HUMAN OR CHATBOT SERVICE AGENTS IN ONLINE SERVICE RECOVERY? AN EXPERIMENTAL STUDY ON CUSTOMER JUSTICE EVALUATIONS

ABSTRACT

Interactions between consumers and companies have shifted to the digital sphere, resulting in a transition from human-driven to technology-prominent services encounters Chatbots are increasingly used in customer service despite findings indicating that consumers tend to be skeptical of them and often prefer human service agents. Research investigating how consumers perceive chatbots compared to human service agents in a service failure and recovery context is still scarce. The experimental study investigates if customers' perceptions of interactional justice differ depending on whether a human service agent or chatbot agent conveys the service recovery response in a critical or uncritical recovery situation. Moreover, the degree of anthropomorphism of the chatbot agent is considered.

Keywords: Service Agents, Chatbots, Anthropomorphism, Service Recovery, Interactional Justice

INTRODUCTION AND RESEARCH OBJECTIVES

Interactions between consumers and companies have shifted to the digital sphere, resulting in a transition from human-driven to technology-prominent services encounters (Larivière et al., 2017). More and more companies are deploying chatbots instead of human employees for their customer service since they can quickly handle a large amount of consumer requests at any time of day, while being very cost-effective to operate (Wirtz et al., 2018). However, chatbots are often considered less competent and incapable of expressing and understanding emotions to the same extent as human service agents (Mozafari, Weiger, & Hammerschmidt, 2022). This negative customer impression is particularly relevant when it comes to service failures and service recovery encounters as these often involve upset customers (Crolic, Thomaz, Hadi, & Stephen, 2022). These encounters, accordingly, call for perceptions of empathy on the service agent's side (Crolic et al., 2022). Tsai, Lun, Carcioppolo, and Chuan (2021) show that interaction satisfaction with a chatbot is significantly lower compared to a human service agent for disgruntled customers. Given these research findings, it seems pertinent for companies to know how consumers react to service recovery responses provided by chatbots in comparison to humans.

One of the most widely applied theoretical framework to examine the effects of service recovery on consumers is justice theory. The important dimension of *interactional* justice entails customers' perceptions of the manner and treatment during the resolution of the service issued by the company and its service agents (McColl-Kennedy & Sparks, 2003). It is associated with empathy, politeness, effort, and honesty on the part of the service agent (del Río-Lanza, Vázquez-Casielles, & Díaz-Martín, 2009). Justice-based service recovery research so far has solely focused on service recovery delivered by human service agents. Investigation of chatbots within the scope of service recovery remains lacking despite prior calls for research (Blut, Wang, Wünderlich, & Brock, 2021). Moreover, Singh and Crisafulli (2016) point out that the role of technology-mediated encounters and associated perceptions of interactional justice in online service recovery has been scarcely explored. Given consumers' belief that chatbots lack emotional capabilities compared to humans (Cheng, Bao, Zarifis, Gong, & Mou, 2021; Tsai et al., 2021), we first investigate whether consumers perceive interactional justice differently when service recovery is performed by a chatbot compared to a human service agent.

Moreover, it is not only important for firms to know whether to deploy chatbots in a service recovery context but also how to design them (Crolic et al., 2022). Research has paid particular attention to the question of whether chatbots should be imbued with human characteristics and appearance (e.g., Rapp, Curti, & Boldi, 2021). Research has shown that the degree to which a chatbot is viewed as human-like influences how it is perceived by consumers (Rapp et al., 2021). Findings regarding the effect of anthropomorphism, however, have been mixed. Following calls for research on the role of anthropomorphism regarding consumer reactions to perceived justice (Blut et al., 2021), we further investigate whether anthropomorphism of the chatbot influences consumers' perceptions of interactional justice.

Further, prior research illustrates that service criticality can affect customer evaluations of service failure and recovery (e.g., Crisafulli & Singh, 2017; Jafarzadeh, Tafti, Intezari, & Sohrabi, 2021; Webster & Sundaram, 1998). Service criticality refers to the perceived importance of whether the service is successfully provided in a given service encounter (Webster & Sundaram, 1998). "When a purchase occasion is considered very important or critical as determined by the magnitude of the consequences in the event of service failure, customers are likely to view service failure more seriously than when service purchase is less critical" (Webster & Sundaram, 1998, p. 154). It has been found that consumers are also more

reluctant to rely on chatbots when impeccable service delivery is highly relevant to them (Blut et al., 2021; Mozafari et al., 2022). Thus, we additionally investigate whether service criticality influences the relation between the type of service agent delivering the service recovery response and consumers' interactional justice perceptions.

CONCEPTUAL AND THEORETICAL BACKGROUND

Fundamentals of justice theory

Justice theory is one of the most widely applied theoretical frameworks in service failure and recovery research as it has been deemed suitable for examining consumers' reactions to conflict situations, such as service failure and recovery encounters (Orsingher, Valentini, & De Angelis, 2010). While most companies strive to deliver impeccable and high-quality service, even the best service providers cannot entirely avoid failures (Kelley & Davis, 1994; Kuo & Wu, 2012). Justice theory states that consumers' post-recovery attitudes and behavior are determined by whether they feel they are treated fairly as to the outcome (distributive justice), the process (procedural justice) and the interaction (interactional justice) of recovery (McColl-Kennedy & Sparks, 2003). The dimension of interactional justice focuses on whether customers feel that the service organization and its agents treat them fairly while rectifying the service problem (McColl-Kennedy & Sparks, 2003). Empathy, politeness, effort, and honesty on the part of service personnel are usually considered key facets of the fair interpersonal treatment associated with interactional justice (del Río-Lanza et al., 2009). Empathy in this context can be defined as the capability to experience emotions and to discern and understand other individuals' affective states and respond accordingly (Paiva, Leite, Boukricha, & Wachsmuth, 2017; Simon, 2013).

Anthropomorphism

Anthropomorphism refers to the attribution of "humanlike properties, characteristics, or mental states to real or imagined nonhuman agents and objects" (Epley, Waytz, & Cacioppo, 2007, p. 865). Anthropomorphism of chatbots thus describes the extent to which the chatbot is equipped with properties that are perceived as humanlike: a name or avatar picture, for example (e.g., Crolic et al., 2022). There are three main factors driving anthropomorphism (Epley, 2018; Söderlund & Oikarinen, 2021): The first is the innate human need for social relations that serves as a motivation to perceive mind in nonhuman agents and consequently imbue them with humanlike characteristics. The second motivational factor entails the wish to explain, predict and hence to a certain extent control another being. If an agent's behavior requires explanation, no matter whether human or nonhuman, perceptions of mind are evoked. This can be linked back to the tendency of humans to assign mind to other agents, including notions such as intentions, desires, attitudes, and beliefs, to make sense of the behavior of the other being. Perceived similarity represents the third factor guiding anthropomorphism. The more an agent exhibits typical human behavior and appearance, the more likely it is to trigger perceptions of a humanlike mind, thus fostering anthropomorphism.

Mind Perception Theory

Theory of mind perception suggests that individuals assign mind to both humans and nonhumans along two dimensions. One dimension is called experience and alludes to the capacity to feel emotions and sensations. The other dimension is named agency and describes the ability to think, act with purpose, and exercise self-control (H. M. Gray, Gray, & Wegner, 2007; Wang & Krumhuber, 2018).

Out of these two dimensions, experience is considered more integral to being human than agency (K. Gray & Wegner, 2012; Söderlund & Oikarinen, 2021). There are several reasons

for this: First, a person without the ability to feel emotions and sensations is perceived as more unsettling than a person without agency, indicating that experience is more important to what constitutes being human (Gray and Wegner 2012). Second, in contrast to agency, experience is less often ascribed to non-human entities and thus seems more indicative of humanness than agency (Söderlund & Oikarinen, 2021). Consequently, adult humans are perceived as having both a high level of agency and experience while robots are viewed as having some agency but only little experience (H. M. Gray et al., 2007). Hence, while people tend to attribute certain cognitive capabilities to robots and thus chatbots, they believe them to be less capable of experiencing emotions, even if they serve a social function (Wang & Krumhuber, 2018).

HYPOTHESES

Since empathy of the service agent is critical to perceptions of interactional justice (del Río-Lanza et al., 2009), a service agent whom consumers believe can provide an empathetic service recovery response should elicit higher interactional justice than a service agent who cannot. According to Gray, Gray, and Wegner's (2007) theory of mind perception, people ascribe a high level of both agency and experience to adult humans but only some agency and little experience to robots. Experience encompasses the ability to feel sensations and emotions. It is thus highly related to empathy (Simon, 2013). Several studies indicate that consumers consider chatbots to be less capable of understanding emotions and expressing empathy than human service agents. For instance, in their in-depth qualitative study with forty participants, Pitardi, Wirtz, Paluch, and Kunz (2021) find out that respondents did not expect robots to exhibit em pathy. Similarly, Luo, Tong, Fang, and Qu (2019) reveal that chatbot identity disclosure lowers purchase rates and shortens the duration of a sales call because customers view the chatbot as less empathetic compared to human agents. People might view emotion as a "biological reaction und subjective experience" that cannot be replicated through technical programming, thus explaining their inability to consider machines capable of feeling genuine emotions (Huang & Rust, 2018, p. 159). Overall, it can therefore be concluded that consumers will perceive a human service agent as being more capable of showing empathy than a chatbot during service recovery, thus assigning a higher level of perceived interactional justice to the human service agent. Hence:
H1: The service agent conveying the service recovery response has an effect on perceived interactional justice, such that interactional justice is higher when a human service agent conveys the recovery response compared to a chatbot agent.

As previously outlined, robots are usually viewed as having low levels of agency and experience while adult humans are seen as having the capacity to act and feel to a high extent (H. M. Gray et al., 2007). Anthropomorphism, however, can trigger mental processes in people that cause them to react to anthropomorphic beings in a similar way as they do to humans (Epley, 2018). Prior research findings indicate that people attribute more agency to anthropomorphic entities because they associate humanlike characteristics and appearance with higher levels of cognitive intelligence and corresponding performance expectations (Blut et al., 2021). Considering that anthropomorphism increases perceptions of the chatbot having a higher level of agency, it stands to reason that it also fosters the impression of the chatbot having more experience and thus being more capable of empathy. This assumption is supported by Gray and Wegner (2012) who found out that a humanlike appearance heightens the level of experience humans assign to robots. Since being human is even more strongly associated with experience than with agency, it is likely that the more human a chatbot seems, the more it would appear to be capable of feeling and understanding emotions (K. Gray & Wegner, 2012). Since empathetic behavior by the service employee plays a critical role in ensuring interactional justice, anthropomorphism of the chatbot should lead to higher levels of interactional fairness (del Río-

Lanza et al., 2009). Hence, we propose that anthropomorphism of the chatbot has a positive influence on perceptions of interactional justice:
H2: Interactional justice will be higher when an anthropomorphized chatbot conveys the service recovery response compared to a non-anthropomorphized chatbot.

Prior research illustrates that the evaluation of a service is also influenced by the criticality of the service. Since the importance of service provision depends on how serious the consequences of a service failure are perceived to be, customers are likely to view a service error more seriously in high service criticality settings (Ostrom & Iacobucci, 1995). Consumers therefore appear to be less susceptible to service recovery efforts in service encounters when they urgently require a flawless delivery of service (Webster & Sundaram, 1998). Research shows that criticality of service consumption not only affects consumers' evaluations of service recovery (e.g., Crisafulli & Singh, 2017) but also their perception and evaluation of chatbots (e.g., Blut et al., 2021). Consumers are more reluctant to rely on chatbots when flawless service consumption plays an important role for them. Mozafari et al. (2022) show that in highly critical service encounters, stereotypes regarding chatbots are activated and negatively affect customer trust. In particular, the decrease of trust stems from lower perceptions of the chatbot's competence and benevolence. The decline in benevolence belief can be attributed to lower perceived empathy. These findings indicate that the preference for a human service agent might be higher in the case of high service criticality. Thus, we propose:
H3: For high service criticality situations (vs. low service criticality situations), the difference between the human service agent and the chatbot agent regarding the level of interactional justice is greater.

EMPIRICAL STUDY

We conducted a 3x2 between-subjects scenario-based experiment. 237 participants took part in the main study (mean age = 37.89; 30% male) The service agent conveying the service recovery response was varied at three levels (human service agent, anthropomorphized chatbot, non-anthropomorphized chatbot) and service criticality at two levels (low vs. high). Participants were randomly assigned to one of the six groups. A fashion retail setting was chosen because online apparel retailers are among the service providers where consumers frequently face service failure and recovery situations. A late delivery of clothing items purchased online was selected as the service failure because it is one of the most commonly reported online service failures. Participants of each scenario were told that they have placed an order with an online retailer that is delayed by several days. The scenario further states that they have contacted the retailer via an online chat on the retailer's website and explained their problem to a customer service agent. The ensuing chat history between the customer and the customer service agent was then displayed. We focused on text-based chatbots because consumers interact with them almost twice as often as they do with voicebots (Monard et al., 2021). The scenario explicitly mentioned that the ordered clothes are delivered soon after the customer contacted the retailer, signifying a resolution of the service failure.

To manipulate the service agent, participants in the human service agent condition were shown a realistic picture of a female service agent. To depict an anthropomorphized chatbot, previous studies have used cartoon-like illustrations of a human (Crolic et al., 2022). Thus, the anthropomorphized chatbot was represented by a semi-realistic Memoji that was based on the image of the human service agent. The non-anthropomorphized chatbot was depicted by an avatar image that resembled a robot (Go & Sundar, 2019). The human service agent and anthropomorphized chatbot were chosen to be female because there a higher proportion of women in customer service (Al-Hussaini, 2022). Additionally, each service agent had a unique

text when introducing themselves to the customer (e.g., Crolic et al., 2022). For instance, the human service agent introduced themselves with the line "Hi, I'm Alex, your customer service agent for today" while the anthropomorphized chatbot used the words "Hi, I'm Alex, the customer service chatbot". To manipulate the level of service criticality, participants in the high service criticality condition were told that they urgently need the clothing items for an important social gathering and are worried that the order will not arrive in time for this special event. In the low criticality condition, this information was left out.

Items for the manipulation checks and the dependent variables are based on prior research (Crisafulli & Singh, 2017; Crolic et al., 2022; Mozafari et al., 2022; Simon, 2013). 7-point scales were used wherever possible (with 7 = strongly agree). In the experiment, manipulations were successful. Participants correctly identified the extent to which they interacted with a human service agent or a chatbot (M_{Human} = 3.65, M_{Anthro} = 5.11, $M_{Non\text{-}anthro}$ = 5.37). Moreover, respondents in the anthropomorphized chatbot scenario perceived the chatbot as more humanlike than the chatbot in the non-anthropomorphized chatbot scenario (M_{Anthro} = 4.10, $M_{Non\text{-}anthro}$ = 3.54; $p < .01$). Participants in the high service criticality condition considered the service delivery as more critical to them than the respondents in the low criticality condition ($M_{HighCrit}$ = 6.18, $M_{LowCrit}$ = 4.05; $p < .001$).

Major ANOVA results demonstrate that interactional justice is higher for the human service agent (M_{Human} = 5.20) compared to an anthropomorphized chatbot (M_{Anthro} = 4.70) and a non-anthropomorphized chatbot ($M_{Non\text{-}anthro}$ = 4.44) ($p = .001$). Thus, confirming H1 and H2. Moreover, there is a significant interaction effect between service criticality and the agent conveying the service recovery response ($p = .062$), such that in high service criticality situations, the level of perceived interactional justice is higher for all service agents types (M_{Human} = 5.53; M_{Anthro} = 4.94, $M_{Non\text{-}anthro}$ = 4.39) than in low service criticality situations M_{Human} = 4.89; M_{Anthro} = 4.45, $M_{Non\text{-}anthro}$ = 4.61). H3 is thus also confirmed.

CONCLUSION

As companies increasingly use chatbots for customer service, it is critical for them to gain a better understanding of how this technology impacts service interactions with consumers. The present study seeks to provide empirical insight on consumers' reactions to chatbots compared to human service agents in a service recovery context. It is the first study to bridge the extensive service failure and recovery literature to the field of chatbot research by examining if customers' perceptions of interactional justice differ depending on whether a human service agent or a chatbot delivers the service recovery response. Service providers need to be cautious with regard to perceptions of interactional justice when using chatbots in service recovery. Since interactional justice is significantly higher for the human service agent compared to a non-anthropomorphized chatbot, companies should refrain from using non-anthropomorphic chatbots in situations where interactional justice perceptions might be particularly important. Prior research indicates that interactional justice is especially relevant for long-term service relationships (Gelbrich & Roschk, 2011). It is therefore recommended that either human service agents or anthropomorphized chatbots handle service recovery for customers whose long-term retention is particularly important. As to limitations and future research, one has to keep in mind that study participants were instructed to read a conversation with the service agent based on a hypothetical scenario instead of engaging in an actual texting exchange in order to establish high internal validity. However, this experimental design may not accurately capture the conditions in the real world. Future studies should therefore conduct field experiments comparing actual service recovery interactions between consumers and chatbots as well as human service agents to raise external validity.

REFERENCES

Al-Hussaini, S. (2022). Why women should shape the future of customer service. Retrieved from https://www.ultimate.ai/blog/ultimate-life/why-women-should-shape-the-future-of-customer-service

Blut, M., Wang, C., Wünderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: a meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science, 49*(4), 632-658.

Cheng, X., Bao, Y., Zarifis, A., Gong, W., & Mou, J. (2021). Exploring consumers' response to text-based chatbots in e-commerce: the moderating role of task complexity and chatbot disclosure. *Internet Research, 32*(2), 496-517.

Crisafulli, B., & Singh, J. (2017). Service failures in e-retailing: Examining the effects of response time, compensation, and service criticality. *Computers in Human Behavior, 77*, 413-424.

Crolic, C., Thomaz, F., Hadi, R., & Stephen, A. T. (2022). Blame the bot: Anthropomorphism and anger in customer–chatbot interactions. *Journal of Marketing, 86*(1), 132-148.

del Río-Lanza, A. B., Vázquez-Casielles, R., & Díaz-Martín, A. M. (2009). Satisfaction with service recovery: Perceived justice and emotional responses. *Journal of Business Research, 62*(8), 775-781.

Epley, N. (2018). A mind like mine: The exceptionally ordinary underpinnings of anthropomorphism. *Journal of the Association for Consumer Research, 3*(4), 591-598.

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

Gelbrich, K., & Roschk, H. (2011). A meta-analysis of organizational complaint handling and customer responses. *Journal of Service Research, 14*(1), 24-43.

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior, 97*, 304-316.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *science, 315*, 619.

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition, 125*(1), 125-130.

Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research, 21*(2), 155-172.

Jafarzadeh, H., Tafti, M., Intezari, A., & Sohrabi, B. (2021). All's well that ends well: Effective recovery from failures during the delivery phase of e-retailing process. *Journal of Retailing and Consumer Services, 62*(September), 102602.

Kelley, S. W., & Davis, M. A. (1994). Antecedents to customer expectations for service recovery. *Journal of the Academy of Marketing Science, 22*(1), 52-61.

Kuo, Y.-F., & Wu, C.-M. (2012). Satisfaction and post-purchase intentions with service recovery of online shopping websites: Perspectives on perceived justice and emotions. *International Journal of Information Management, 32*(2), 127-138.

Larivière, B., Bowen, D., Andreassen, T. W., Kunz, W., Sirianni, N. J., Voss, C., . . . De Keyser, A. (2017). "Service encounter 2.0": An investigation into the roles of technology, employees and customers. *Journal of Business Research, 79*(October), 238-246.

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.

McColl-Kennedy, J. R., & Sparks, B. A. (2003). Application of fairness theory to service failures and service recovery. *Journal of Service Research, 5*(3), 251-266.

Monard, F., Uebersax, H.-P., Müller, M., Hannich, F. M., Furchheim, P., & Kaya, T. (2021). Chatbot Studie ´21 [Chat bot study ´21]. Retrieved from https://www.intre.cc/wp-content/uploads/2021/09/Chatbot-Studie-2021_aiaibot_ZHAW.pdf.

Mozafari, N., Weiger, W. H., & Hammerschmidt, M. (2022). Trust me, I'm a bot–repercussions of chatbot disclosure in different service frontline settings. *Journal of Service Management, 33*(2), 221-245.

Orsingher, C., Valentini, S., & De Angelis, M. (2010). A meta-analysis of satisfaction with complaint handling in services. *Journal of the Academy of Marketing Science, 38*, 169-186.

Ostrom, A., & Iacobucci, D. (1995). Consumer trade-offs and the evaluation of services. *Journal of Marketing, 59*(1), 17-28.

Paiva, A., Leite, I., Boukricha, H., & Wachsmuth, I. (2017). Empathy in virtual agents and robots: A survey. *ACM Transactions on Interactive Intelligent Systems (TiiS), 7*(3), 1-40.

Pitardi, V., Wirtz, J., Paluch, S., & Kunz, W. H. (2021). Service robots, agency and embarrassing service encounters. *Journal of Service Management, 33*(2), 389-414.

Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies, 151*, 1-24.

Simon, F. (2013). The influence of empathy in complaint handling: Evidence of gratitudinal and transactional routes to loyalty. *Journal of Retailing and Consumer Services, 20*(6), 599-608.

Singh, J., & Crisafulli, B. (2016). Managing online service recovery: procedures, justice and customer satisfaction. *Journal of Service Theory and Practice, 26*(6), 764-787.

Söderlund, M., & Oikarinen, E.-L. (2021). Service encounters with virtual agents: an examination of perceived humanness as a source of customer satisfaction. *European Journal of Marketing, 55*(13), 94-121.

Tsai, W. H. S., Lun, D., Carcioppolo, N., & Chuan, C. H. (2021). Human versus chatbot: Understanding the role of emotion in health marketing communication for vaccines. *Psychology & Marketing, 38*(12), 2377-2392.

Wang, X., & Krumhuber, E. G. (2018). Mind perception of robots varies with their economic versus social function. *Frontiers in psychology, 9*, 1-10.

Webster, C., & Sundaram, D. S. (1998). Service consumption criticality in failure recovery. *Journal of Business Research, 41*(2), 153-159.

Wirtz, J., Patterson, P. 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), 907-931.