwin & Ross, 1992; Gregoire, Tripp, & Legoux, 2009; Ohbuchi et al., 1989), with affect-laden apologies proving to be more effective (Coombs, 2014; Moon & Rhee, 2012). Our research extends this discussion by showing that although apologies from chatbots increased consumers' service evaluation, affect-laden apologies made by chatbots had the opposite effect, because consumers perceive those apologies as insincere. Our results also corroborate prior research findings that algorithms are perceived as less authentic than humans (Jago, 2019). We show that this perception of a lack of authenticity in robots might be particularly true when a robot expresses emotions in an apology.

Managerial Implications

Our work has managerial implications for companies that currently use or plan to use chatbots. First, our findings suggest that consumers still prefer human agents. However, this does not mean that a human agent is always the best choice. For companies that prioritize consumer service evaluation, introducing chatbots could be effective when offering explanations for failed service delivery. Companies can take advantage of chatbots to lower costs; provide quick, consistent service; and alleviate the negative effect of service failure.

Second, companies often empower employees to provide personalized services to meet consumers' requests. Although such a strategy can often increase consumers' satisfaction, our research suggests that this may also affect consumers' expectations about the flexibility of the services. If their request cannot be satisfied, they might feel more negative about the service. Therefore, managers need to keep in mind that personalized services may sometimes reduce consumer satisfaction when the service request is rejected.

Third, our findings also suggest that consumers are less likely to expect chatbots to handle their service requests flexibly, which makes service rejection more acceptable and justifiable. Therefore, consumers are less dissatisfied when their requests are rejected by a chatbot versus by a human. With this in mind, when companies have difficulty addressing unsolvable service requests, a better strategy than sending human service agents might be to use chatbots to interact with customers. Nonetheless, we cannot ignore the fact that in situations where companies need to apologize for a service failure and demonstrate sincerity, using chatbots might backfire, as the chatbots' inflexibility and lack of emotions would be perceived as insincere.

Limitations and Future Research

Some limitations and future research are worth noting. First, our proposed effect might be more applicable to scenarios in which the service provider makes a “yes/no” decision that can be easily governed by a set of pre-programmed rules. In reality, however, the services consumers receive may vary in many dimensions. For example, a consumer may find that the service provider does not reply to them quickly or does not provide them enough information. If consumers do not perceive those responses to be governed by a set of rules, they might not necessarily react to a robot less negatively than a human. Future research could examine those contexts.

Second, we do not claim that marketers should use chatbots to replace human service providers altogether. Using chatbots might not be desirable for services that involve providing emotional support or building emotional connections. Future research could examine consumers’ satisfaction with different types of service agents (chatbots or humans) within different types of service requests.

Finally, it would also be interesting to explore consumers’ relationship with the robot if they receive a programmed emotional expression from a robot during the service experience. On the one hand, consumers might build a close relationship with a robot if it appears to be affect-driven. On the other hand, previous research suggests that consumers might be uncomfortable interacting with a robot that has emotions, as it poses a threat to human distinctiveness (Gray &Wegner, 2012; Loughnan & Haslam, 2007). Future research could investigate different effects of robots’ emotional expressions on consumers’ relationship with the robot agent.

REFERENCES

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, 183–89.

Basso, K., & Pizzutti, C. (2016). Trust Recovery Following a Double Deviation. *Journal of Service Research*, 19(2), 209–223.

Bitner, M. J. (1990). Evaluating Service Encounters: The Effects of Physical Surroundings and Employee Responses. *Journal of Marketing*, 54, 69-82.

Bitner, M. J., Booms, B. H., & Tetreault, M. S. (1990). The Service Encounter: Diagnosing Favorable and Unfavorable Incidents. *Journal of Marketing*, 54(1), 71–84.

Bolkan, S., & Daly J. A. (2009). Organizational Responses to Consumer Complaints: An Examination of Effective Remediation Tactics. *Journal of Applied Communication Research*, 37 (1), 21–39.

Brain [BRN.AI] CODE FOR EQUITY (2018). Chatbot Report 2018: Global Trends and Analysis. *Chatbots Magazine.* 2018, March 17.

Bramel, D., Taub, B., & Barbara Blum (1968). An Observer's Reaction to the Suffering of His Enemy. *Journal of Personality and Social Psychology*, 8 (4, Pt.1), 384–92.

Broeck, E.V., Zarouali, B., & Poels, K. (2019). Chatbot Advertising Effectiveness: When Does the Message Get Through? *Comput. Hum. Behav*., 98, 150-157.

Brown, T., Sautter, J., Littvay, L., Sautter, A., & Bearnes, B. (2010). Ethics and Personality: Empathy and Narcissism as Moderators of Ethical Decision Making in Business. *The Journal of Education for Business*, 85(4), 203–208.

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-Dependent Algorithm Aversion. *Journal of Marketing Research*, 56 (5), 809–25.

Choi, Y., Choi, M., Oh, M., & Kim, S. (2020). Service Robots in Hotels: Understanding the Service Quality Perceptions of Human-Robot Interaction. *Journal of Hospitality Marketing and Management*, 29(6), 613-635.

Chung, M., Ko, E., Joung, H., & Kim, S. J. (2020). Chatbot E-service and Customer Satisfaction Regarding Luxury Brands. *Journal of Business Research*, 117, 587–95.

Coombs, W. T. (2014). *Ongoing Crisis Communication: Planning, Managing, and Responding*. SAGE Publications.

Curran, J. M., Meuter, M. L., & Surprenant, C. F. (2003). Intentions to Use Self-Service Technologies: A Confluence of Multiple Attitudes. *Journal of Service Research*, 5 (3), 209–24.

Dana, J., & Thomas, R. (2006). In Defense of Clinical Judgment...and Mechanical Prediction. *Journal of Behavioral Decision Making*, 19(5), 413–428.

Davis, M. H. (1983). The Effects of Dispositional Empathy on Emotional Reactions and Helping: A Multidimensional Approach. *Journal of Personality*, 51 (2), 167–84.

Dawes, R. M., David Faust, & Paul E. Meehl (1989), "Clinical versus Actuarial Judgment," *Science*, 243(4899), 1668-1674. doi:10.1126/science.2648573.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. *Journal of Experimental Psychology: General*, 144(1), 114-126.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2016). Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms if They Can (Even Slightly) Modify Them. *Management Science: Journal of the Institute of Management Sciences*, 64(3), 1155-1170.

Ding, Y., & Keh, H. T. (2016). A Re-Examination of Service Standardization versus Customization from the Consumer's Perspective. *Journal of Services Marketing*, 30 (1), 16–28.

Fehr, R., & Gelfand, M. J. (2010). When Apologies Work: How Matching Apology Components to Victims' Self-Construals Facilitates Forgiveness. *Organizational Behavior and Human Decision Processes*, 113 (1), 37–50.

Frey, C. B., Osborne, M. A., & Holmes, C. (2016). Technology at work v2.0: The future is not what it used to be (Citi GPS: Global Perspectives & Solutions). Oxford. Retrieved from http://www.oxfordmartin.ox.ac.uk/downloads/reports/Citi_GPS_Technology_Work_2.pdf.

Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and Sustaining Interest in a Language Course: An Experimental Comparison of Chatbot and Human Task Partners. *Computers in Human Behavior*, 75, 461–68.

Goodwin, C., & Ross, I. (1992). Consumer Responses to Service Failures: Influence of Procedural and Interactional Fairness Perceptions. *Journal of Business Research*, 25, 149-163.

Granulo, A., Fuchs, C., & Puntoni, S. (2020). Preference for Human (vs. Robotic) Labor is Stronger in Symbolic Consumption Contexts. *Journal of Consumer Psychology*.

Gray, K., & Wegner, D. M. (2012). Feeling Robots and Human Zombies: Mind Perception and the Uncanny Valley. *Cognition*, 125 (1), 125–30.

Gregoire, Y., Tripp, T.M. and Legoux, R. (2009). When Customer Love Turns into Lasting Hate: The Effects of Relationship Strength and Time on Customer Revenge and Avoidance. *Journal of Marketing*, 73, 18-32.

Haslam, N. (2006). Dehumanization: An Integrative Review. *Personality and Social Psychology Review*, 10 (3), 252–64.

Haslam, N., & Loughnan. S. (2014). Dehumanization and Infrahumanization. *Annual Review of Psychology*, 65 (1), 399–423.

Howell, A. J., Turowski, J. B., & Buro, K. (2012). Guilt, Empathy, and Apology. *Personality and Individual Differences*, 53 (7), 917–22.

Huang, Ming-Hui, & Roland T. Rust (2017), Technology-Driven Service Strategy. *Journal of the Academy of Marketing Science*, 45 (6), 906-924.

Huang, Ming-Hui, & Roland T. Rust (2018). Artificial Intelligence in Service. *Journal of Service Research*, 21(2), 155–172.

Inbar, Y., Cone, J., & Gilovich, T. (2010). People's Intuitions About Intuitive Insight and Intuitive Choice. *Journal of Personality and Social Psychology*, 99(2), 232–247.

Jago, Arthur (2019), "Algorithms and Authenticity," *Academy of Management Discoveries*, 5 (1), 38–56.

Kim, T. W., & Duhachek, A. (2020). Artificial intelligence and persuasion: a construal-level account. *Psychological Science*, 31(4).

Komiak, S. Y. X., Wang, W., & Benbasat, I. (2004). Trust Building in Virtual Salespersons Versus in Human Salespersons: Similarities and Differences. *e-Service Journal*, 3 (3), 49.

Leung, E., Paolacci, G., & Puntoni, S. (2018). Man Versus Machine: Resisting Automation in Identity-Based Consumer Behavior. *Journal of Marketing Research*, 55(6), 818-831. doi:10.1177/0022243718818423

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm Appreciation: People Prefer Algorithmic to Human Judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103.

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to Medical Artificial Intelligence. *Journal of Consumer Research*, 46 (4), 629–50.

Longoni, Chiara & Luca Cian (2022). Artificial Intelligence in Utilitarian vs. Hedonic Contexts: The 'Word-of-Machine' Effect. *Journal of Marketing*, 86 (1), 91–108.

Loughnan, S., & Haslam, N. (2007). Animals and Androids: Implicit Associations Between Social Categories and Nonhumans. *Psychological Science*, 18 (2), 116–21.

Lovelock, C. H. (1983). Classifying Services to Gain Strategic Marketing Insights. *Journal of Marketing*, 47(3), 9-20.

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. *Marketing Science*, 38 (6) 937-947

Malle, B. F., Scheutz, M., Arnold, T., Voiklis, J., & Cusimano, C. (2015). Sacrifice One For the Good of Many? People Apply Different Moral Norms to Human and Robot Agents. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15*, Portland, Oregon, USA: ACM Press, 117–24.

Mende, Martin, Maura L. Scott, Jenny van Doorn, Dhruv Grewal, & Ilana Shanks (2019), "Service Robots Rising: How Humanoid Robots Influence Service Experiences and Elicit Compensatory Consumer Responses," *Journal of Marketing Research*, 56(4), 535-556.

Monroe, A. E., Dillon, K. D., & Malle, B. F. (2014). Bringing Free Will Down to Earth: People's Psychological Concept of Free Will and Its Role in Moral Judgment. *Consciousness and Cognition*, 27, 100–108.

Montagu, A., & Matson, F. W. (1983). *The Dehumanization of Man*, New York: McGraw-Hill.

Moon, B. B., & Rhee, Y. (2012). Message Strategies and Forgiveness during Crises: Effects of Causal Attributions and Apology Appeal Types on Forgiveness. *Journalism & Mass Communication Quarterly*, 89 (4), 677–94.

Morgan, B. (2017). How Chatbots Improve Customer Experience in Every Industry: An Infograph. *Forbes*.

Nissenbaum, H., & Walker, D. (1998). A Grounded Approach to Social and Ethical Concerns about Technology and Education. *Journal of Educational Computing Research*, 19 (4), 411–32.

Ohbuchi, K., Kameda, M., & Agarie, N. (1989). Apology as Aggression Control: Its Role in Mediating Appraisal of and Response to Harm. *Journal of Personality and Social Psychology*, 56 (2), 219–27.

Oliver, R. L., & DeSarbo, W. S. (1988). Response Determinants in Satisfaction Judgments. *Journal of Consumer Research,* 14(4), 495–507. https://doi.org/10.1086/209131

Önkal, D., Goodwin, P., Thomson, M.E., Gönül, S., & Pollock, A.C. (2009). The Relative Influence of Advice from Human Experts and Statistical Methods on Forecast Adjustments. *Journal of Behavioral Decision Making*, 22, 390-409.

Paluch, Stefanie, Jochen Wirtz, &Werner Kunz (2022), "The Service Robot Revolution", forthcoming in Research Handbook on Services Management.

Paulhus, D. L., & Martin, C. L. (1988). Functional Flexibility: A New Conception of Interpersonal Flexibility. *Journal of Personality and Social Psychology*, 55(1), 88–101.

Promberger, M., & Baron, J. (2006). Do Patients Trust Computers? *Journal of Behavioral Decision Making*, 19 (5), 455–68.

Research and Markets (2021). Mordor Intelligence (2021, January). *Chatbot Market – Growth, Trends, COVID-19 Impact, and Forecasts (2021 – 2026).*

Shostack, G. L. (1987). Service Positioning through Structural Change. *Journal of Marketing*, 51 (1), 34–43.

Walster, E., Berscheid, E., & Walster, G. W. (1973). New Directions in Equity Research. *Journal of Personality and Social Psychology*, 25 (2), 151–76.

Weiner, B. (2000). Attributional thoughts about consumer behavior. *Journal of Consumer Research*, 27, 382-387.

Wirtz, J. P., & Mattila, A. S. (2004). Consumer Responses to Compensation, Speed of Recovery and Apology After a Service Failure. *International Journal of Service Industry Management*, 15(2), 150-166.

Wirtz, J. P., Patterson, G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave New World: Service Robots in the Frontline. *Journal of Service Management*, 29 (5), 907–31.

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making Sense of Recommendations. *Journal of Behavioral Decision Making*, 32 (4), 403–14.

Yogeeswaran, K. J. Z., Livingstone, M., Bartneck, C., Sumioka, H., & Ishiguro, H. (2016). The Interactive Effects of Robot Anthropomorphism and Robot Ability on Perceived Threat and Support for Robotics Research. *Journal of Human-Robot Interaction*, 5 (2), 29.

Table 1. Means (SD) as a Function of Service Agent Type, Study 1

Dependent Variable	Service Agent Type	Means (SD)
Service Evaluation	Chatbot	2.77^{a} (1.48)
	Human	2.04^{b} (1.29)
Expected Flexibility	Chatbot	3.33^{a} (1.50)
	Human	4.15^{b} (1.51)

Note. The numbers with different superscripts are significantly different from each other at $p < .05$.

Table 2. Means (SD) as a Function of Service Agent Type, Study 2

Dependent Variable	Service Agent Type	Means (SD)
Service Evaluation	Chatbot	4.57^{a} (1.37)
	Human	3.95^{b} (1.51)

Note. The numbers with different superscripts are significantly different from each other at $p < .05$.

Table 3. Means (SD) as a Function of Service Agent Type, Study 3

Dependent Variable	Service Agent Type	Means (SD)
Service Evaluation	Chatbot	5.16^{a} (1.20)
	Human	4.46^{b} (1.47)
Recommendation Intention	Chatbot	4.34^{a} (1.44)
	Human	3.83^{b} (1.33)
Attitude Toward the Company	Chatbot	4.61^{a} (1.16)
	Human	4.14^{b} (1.28)

Note. The numbers with different superscripts are significantly different from each other at $p < .05$.

Table 4. Means (SD) as a Function of Service Agent Type, Study 4A

Dependent Variable	Service Agent Type	Means (SD)
Service Evaluation	Chatbot	4.03^{a} (1.39)
	Human	4.91^{b} (1.26)
Expected Flexibility	Chatbot	3.78^{a} (1.51)
	Human	4.95^{b} (1.37)

Note. The numbers with different superscripts are significantly different from each other at $p < .05$.

Table 5. Means (SD) as a Function of Service Agent Type and Service Handling Status, Study 4B

		Service Agent Type	
Dependent Variable	Service Handling Status	Human	Chatbot
Service Evaluation	Success	6.11[a] (.98)	5.63[b] (1.36)
	Failure	1.62[c] (.95)	2.04[d] (1.56)
Attitude Toward the Online Store	Success	6.04[a] (1.09)	5.66[b] (1.40)
	Failure	1.59[c] (.84)	2.00[d] (1.34)

Note. The numbers with different superscripts are significantly different from each other at $p < .05$.

Table 6. Means (SD) as a Function of Service Agent Type and Apology Condition, Study 5

		Service Agent Type	
Dependent Variable	Apology	Human	Chatbot
Service Evaluation	Emotional Message	5.32[a] (1.44)	4.62[b] (1.76)
	Non- Emotional Message	4.31[c] (1.65)	4.75[b] (1.43)
Intention to Contact the Service Agent in the Future	Emotional Message	5.22[a] (1.32)	4.95[a] (1.68)
	Non- Emotional Message	4.00[b] (1.89)	4.59[a] (1.65)
Perception of the Apology	Emotional Message	5.33[a] (1.45)	4.89[b] (1.80)
	Non- Emotional Message	4.09[b] (1.97)	4.44[a, b] (1.83)

Note. The numbers with different superscripts are significantly different from each other at $p < .05$.