

Tsekouras, Dimitrios; Gutt, Dominik; Heimbach, Irina

**Article — Published Version**

## The robo bias in conversational reviews: How the solicitation medium anthropomorphism affects product rating valence and review helpfulness

Journal of the Academy of Marketing Science

**Provided in Cooperation with:**

Springer Nature

*Suggested Citation:* Tsekouras, Dimitrios; Gutt, Dominik; Heimbach, Irina (2024) : The robo bias in conversational reviews: How the solicitation medium anthropomorphism affects product rating valence and review helpfulness, Journal of the Academy of Marketing Science, ISSN 1552-7824, Springer US, New York, NY, Vol. 52, Iss. 6, pp. 1651-1672, <https://doi.org/10.1007/s11747-024-01027-8>

This Version is available at:

<https://hdl.handle.net/10419/315700>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>



# The robo bias in conversational reviews: How the solicitation medium anthropomorphism affects product rating valence and review helpfulness

Dimitrios Tsekouras<sup>1</sup> · Dominik Gutt<sup>1</sup> · Irina Heimbach<sup>2</sup>

Received: 26 July 2022 / Accepted: 10 May 2024 / Published online: 6 June 2024  
© The Author(s) 2024

## Abstract

Companies are increasingly introducing *conversational reviews*—reviews solicited via chatbots—to gain customer feedback. However, little is known about how chatbot-mediated solicitation influences rating valence and review helpfulness compared to conventional online forms. Therefore, we conceptualized these review solicitation media on the continuum of anthropomorphism and investigated how various levels of anthropomorphism affect rating valence and review helpfulness, showing that more anthropomorphic media lead to more positive and less helpful reviews. We found that moderate levels of anthropomorphism lead to increased interaction enjoyment, and high levels increase social presence, thus inflating the rating valence and decreasing review helpfulness. Further, the effect of anthropomorphism remains robust across review solicitors' salience (sellers vs. platforms) and expressed emotionality in conversations. Our study is among the first to investigate chatbots as a new form of technology to solicit online reviews, providing insights to inform various stakeholders of the advantages, drawbacks, and potential ethical concerns of anthropomorphic technology in customer feedback solicitation.

**Keywords** Product reviews · Chatbots · Review solicitation · Anthropomorphism · Social presence · Interaction enjoyment

## Introduction

Conversational agents or, more colloquially, chatbots, are “software-based system[s] designed to interact with humans using natural language” (Feine et al., 2019, p. 3). They aim to simulate human conversation by integrating a language model and computational algorithms to mimic informal communication, or chats, and execute tasks between a human user and a computer using natural language (Araujo, 2018). Firms widely use chatbots across various functions and contexts, including marketing (Thomaz et al., 2020), sales (Gartner, 2018; Luo et al., 2019), finance (Luo et al., 2019), health care (Health Europa, 2019), and education (Tourangeau et al., 2003). Firms that introduce chatbots in the context of customer-firm interactions are often trying to improve their operational efficiency (reducing costs and customer response times) and increase their customer satisfaction and revenues (Radziwill & Benton, 2017; Reddy, 2017).

---

Dimitrios Tsekouras, Dominik Gutt and Irina Heimbach contributed equally.

---

Yany Gregoire served as Area Editor for this article.

---

✉ Dominik Gutt  
gutt@rsm.nl  
Dimitrios Tsekouras  
dtsekouras@rsm.nl  
Irina Heimbach  
irina.heimbach@whu.edu

<sup>1</sup> Erasmus University Rotterdam, 3000 DR, Rotterdam Postbus 1738, Netherlands

<sup>2</sup> WHU – Otto Beisheim School of Management, Burgplatz 2, 56179 Vallendar, Germany

Notwithstanding these motives, if the recently emerging literature on chatbots has taught us one thing, it is that their introduction rarely comes without side effects, be they negative or positive. On the encouraging side, firms that introduce chatbots may experience positive investor responses (Fotheringham & Wiles, 2022). Also, chatbots increase affective trust and improve consumer perceptions (Hildebrand & Bergner, 2021). On the downside, chatbots can prompt unethical consumer behavior (Kim et al., 2022), decrease customer satisfaction among angry customers (Crollic et al., 2022), increase customers' assertiveness in negotiations (Schanke et al., 2021), and decrease customer purchases when chatbot use is disclosed (Luo et al., 2019). Therefore, firms should carefully consider whether to introduce chatbots in a particular domain.

One trending area that deserves attention is that of *conversational reviews*—the acquisition of customer reviews via chatbots (Haptik, 2018)—because multiple e-commerce firms are increasingly implementing this practice, both on their platforms and via third-party ones, such as Facebook Messenger (Zhang, 2018). Online firms have always been interested in motivating their consumers to share their product or service experiences (Burtch et al., 2018; Gutt et al., 2019; Tsekouras, 2017) because the online review volume and valence as well as the informational value (e.g., review helpfulness) can facilitate consumer decision-making and increase future sales (You et al., 2015; Forman et al., 2008). Consequently, the goal of this study is to examine whether and how the use of chatbots for online review collection affects these key metrics of the resulting reviews. Motivated by the previous research on chatbots, we concentrate on the level of human likeness—hereafter, anthropomorphism—which is the central aspect in the design of chatbots. One can equip chatbots with a humanlike identity and visual appearance, thus adding a personal touch and enabling them to hold a humanlike conversation (Feine et al., 2019). Because previous research has shown that the effects of anthropomorphic design of products and technology on consumer responses are ambiguous (Luo et al., 2019; Schanke et al., 2021; Crollic et al., 2022), this study aims to answer the following research question: *How does the anthropomorphism of review solicitation media affect online rating valence and review helpfulness?*

Across four studies, we show that anthropomorphism engendered in chatbots positively biases online ratings compared to a conventional review form. We trace two mechanisms through which this bias operates. At a moderate level of anthropomorphism, the bias operates through perceived interaction enjoyment—“the extent to which the activity of using the computer is perceived

to be enjoyable in its own right, apart from any performance consequences that may be anticipated” (Davis et al., 1992, p. 1113), whereas at a high level of anthropomorphism, the bias operates through increased social presence—the “degree of salience of the other person in a mediated communication and the consequent salience of their interpersonal interaction” (Short et al., 1976, p. 65). We also demonstrate that plausible behavioral interventions intended to mute the interaction enjoyment mechanism and the social presence mechanism are unsuccessful at mitigating the bias. Finally, we show that readers assess chatbot-mediated reviews as less helpful than reviews generated using conventional forms. We suggest the fact that human-chatbot interaction leads to shorter reviews, which in turn convey fewer rich arguments, as a potential explanation for this effect.

With these findings, we argue for several theoretical contributions to literature. First, we contribute to the stream of literature on the effects of anthropomorphism on consumer behavior. To the best of our knowledge, we are one of the first to conceptualize and present empirical evidence for how anthropomorphism affects consumers' online reviewing behavior. Although past studies predominantly focused on how anthropomorphism affects consumer perception of firms, such as in terms of affective trust (Hildebrand & Bergner, 2021) and satisfaction (Crollic et al., 2022), this study extends the anthropomorphism effects to consumer perception of products.

Second, we unveil increased social presence and interaction enjoyment as the two mechanisms responsible for the anthropomorphism effect. Although social presence has been a mainstay in studies on anthropomorphism (Blut et al., 2020), we introduce the construct of interaction enjoyment to the discussion of anthropomorphism effects. In particular, we identify boundary conditions of anthropomorphism for when the interaction enjoyment mechanism actually dominates social presence and vice versa.

Third, the results respond to the call for research on the effects of artificial intelligence (AI)-enabled technology on customer-firm interactions (Luo et al., 2019). Our findings aid in understanding the role of novel review solicitation media and extend the evidence on consumer susceptibility to the reviewing environment when generating product ratings (Gutt et al., 2019; Ransbotham et al., 2019; Tsekouras, 2017).

Beyond the contributions to literature, our findings provide meaningful practical implications to managers, consumers, and policymakers. Managers need to be aware that using anthropomorphic chatbots to solicit reviews positively biases their products' ratings. At the surface, positive ratings may be desirable due to their

widely documented effects on sales (Zhu & Zhang, 2010), but managers risk misrepresenting product quality to their customers, raising serious ethical concerns and risking long-term damage to a firm's brand positioning and image (Bazaarvoice, 2022). As shown in the field of fake reviews (He et al., 2022; Luca & Zervas, 2016; Mayzlin et al., 2014), sellers of weak brands and non-branded products, especially, might disproportionately deploy chatbots to receive better ratings and thus improve sales. What may further disincentivize managers from employing anthropomorphic chatbots, though, is our finding that they decrease review helpfulness. This latter discovery may work as a natural deterrent for the use of anthropomorphic chatbots. If consumers use positively biased ratings that are unhelpful toward informing their decision-making, they can make bad purchase decisions. To mentally discount the positivity bias when screening a list of reviews, consumers would need to know which ratings were collected by means of chatbots.

Finally, policymakers might need to step in to establish regulations that keep fraudulent businesses from positively biasing their ratings and harm customer decision-making. Although many government bodies, such as the US Federal Trade Commission's (FTC) and German Federal Cartel Office (Bundeskartellamt), have issued guidelines arguing for keeping online reviews accurate and unbiased, guidelines such as the FTC's (FTC, 2024) guide on soliciting reviews may be adapted to incorporate guidelines on the ethical use of chatbots for review solicitation.

The remainder of this paper is structured as follows. First, we locate our research within the context of existing literature and supply a theoretical background. Next, we present the methods and results from four empirical studies. Finally, we discuss the implications of our research, comment on the limitations, and offer directions for future research.

## Related literature

Our study pertains to two streams of literature. The first one is concerned with ethical issues in online review collection, and the second is concerned with the effects of anthropomorphic conversational agents on consumer behavior.

### Ethical issues in online review collection

In light of the beneficial sales effect of online reviews for businesses (Zhu & Zhang, 2010), some may disregard ethical concerns and increase their average rating valence or the number of reviews by malicious practices. The most

popular practice is to collect positive fake reviews. Studies report that between 5% (Anderson & Simester, 2014) and 15% (Luca & Zervas, 2016) of all reviews might be fake. Despite ethical concerns, fake reviews are effective for businesses. Faking positive reviews can increase a business's visibility in review-based rankings by up to 40% (Lappas et al., 2016) and can increase a business's sales by up to 16% (He et al., 2022). This raises serious ethical concerns because it shows that engaging in fraudulent review collection can pay off for businesses. Businesses most likely to collect fake reviews include those that are non-branded (Mayzlin et al., 2014), those that rely heavily on reviews (Luca & Zervas, 2016), and those that face high competition (Luca & Zervas, 2016). Also, sellers oftentimes collect positive fake reviews, especially for low-quality products (He et al., 2022), which is particularly detrimental to the consumer who may be lured into buying poorly made products.

Unlike fake reviews, which represent an intentional deception of buyers, our study aims to show how a deliberate anthropomorphic design of chatbots to collect customer reviews might generate unintended negative side effects in the form of inflated product ratings. Given that online sellers often experiment with new technologies via A/B tests, the discoverers of the effect will be incentivized to keep their discovery a secret and utilize it for their purposes. This utilization of highly anthropomorphic chatbots to collect product reviews raises ethical concerns because other sellers will have a competitive disadvantage and buyers will be exposed to distorted product reviews. Therefore, our research aims at raising awareness for this phenomenon and at informing all stakeholders, buyers, sellers, and policymakers about the existence of these effects such that each stakeholder group can make an informed response to the usage of anthropomorphic technology.

### The effects of chatbot anthropomorphism on consumer behavior

A growing number of empirical studies examine how conversational agent anthropomorphism affects consumer perception of the firm and subsequent consumer behavior (see Table 1). Crollic et al. (2022) showed that high levels of anthropomorphism are less effective in interactions with angry customers because of the heightened efficacy expectancies of those chatbots. When facing angry consumers, anthropomorphic chatbots decrease customer satisfaction, firm evaluations, and purchase intention. Regardless of the current emotional status, chatbots can also reduce consumers' perceptions of anticipatory guilt toward firms when making false claims (Kim et

**Table 1** Summary of previous research on the effect of chatbot anthropomorphism on consumer behavior

Study	Context	Theory Basis	Experimental chatbot conditions	Psychological process	Dependent variable	Finding
Crolic et al. (2022)	Conversational agent for customer service of angry customers	Functionalism theory of emotion, Appraisal theory	Low vs. high anthropomorphic bot	Expectations	Customer satisfaction, firm evaluation, purchase intention	Chatbots decrease the customer satisfaction, firm evaluation, and purchase intention of angry customers.
Kim et al. (2022)	Conversational agent for customer service	Unethical consumer behavior	Human vs. chatbot	Anticipatory guilt	Providing false information, claim coupon eligibility	Chatbots make customers provide information that is manufactured to their advantage. The effect decreases as chatbots become more anthropomorphic.
Hildebrand and Bergner (2021)	Conversational agent for financial advice	Speech formation, turn-taking	Low vs. medium vs. high anthropomorphic bot	Affective trust	Firm perception, investment behavior	Conversational agents positively affect firm perception and investment behavior.
Sachdeva et al. (2024)	Conversational agents to solicit online reviews	Genre rules	Web form vs. chatbot	Cognitive effort	Perception of review process, review characteristics	Chatbots are perceived as more efficient than forms. Structuring review text boxes may increase review length.
Araujo (2018)	Conversational shopping assistant	Embodiment	Low vs. high anthropomorphic bot	Social presence	Emotional connection, service satisfaction	Higher anthropomorphism leads to stronger emotional connection with a firm.
Adam et al. (2019)	Conversational agent for financial advice	Anthropomorphism, anchoring	Low vs. medium vs. high anthropomorphic bot	Social presence	Investment volume	Anthropomorphism leads to higher investment volume through social presence.
Gnewuch et al. (2022)	Conversational agent for customer service	Social Response Theory, Expectancy Violation Theory	Low vs. high anthropomorphic bot	Social presence	Intention to use	Response time delay increases intention to use through increased social presence
<b>This study</b>	Conversational agent to solicit online reviews	Impression management, evaluative conditioning theory, common ground	Review form vs. moderate vs. high anthropomorphic bot	Social presence, interaction enjoyment	Rating valence, review helpfulness	Anthropomorphism increases rating valence, especially for low-quality products, and decreases review helpfulness.

al., 2022). As a result of reduced anticipatory guilt, consumers tend to provide information that is manufactured to their advantage or claim coupons they are ineligible for (Kim et al., 2022). However, this effect dampens as anthropomorphism increases, suggesting that chatbot anthropomorphism may instill a sense of accountability with the consumer. Yet, customer responses to chatbots may not always be detrimental. In the context of financial services, conversational agents can improve firm perceptions in terms of affective trust and perceived firm benevolence (Hildebrand & Bergner, 2021). Although all these studies have delivered important insights on the effect of chatbots on consumer behavior, they have primarily focused on the perception of the firm. None of these studies has examined whether chatbots can affect the perceptions of a firm's products. The only exception is a recent study which shows that consumers feel more satisfied with the review process when they use chatbots, and the review length depends on the structure of the review text environment (Sachdeva et al., 2024).

Besides the effects of chatbots on customers' firm perceptions, a recurring theme in the literature is that a growing number of studies build on social presence theory (Short et al., 1976). Even though the literature converges on anthropomorphic cues in chatbots engendering social presence, nuanced, context-dependent effects are visible. For example, chatbots with anthropomorphic cues exert significantly higher social presence than chatbots without them, but only when the anthropomorphic chatbots are also framed as intelligent (Araujo, 2018). By contrast, in the context of investing, even static bots without the capability to interact with humans can be perceived as significantly more socially present when they are equipped with anthropomorphic cues (Adam et al., 2019). Customers perceive chatbots that have response time delays as more socially present than those without anthropomorphic cues (Gnewuch et al., 2022). Similar to this literature stream, we build on social presence theory, but we complement our theory basis with a new psychological process—interaction enjoyment—next to social presence. Prior studies have predominantly compared

chatbots with and without anthropomorphic cues. By adding a conventional review form as the low anthropomorphic condition, we cover a broader range of anthropomorphism on distinct levels (low, moderate, high) than has been typically used in past literature.

### Conceptual model and hypotheses

In this section, we delineate the study's hypotheses. Figure 1 depicts our conceptual model.

We follow recent marketing literature in building on the “computers-are-social-actors” (CASA) paradigm (Nass et al., 1994) to theorize on how consumers interact with anthropomorphized technology (Miao et al., 2022; Novak & Hoffman, 2019; Noble et al., 2022). According to CASA, anthropomorphic cues lead humans to relate to computers as they do with other humans (Reeves & Nass, 1996) and shape their interactions with chatbots with the same social and psychological dynamics that characterize human-to-human interactions. Past evidence has suggested that humans perceive computers as teammates when performing online tasks (Nass et al., 1996); they react to computer movements as they would to human motions (Reeves & Nass, 1996) and apply gender stereotypes and social traits, such as politeness, to computers (Nass et al., 1994). The main reason for treating computers as humans lies in the mindless reactions to social traits that computers exhibit (Nass & Moon, 2000), where the term *mindless* refers to “an overreliance on categories and distinctions drawn in the past” (Langer, 1992, p. 289). Humans can distinguish humans based on characteristic social traits and behaviors; these include physical traits, psychological cues, language sophistication, social dynamics, and social roles (Epley et al., 2007; Fogg, 2002; Reeves & Nass, 1996). Thus, the

more anthropomorphic that communication technologies are, the more humans treat them as social actors, relating to them in the same manner as they would with humans (Nass & Moon, 2000; Oh et al., 2018). The CASA paradigm represents an important premise to the development of the hypotheses that follow. The same principles that govern humans interacting with each other, such as social presence and impression management (H1, H2), enjoyment through human interaction (H3), and the feeling of common ground (H6, H7), can be applied to the human-chatbot context.

### Social presence

Designers purposefully endow technology with various anthropomorphic cues to increase social presence, which scholars believe positively affects social reactions and communication outcomes (Oh et al., 2018; Thomaz et al., 2020; Von der Pütten et al., 2010). Increased social presence makes conversational partners more salient, converging toward face-to-face communication. That shift, in turn, triggers the activation of impression management, which describes a communicator's desire to make a favorable impression on the audience (Berger, 2014; Tedeschi, 2013). Humans often conform to audience expectations to receive a reward, such as being liked, or to avoid punishment, such as social disapproval (Cialdini & Goldstein, 2004). Adjusting one's expressed opinions toward the audience's perspectives can prevent negative evaluations of oneself by others. Feeling the presence of a “real” person makes the conversational partner more identifiable and accountable. As people strive to maintain a positive self-concept (Leary & Kowalski, 1990), they tend to give favorable evaluations and refrain from unfavorable news in exchanges with a salient conversational

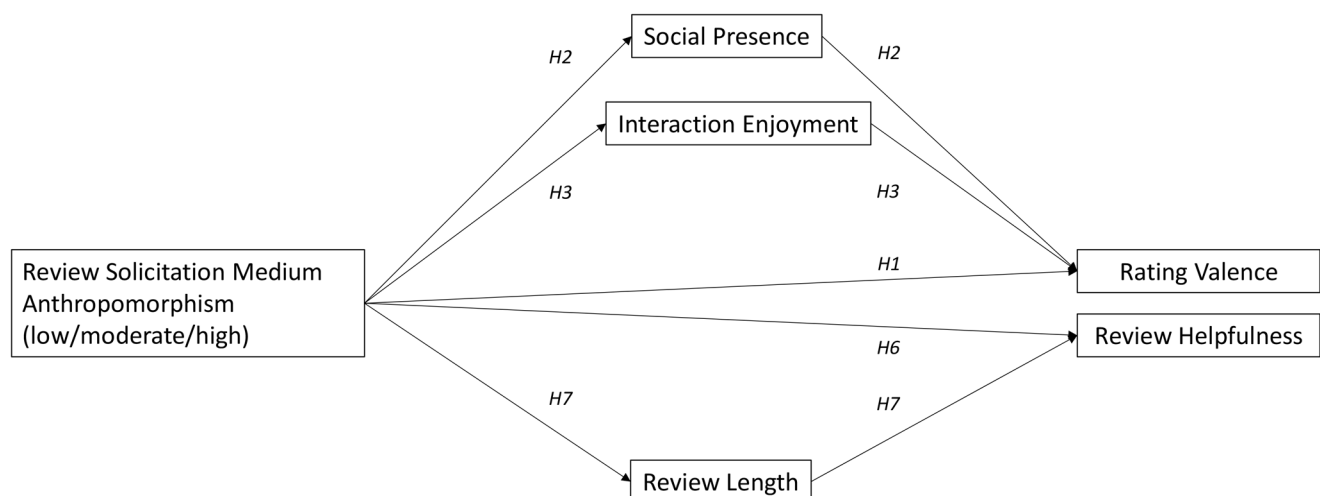


Fig. 1 Conceptual model



partner (Heath, 1996; Rosen & Tesser, 1972). Previous research has revealed how increases in perceived social presence generated with the use of a chatbot may lead to socially desirable responses (Schuetzler et al., 2018). Consequently, we expect reviewers who are interacting with chatbots to adapt their reviews positively, such that they are perceived favorably and circumvent negative judgment from the review solicitor. Further, we anticipate that a higher level of anthropomorphism in the review solicitation medium (operationalized by the deployment of chatbots) positively affects reviewers' evaluations via increased social presence. Therefore:

**H1** A higher level of anthropomorphism in the review solicitation medium leads to higher rating valence.

**H2** The positive effect of anthropomorphism in the review solicitation medium on rating valence is mediated by social presence.

### Interaction enjoyment

In general, a communication medium endowed with rich anthropomorphic features not only evokes a feeling of being in the presence of another person but also leads to perceptions of the conversation as fun and enjoyable (Blut et al., 2020; Van Doorn et al., 2017; Van Pinxteren et al., 2020). One can explain this effect by considering people's preference of interacting with real humans. Qiu and Benbasat (2009) showed that the humanlike appearance and voice output of virtual recommendation agents are associated with increased enjoyment. Jin (2010) revealed that the presence of an educational virtual agent in an interactive test is associated with higher student enjoyment. Therefore, we expected the review solicitation medium endowed with anthropomorphic cues to enhance the perceptions of enjoyment.

Further, the enjoyment of interacting with an anthropomorphic medium might have a positive effect on the review rating. Evaluative conditioning theory predicts that a positively laden stimulus presented with another stimulus positively affects the evaluation of the second stimulus (De Houwer et al., 2001; Hofmann et al., 2010). In this vein, a solicitation medium deliberately designed as an interactive, likable, and humanlike conversational partner might affect the resulting product review, a general phenomenon widely observed in the field of relationship marketing (Palmatier et al., 2006). For example, customers who experience enjoyable interactions with service employees are more likely to provide positive word-of-mouth reviews (Gremler & Gwinner, 2000). Howard and Gengler (2001) showed how positive emotions from a gift

sender spill over to the gift receiver and, subsequently, to the gift itself. In the context of product reviews, Woolley and Sharif (2021) showed that small financial incentives increase the joy of writing reviews and, subsequently, positively affect rating valence. Based on this evidence, we formulated hypothesis H3 as follows:

**H3** The positive effect of anthropomorphism in the review solicitation medium on rating valence is mediated via perceived interaction enjoyment.

Naturally, the question of how to compare the two mechanisms of social presence and interaction enjoyment emerges. Under which circumstances will one mechanism dominate the other and vice versa? In this direction, prior literature suggests that the interactivity of a medium is a strong predictor of perceived interaction enjoyment (Coursaris & Sung, 2012). This relationship has been found in a variety of domains, such as website use (Coursaris & Sung, 2012), virtual reality use (Jang & Park, 2019), and digital artistic experiences (Gonzales et al., 2019). Interactivity is part of anthropomorphism (Kim & Sundar, 2012), but its increase may not be linear over our three conditions. The investigations that follow are designed to explore three levels of anthropomorphism (low vs. moderate vs. high) conditions. The increase in interactivity is arguably strongest comparing a low anthropomorphic medium (a conventional review form) to a moderately anthropomorphic medium (a chatbot). The chatbot is much more interactive than the form because of its ability to mimic a conversation. By contrast, both the moderately anthropomorphic and the highly anthropomorphic chatbots are interactive. Social presence, on the other hand, should increase continuously over three anthropomorphism conditions because each condition adds features that make the medium more human-like. Hence, we hypothesize that the interaction enjoyment mechanism is stronger than the social presence mechanism when comparing low to moderately anthropomorphic conditions. However, the social presence mechanism will be stronger than the interaction enjoyment mechanism when comparing moderate to highly anthropomorphic conditions. Consequently, H4 and H5 are formalized as follows:

**H4** The mediation through interaction enjoyment is stronger than the mediation through social presence when comparing low to moderately anthropomorphic review solicitation media

**H5** The mediation through social presence is stronger than the mediation through interaction enjoyment when

comparing moderate to highly anthropomorphic review solicitation media.

## Review helpfulness

Finally, anthropomorphism may affect the perception of reviews, particularly their helpfulness. A large body of literature discusses the determining characteristics of helpfulness for online reviews (Hong et al., 2017). Among these characteristics are the review's age (Archak et al., 2011), the rating valence (Kuan et al., 2015), the readability (Archak et al., 2011), and the review length (Mudambi & Schuff, 2010). Among these four, the most important precursor for helpfulness is the review length. Other things equal, past studies suggest that longer reviews are more helpful because they contain more information and have improved diagnosticity (Mudambi & Schuff, 2010). Social presence likely affects review length. Triggered by

the increased social presence of chatbots, the feeling of increased common ground can unfold. *Common ground* describes the knowledge, beliefs, and suppositions that communicators share and know that they share (Krauss and Fussell 1996). For example, we describe an object differently to a person who is in our physical or conversational proximity ("Take the red one") rather than one who is not ("Take the red ball when you enter the room"). Therefore, people need fewer words to communicate the same message in a condition that is more anthropomorphic (higher increased feeling of common ground) than in one with less anthropomorphism (decreased feeling of common ground). Therefore:

**H6** A higher level of anthropomorphism in the review solicitation medium leads to lower review helpfulness.

**H7** The negative effect of anthropomorphism in the review solicitation medium on review helpfulness is mediated via review length.

**Table 2** Summary of studies

Study	Conditions	Product	Findings
1	Low vs. high anthropomorphism	Teaching video evaluations	Product ratings are higher for review solicitation media with higher anthropomorphism.
2	3 (low vs. moderate vs. high anthropomorphism) × 2 (emotional vs. non-emotional)	Short movies	Product ratings are higher for review solicitation media with higher anthropomorphism. Interaction enjoyment (social presence) mediates the effect of anthropomorphism strongest for low vs. moderate (moderate vs. high) anthropomorphism. Emotionality does not affect the results.
3	2 (moderate vs. high anthropomorphism) × 2 (seller vs. platform)	Short movies	Product ratings are higher for review solicitation media with higher anthropomorphism. Social presence mediates the effect of anthropomorphism, but this mediation cannot be muted by a moderation with solicitor salience.
4	2 (low vs. moderate anthropomorphism)	AMT worker survey reviews	Reviews solicited via review solicitation media with higher anthropomorphism are less helpful. This effect is mediated by review length.

*Note* An AMT worker refers to a worker on Amazon Mechanical Turk

## Overview of studies

We conducted four studies (see Table 2 for an overview) to examine the effects of the level of anthropomorphism in review solicitation media. We conceptualized them along the continuum of anthropomorphism, with conventional forms and highly humanlike chatbots on the lower and higher ends, respectively. (See Table 3 for the overview of anthropomorphism manipulations.)

In Study 1, we tested the main effect of anthropomorphism in the review solicitation medium on rating valence in a field experiment. We found that chatbots, representing more anthropomorphic solicitation media, generate higher ratings than conventional forms (H1). In Study 2, using a controlled experimental setting, we confirmed the positive effect of anthropomorphism on rating valence (H1) and showed that this effect is mediated by interaction enjoyment (H3) for low versus moderate levels of anthropomorphism (H4) and by social presence (H2) for moderate versus high levels of anthropomorphism (H5). We tested the effects of expressed emotionality for two reasons. First, we wanted to rule out the confounding effects of anthropomorphism and expressed emotionality. Second, we tested expressed emotionality's role as a potentially actionable moderator to mute the effects of increased interaction enjoyment. In Study 3, we tested solicitor salience as a potential actionable moderator to mute the effects of increased social presence for moderate versus high levels of anthropomorphism. Finally, in Study 4, we examined the effect of solicitation medium



**Table 3** Overview of anthropomorphic cues of the solicitation media across studies

	Level of Anthropomorphism	Low	Moderate	High	
	Anthropomorphic Cue	Design Elements	Design Elements	Design Elements	Source
Visual	Humanlike appearance	No	Chatbot icon	Real photo of service agent	Feine et al. (2019)
	Gender	No indication	No indication	Female agent	Fogg (2002)
	Human name	No	No	Yes (Emma)	Araujo (2018)
Verbal	Emoticons	No use	Moderate use	Frequent use	Feine et al. (2019)
	Conversational skill	No	Yes, no first person (e.g., “Let’s get started.” / “Good to hear you liked the survey.” / “Seems like you were not very pleased.” / “Ok, good to know.”)	Yes, with the use of first person (e.g., “I am happy you liked the movie” / “We would like to hear more!”)	Feine et al. (2019) Schuetzler et al. (2018)
	Self-introduction	No	No	Yes. (“I’m Emma.”)	Feine et al. (2019)
	Opinion conformity	No	Yes, no first person (e.g., “Good feedback.”)	Yes, first person (e.g., “I totally agree with you.”)	Feine et al. (2019)
	Thanking	No	Yes (e.g., “Thank you very much for providing your valuable feedback!”)	Yes, with use of user’s name. (e.g., “Amazing, thank you so much for your input, [name].”)	Feine et al. (2019)
	Greeting and farewell	No	Yes (e.g., “Hello there!”)	Yes, with use of user’s name. (e.g., “Nice to meet you, [name]!”)	Feine et al. (2019)
	Lexical diversity	No	Yes	Yes	Feine et al. (2019)
Invisible	Response time	Immediate proceeding to the next page	Dots blinking before each message is sent	Dots blinking before each message is sent / “Is typing...” shown before message is sent	Feine et al. (2019)

anthropomorphism on review readers’ perceptions of review helpfulness (H6 and H7).

Our studies offer comprehensive evidence of how anthropomorphism influences product ratings and helpfulness and identify the potential behavioral mechanisms. The general aim of our research was not to test the effects of single anthropomorphic cues on social presence or perceived interaction enjoyment but how increases in these two factors, which are triggered by the anthropomorphism of the review solicitation medium, can lead to higher ratings. For the design of various chatbot configurations, we used a combination of cues to ensure a certain level of anthropomorphism. Our approach is similar to that of Schanke et al. (2021) and is justified by the findings of Seeger et al. (2021), who showed that single cues are insufficient in achieving satisfactory levels of anthropomorphism. Table 3 presents an overview of the anthropomorphic cues used to configure the chatbots in the three studies, and Appendix I provides screen shots of the experimental manipulations of each study.

In each study, the participants were paid according to the platform guidelines at the time of data collection, based on the expected length of the studies (Palan & Schitter, 2018). Prior literature has revealed that financial incentives can induce a positive bias in ratings (Burtch et al., 2018; Khern-am-nuai et al., 2018). However, in our study, all the participants received the

same payment, and the experiments differed only in the deployed review solicitation media. Hence, the issue of payment did not confound the identification of the effects of anthropomorphism.

### Study 1

In Study 1, we conducted a field experiment to test whether higher anthropomorphism in the review solicitation medium affects rating valence (H1). The context of the study was course evaluations that students provided. We conducted the study at a medium-sized German university in the midterm of an undergraduate course in statistics between November 22 and November 26, 2022. Because the course introduced learning videos for students, and to ensure consistency with later studies, we asked students to evaluate those videos.

### Experimental design

The instructor asked the students via the course platform to click on a link to rate the course videos. Once students clicked on the provided link, they were randomly allocated to a high (using a chatbot with a high level of anthropomorphism) or a low (using a conventional form) anthropomorphism condition (see Appendix I). We manipulated anthropomorphism via the use of a chatbot, which equaled 1 if the students gave their ratings through

a chatbot and equaled 0 otherwise. We followed prior literature (Feine et al., 2019) and used a combination of typical visual, verbal, and invisible anthropomorphic cues to ensure a high level of anthropomorphism (see Table 3 for design cues) of chatbots. These cues in the chatbot-mediated review solicitation simulated a human conversation. The implemented chatbot applied typical conversational norms such as greetings, farewells, and thank-yous. We implemented blinking dots to simulate time needed for turn taking. Students could not identify the condition before clicking on the link. The rating questions were identical across both groups. We conducted the experiment on days that students had no classes to avoid interference between conditions and prevent contamination of the results. We measured the rating of the videos used in the course on a 5-point scale to comply with the standards of the host university.

## Results

### Manipulation checks

Due to the field experiment nature of Study 1, we were not able to conduct a check of a successful manipulation. Instead, we used an independent Prolific sample ( $N=196$ ) to perform a post-hoc manipulation check of anthropomorphism of the review environment on two seven-point Likert scales (Crolic et al., 2022; Kim et al., 2016): “Please rate the extent to which the [review environment / chatbot] you used when providing the review: came alive (like a person) in your mind; has some humanlike qualities”). We found that participants evaluated the high anthropomorphic condition (the chatbot) as more anthropomorphic than the low anthropomorphic condition (the form) ( $M_{\text{low}} = 3.02$  vs.  $M_{\text{high}} = 4.97$ ;  $F(1, 195) = 47.37$ ,  $p = .00$ ,  $\alpha = 0.91$ ).

### Key findings

A total of 68 students (33 in the low anthropomorphism condition, and 35 in the high anthropomorphism condition) completed the video evaluation for the course. Our analysis confirmed that students in the chatbot condition provided significantly higher ratings than those in the conventional form condition ( $M_{\text{low}} = 3.79$  vs.  $M_{\text{high}} = 4.38$ ;  $F(1, 66) = 5.81$ ,  $p = .02$ ). Due to the field experiment nature of the study and to ensure the ecological validity, we did not collect further control variables. Also, due to data protection laws, we could not access administrative student data from the host university.

## Discussion

By leveraging evidence from a real-life field study, we found support for the hypothesis that anthropomorphism of the review solicitation medium increases the ratings valence. The differences in means were non-negligible (16% increase in ratings). Because student evaluations can have important implications for faculty promotion or course redesign, biases arising from anthropomorphic solicitation media may have detrimental effects in this context.

### Study 2

Whereas Study 1 delivered field evidence for the main effect, Study 2 tested this effect in a controlled experimental setting. In particular, the study served three purposes. First, we obtained a deeper understanding of the effect of different levels of anthropomorphism by extending the previous chatbot design, adding a moderately anthropomorphic chatbot. Second, based on the different levels of anthropomorphism, we elucidated the theoretical mechanism behind the main effect ( $H2 / H3$ ) and examined the prevailing mechanism at different levels of anthropomorphism. Third, we ruled out a potential *chameleon effect*—a facial, verbal, or behavioral mimicry (LaFrance, 1979; Chartrand & Bargh, 1999)—in the rating behavior triggered by expressed emotionality of the chatbots. Research from various fields shows that people “catch” the emotional states of those around them (Bagozzi et al., 1999; Howard & Gengler, 2001; Neumann & Strack, 2000). Goldberg and Gorn (1987) demonstrated that happy TV shows evoke positive audience emotions. Similarly, Howard and Gengler (2001) found that receivers evaluate gifts from happy senders more favorably. The roots of emotional contagion lie in the biological mechanism of a mirror neuron system (Iacoboni, 2009; Rizzolatti & Craighero, 2004) that lets people mimic others, a feature likely essential for survival of the human species (Lakin et al., 2003). Although emotions are part of what makes humans human (Haslam, 2006), we intended to mitigate participants simply mimicking positive emotions engendered in a chatbot by giving high ratings. In that case, we would expect a positive moderated mediation of emotionality on the path from anthropomorphism to interaction enjoyment and social presence.

### Experimental design

In Study 2, we implemented a 3 (anthropomorphism: low vs. moderate vs. high)  $\times$  2 (emotionality: non-emotional vs. emotional) experimental design. We divided

the experiment into two phases: product experience and review generation. In the product experience phase, we used a short clay animation movie as the focal item to be rated (Schneider et al., 2021). We selected and evaluated this video during a pre-study, as shown in Appendix E. After completing the product experience phase of watching the movie, participants were randomly allocated to one of the experimental conditions. We disabled the display of progress bars in each of the conditions.

To calculate the sample size required to detect a significant difference between the overall ratings in the conditions, we used G\*Power (Faul et al., 2009). Based on an effect size of  $d = 0.4$ ,  $\alpha = 0.05$ , and  $\beta = 0.8$ , we determined that, to observe a significant difference, we needed a sample size of 100 participants per condition (600 in total). We used Prolific to recruit US participants and gave them monetary compensation (£0.16 per minute) upon completion of both phases. We prescreened participants based on approval rate and previous submission to ensure a suitable pool for our study. As an attention check, we displayed a number at the end of the video to confirm that participants had watched it to completion. Participants who could answer the number correctly at the end of the video passed the attention check.

Next, we designed three levels of anthropomorphism (low vs. moderate vs. high). The low anthropomorphism condition was a conventional review form. The moderate and high anthropomorphism conditions were chatbots with different levels of anthropomorphic cues. In designing the chatbot, we followed prior literature (Feine et al., 2019) and used a combination of typical visual, verbal, and invisible anthropomorphic cues to ensure various levels of anthropomorphism (Table 3). These cues in the chatbot-mediated review solicitation simulated a human conversation.<sup>1</sup> The three conditions differed only in their format—either a form or a chatbot—and both review solicitation media incorporated the same questions and flow. To ensure the manipulation was successful, we asked participants to indicate the anthropomorphism of the review environment on two seven-point Likert scales (Crolic et al., 2022; Kim et al., 2016): “Please rate the extent to which the [review environment / chatbot] you used when providing the short movie review: came alive (like a person) in your mind; has some humanlike qualities.”

<sup>1</sup> Disclosing the fact that consumers are interacting with a chatbot and not with a human collocutor can bias human behavior (Luo et al., 2019). However, we conducted a pretest and found that such disclosure did not significantly affect respondents’ rating responses ( $M_{\text{no\_disclosure}} = 5.38$  vs.  $M_{\text{disclosure}} = 5.23$ ;  $F(1, 557) = 0.46$ ,  $p = .50$ ). Nevertheless, in all our studies, we did not explicitly inform the respondents of the chatbot conditions and that they were interacting with robots rather than humans.

Following Han et al. (2022), we manipulated emotionality using adjectives, exclamation marks, emojis, and responsiveness. (See Appendix A for an overview of emotionality cues.) As the manipulation check, we asked participants to indicate the level of emotion of the review environment using the scale of three seven-point Likert scales (Puntoni et al., 2009): “In your opinion, how much emotion was expressed by the [review environment / chatbot] customer service agent when providing the short movie review: expressed a great deal of emotion; expressed a lot of feelings; expressed many sentiments.”

We measured rating valence as the average of three numerical ratings on a 1- (worst) to 10-star (best) scale for (a) visual appeal, (b) story line, and (c) overall quality (Cui et al., 2012). We selected a 10-star scale because larger scales lead respondents to utilize more response points, lower their mean ratings, and reduce their extreme response tendency (Tsekouras, 2017). We also asked participants to provide review texts. We included three mandatory sections in the review texts: positive aspects, negative aspects, and overall evaluation. Dividing review texts into several categories is a common industry approach that we followed to increase external validity. Importantly, the area to write the review was the same size (370px × 100px) in all conditions.

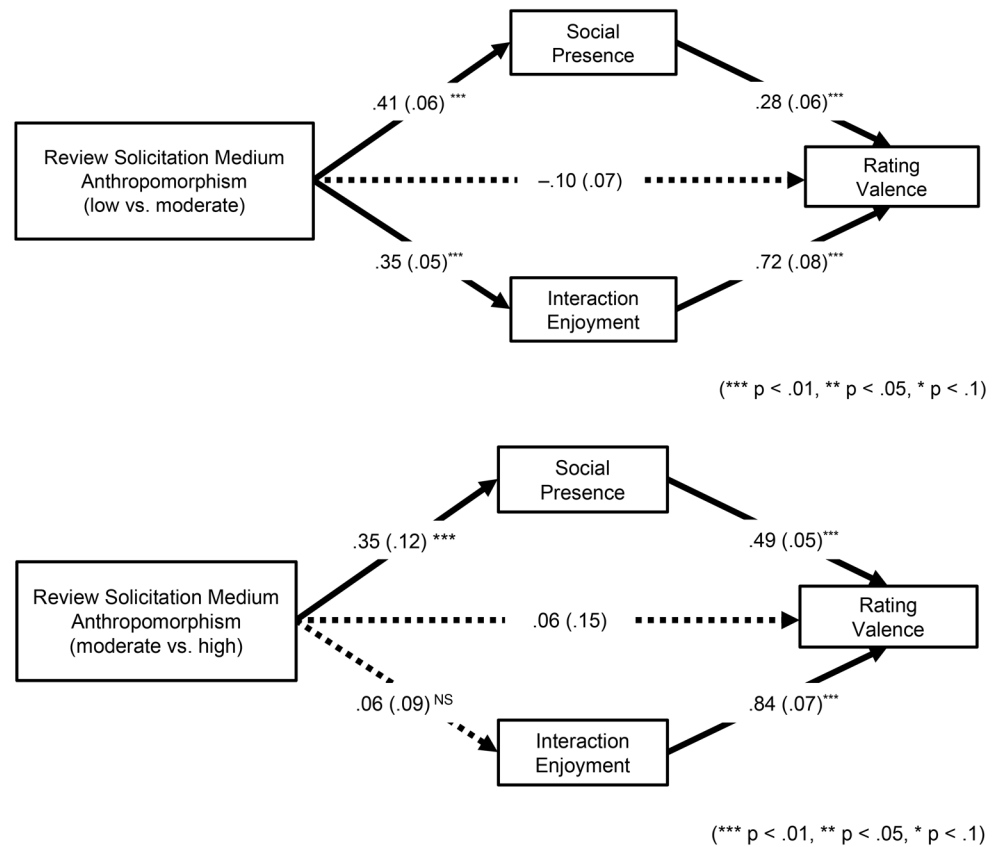
The mediators in H2 and H3 were interaction enjoyment and social presence. We measured these variables using the scales of Schuetzler et al. (2018), Gefen and Straub (2003), Mimoun and Poncin (2015), Qiu and Benbasat (2009), and Verhagen et al. (2014). An overview of all the survey questions is included in Appendix B.

## Results

We had 1,112 participants complete the study. A randomization check revealed that the participants across the conditions were not statistically different (Table C-1 in Appendix C). Table D-2 in Appendix D displays the summary statistics of the data collected in Study 2.

### Manipulation checks

The manipulations worked as intended. We found that perceived anthropomorphism significantly increased in line with greater anthropomorphism ( $M_{\text{low}} = 4.10$  vs.  $M_{\text{moderate}} = 4.19$  vs.  $M_{\text{high}} = 4.60$ ;  $F(2, 1109) = 10.66$ ,  $p = .00$ ;  $\alpha = 0.84$ ). Pairwise comparisons revealed that perceived anthropomorphism in the high condition was significantly higher than in the moderate ( $F(1, 733) = 14.30$ ;  $p < .001$ ) and low conditions ( $F(1, 742) = 18.37$ ;  $p = .00$ ), yet the difference between the low and the moderate conditions was

**Fig. 2** Study 2 mediation analyses

not significant ( $F(1, 746)=0.45$ ;  $p=.50$ ).<sup>2</sup> Also, the introduction of emotional cues significantly increased perceived emotionality ( $M_{\text{non-emotional}} = 3.46$  vs.  $M_{\text{emotional}} = 4.15$ ;  $F(1, 1110)=49.37$ ,  $p=.00$ ;  $\alpha=0.96$ ).

## Key findings

ANOVA results revealed a significant main effect of review medium anthropomorphism on rating valence, where participants provided higher ratings for the product the more anthropomorphic the review environment ( $M_{\text{low}} = 4.64$  vs.  $M_{\text{moderate}} = 5.03$  vs.  $M_{\text{high}} = 5.16$ ;  $F(2, 1,108)=5.03$ ,  $p=.00$ ).<sup>3</sup> We further found no direct effect of emotionality and no interaction effect between anthropomorphism and emotionality on rating valence. Further, we found a negative effect of anthropomorphism on review length (measure

as log number of words,  $M_{\text{low}} = 4.32$  vs.  $M_{\text{moderate}} = 4.16$  vs.  $M_{\text{high}} = 3.68$ ;  $F(2, 1108)=52.17$ ,  $p=.00$ ).<sup>4</sup>

## Mediation

We examined the mediating role of social presence (H2) and interaction enjoyment (H3). We performed a mediation analysis based on 5,000<sup>5</sup> bootstrapped samples (Hayes, 2012) to generate the confidence intervals around the indirect effects via the mediators. The results are shown in Fig. 2. We found a significant indirect effect of moderate anthropomorphism (compared to low anthropomorphism) on rating valence through interaction enjoyment ( $b=0.48$ ,  $SE=0.09$ , 95%  $LLCI=0.32$ , and 95%  $ULCI=0.65$ ) and social presence ( $b=0.13$ ,  $SE=0.04$ , 95%  $LLCI=0.06$ , and 95%  $ULCI=0.23$ ). The mechanism through interaction enjoyment was significantly stronger than that through social presence (difference=0.34, 95%  $LLCI=0.18$ , and 95%  $ULCI=0.54$ ), suggesting that interaction enjoyment is the prevailing mechanism when soliciting review via conventional forms compared to a chatbot, including a moderate

<sup>2</sup> We note that this is in line with Hypothesis H4, where we argue that for the comparison between low and moderate anthropomorphism, interaction enjoyment would be the dominant mechanism. As we will see, interaction enjoyment is significantly different between the low and the moderate anthropomorphism conditions, and the construct is less dependent on anthropomorphism compared to social presence.

<sup>3</sup> The pairwise comparisons across conditions show a significant difference in low vs. moderate ( $F(2, 745)=5.29$ ,  $p=.02$ ) and low vs. high ( $F(2, 742)=9.20$ ,  $p=.00$ ), and not in moderate vs. high ( $F(2, 732)=0.57$ ,  $p=.45$ ).

<sup>4</sup> The pairwise differences across conditions are significant in low vs. moderate ( $F(2, 745)=5.37$ ,  $p=.02$ ), low vs. high ( $F(2, 742)=107.87$ ,  $p=.00$ ), and moderate vs. high ( $F(2, 732)=51.09$ ,  $p=.00$ ).

<sup>5</sup> The results were robust to the number of bootstraps and when we introduced mediators independently.

number of human cues. This supports H4. The direct effect when including mediators is nonsignificant ( $b = -0.22$ ,  $SE = 0.14$ ,  $95\% \text{ LLCI} = -0.50$ , and  $95\% \text{ ULCI} = 0.06$ ), suggesting a full mediation via the two mediators. We further found a significant indirect effect of high (compared to moderate) anthropomorphism on rating valence through social presence ( $b = 0.17$ ,  $SE = 0.06$ ,  $95\% \text{ LLCI} = 0.05$ , and  $95\% \text{ ULCI} = 0.30$ ). We found no significant indirect effect through interaction enjoyment ( $b = 0.05$ ,  $SE = 0.08$ ,  $95\% \text{ LLCI} = -0.10$ , and  $95\% \text{ ULCI} = 0.20$ ), demonstrating that social presence is the prevailing mechanism when comparing moderate to high anthropomorphism. This supports H5. The direct effect when including the mediators is nonsignificant ( $b = -0.04$ ,  $SE = 0.16$ ,  $95\% \text{ LLCI} = -0.37$ , and  $95\% \text{ ULCI} = 0.28$ ), suggesting a full mediation via the two mediators.<sup>6</sup> Further, we performed a moderated mediation analysis to examine the extent to which the emotionality of the review solicitation medium increases the perceived enjoyment and social presence and may explain variation on the rating valence (Hayes, 2012, Model 7<sup>7</sup>). We found that the index of moderated mediation was nonsignificant ( $b = 0.23$ ,  $SE = 0.14$ ,  $95\% \text{ LLCI} = -0.04$ , and  $95\% \text{ ULCI} = 0.51$ ), and the indirect effects via social presence and interaction enjoyment remained consistent across the levels of review environment emotionality.

## Discussion

Whereas Study 1 provided initial support for the anthropomorphism effect on rating valence, Study 2 tested our theorizing in a controlled experimental environment. This allowed us to achieve four things. First, it demonstrated the internal validity and increased the credibility of our results. Second, it showed that the anthropomorphism effect operates through two mechanisms with varied importance across different levels of anthropomorphism. The interaction enjoyment mechanism prevails in the low-to-moderate anthropomorphism scenario, and the social presence mechanism dominates in the moderate-to-high anthropomorphism scenario. Third, it allowed us to rule out the alternative explanation that our results are driven by a chameleon effect through emotional language embedded in the chatbots. We showed that stripping chatbots of emotional communication cues does not mitigate the positive anthropomorphism effect. Finally, we showed that anthropomorphism decreases review length.

<sup>6</sup> We also conducted a mediation analysis comparing low vs. high levels of anthropomorphism. Both mechanisms were statistically significant, and the interaction enjoyment mechanism was significantly stronger than the social presence mechanism.

<sup>7</sup> The results are robust to models 8, 15, 58, and 59.

## Study 3

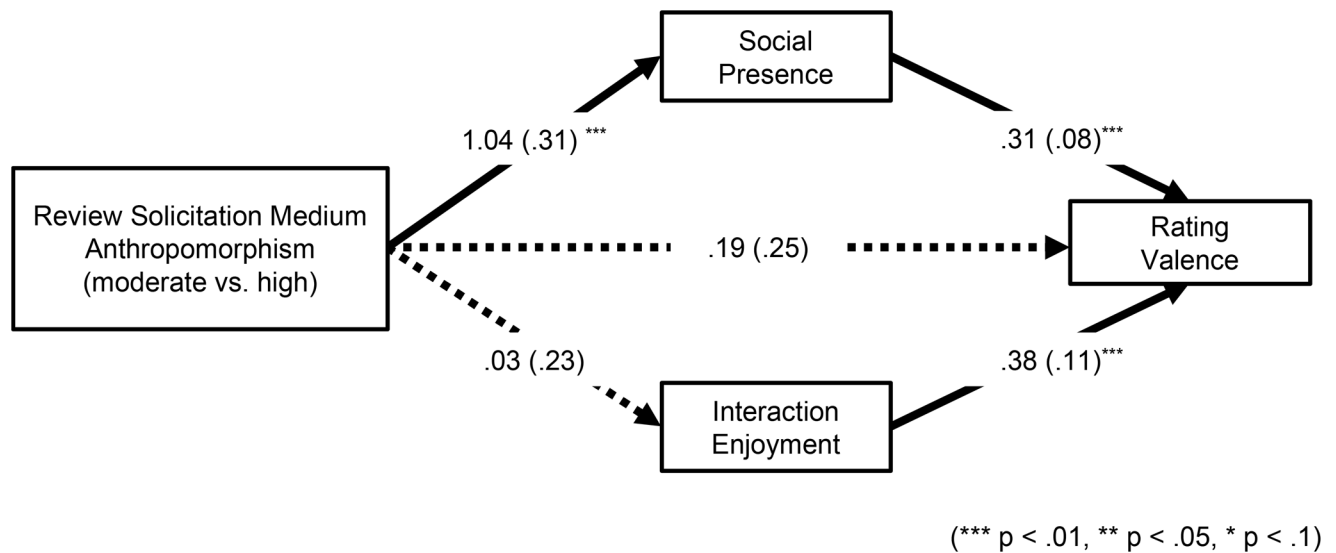
In Study 2, we demonstrated the existence of the interaction enjoyment and social presence channels. Further, we showed that the interaction enjoyment channel cannot be muted via non-emotional chatbot design. The focus of Study 3 was to examine a plausible moderator that could mute or decrease the social presence mechanism by manipulating the salience of the solicitor who deploys the chatbot (high (seller) vs. low (third-party platform)). This is based on the rationale that seller solicitors are more salient to reviewers, whereas platforms, as a group without a personal identity, are less salient (Neumann & Gutt, 2019). As per Short et al. (1976), the salience of the conversation partner is an important element to social presence. A salient solicitor may strengthen the relationship between anthropomorphism and social presence, whereas a less salient solicitor may weaken it. On this basis, we tested whether less solicitor salience can mute the social presence mechanism.

## Experimental design

We compared moderate to high anthropomorphism manipulation. We focused on these two conditions because social presence was the main behavioral mechanism explaining the effect of high compared to moderate anthropomorphism in Study 2.<sup>8</sup> We used a 2 (anthropomorphism: moderate vs. high)  $\times$  2 (solicitor salience: high (seller) vs. low (platform)) experimental design. The moderate and high anthropomorphism conditions were chatbots with different levels of anthropomorphic cues, following Feine et al. (2019) (see Table 3). We used the short movie from Study 2 and divided the study into two phases: product experience and review generation. In the product review phase, the participants were randomly allocated to a review solicitation medium with varied levels of anthropomorphism. We used G\*Power (Faul et al., 2009) and the parameters of Study 2 to calculate the sample size, recruiting a total of 200 participants who would be paid £1.34 (£0.13 per minute) on Prolific upon completion of both phases. We imposed inclusion criteria similar to those in Study 2 to ensure a high level of response quality and recruited participants from the UK. We used two attention checks. The first was a number displayed at the end of each video to confirm whether the participants had watched it to completion, and the second was a question posed to the participants regarding who had asked them to review the video.

<sup>8</sup> We replicated the same design in a low vs. moderate anthropomorphism experiment because social presence had a mediating role (though weaker than interaction enjoyment) in Study 2. The additional study (3b) is presented in Appendix F.





**Fig. 3** Study 3 mediation analyses

To measure rating valence, we used the same rating scales as in Study 2. Next, we used similar anthropomorphic cues for the chatbots as in Study 2 (see Table 3 for an overview). The manipulation of anthropomorphism worked as intended, as the highly anthropomorphic chatbot scored significantly higher on anthropomorphism ( $M_{\text{moderate}} = 5.40$  vs.  $M_{\text{high}} = 6.67$ ;  $F(1, 159) = 31.40$ ,  $p = .00$ ). Regarding the review solicitor's salience, we used two variations: a seller (short filmmaker "ClayProduction") for high salience or a third-party platform (movie comparison platform "Short-MovieCheck") for low salience. To reinforce this variation, we designed logos for all the parties and presented them and a short bio to the participants before the start of the videos. We maintained a neutral tone in the solicitor biographies to avoid affecting participant ratings. (The solicitors asked for "the stars you think it [the movie] deserves.") We did, however, highlight that the filmmaker had created the movie himself to emphasize that it was the first party and that ShortMovieCheck was an independent third-party platform. An illustrative overview of the review solicitors is presented in Appendix E.

To ensure that participants noticed the solicitor source, after they completed the review, we questioned them about who asked them for the review; 0.6% answered wrong. Further, we asked participants about the extent they tailored their review to address the solicitor and found a significant difference for seller solicitation (compared to platform;  $M_{\text{platform}} = 4.26$  vs.  $M_{\text{seller}} = 4.96$ ;  $F(1, 159) = 4.53$ ,  $p = .03$ ). Finally, we measured participants' interaction enjoyment and social presence as well as the same control variables as in Study 2. An overview of the survey questions is included in Appendix B.

## Results

We excluded the participants who had given no consent, provided incomplete responses, failed the attention checks, or had previously seen the videos. The final dataset used for the analyses consisted of 161 participants. A randomization check revealed that the participants across the conditions were not statistically different (Table C-2 in Appendix C). Table D-3 in Appendix D displays the summary statistics.

## Manipulation checks

The manipulations worked as intended. We found a significantly higher perceived anthropomorphism in the high anthropomorphism condition ( $M_{\text{moderate}} = 5.40$  vs.  $M_{\text{high}} = 6.67$ ;  $F(1, 159) = 31.40$ ,  $p = .00$ ;  $\alpha = 0.91$ ).

## Key findings

We confirmed the positive effect of high anthropomorphism on rating valence (H1). ANOVA results revealed that participants provided higher ratings the more anthropomorphic the review environment ( $M_{\text{moderate}} = 4.81$  vs.  $M_{\text{high}} = 5.33$ ;  $F(1, 159) = 3.28$ ,  $p = .07$ ).

## Mediation

Next, in a mediation analysis based on 5,000 bootstrapped samples (Hayes, 2012) we found, in line with Study 2, a significant indirect effect of high anthropomorphism (compared to moderate) on rating valence through social presence ( $b = 0.32$ ,  $SE = 0.12$ , 95%  $LLCI = 0.12$ , and 95%  $ULCI = 0.59$ ). The indirect effect through interaction

enjoyment was nonsignificant, suggesting that social presence is responsible for the positive anthropomorphism effect when comparing moderate to high levels. The direct effect when including mediators also was nonsignificant ( $b = 0.19$ ,  $SE = 0.25$ ,  $95\% LLCI = -0.31$ , and  $95\% ULCI = 0.68$ ), suggesting a full mediation via social presence. Further, we performed a moderated mediation analysis to examine the extent that the solicitor salience can increase social presence and possibly explain the variation in the rating valence (Hayes, 2012, Model 7).<sup>9</sup> We found that the index of moderated mediation was nonsignificant ( $b = -0.29$ ,  $SE = 0.29$ ,  $95\% LLCI = -0.78$ , and  $95\% ULCI = 0.18$ ), and the indirect effect of anthropomorphism via social presence was consistent for the solicitor conditions. The results are shown in Fig. 3.

## Discussion

In Study 3, we confirmed the effect of anthropomorphism on rating valence, noting that the direct effect reaches statistical significance only at the 10% level. Nevertheless, following McShane et al. (2024), who critically discussed the dichotomous treatment of the arbitrary 5% level for statistical significance, we consider this result a contribution to the cumulative evidence compiled in this paper where the direct effect is replicated multiple times. Beyond the direct effect, the mediation results through social presence were statistically significant at the 5% level. We further showed that the salience of the solicitor (seller vs. platform) does not moderate the mediation between anthropomorphism, social presence, and rating valence. This result suggests that the effect of anthropomorphism in the review collection process is not influenced by the solicitor of the review request and can be generalized across seller and third-party platforms. Therefore, decreasing solicitor salience is not a viable way to mitigate the rating bias induced through social presence.

## Study 4

In Study 4, we devoted attention to the effect of anthropomorphism on the helpfulness of reviews and tested H6 and H7. For firms, the goal of collecting reviews is to persuade potential customers to buy their products or use their services. A positive but unhelpful review is, however, of limited value to a firm (Forman et al., 2008). Consequently, it is essential for firms to collect helpful reviews to support customer decision-making. The perceived helpfulness of reviews is also one of the primary metrics to determine

which reviews should be presented to customers (Mudambi & Schuff, 2010; Yin et al., 2014), and it is typically influenced by rating valence and review length. Longer reviews are helpful because of their diagnosticity (Kuan et al., 2015; Mudambi & Schuff, 2010). For our research context, we drew on the theory of common ground to hypothesize that anthropomorphism would decrease the length of reviews. Other things equal, we conjectured that shorter reviews would be associated with decreased review helpfulness.

## Experimental design

We conducted a study on Amazon Mechanical Turk (AMT), a crowdsourcing platform allowing anonymous workers to complete web-based tasks for a predetermined payment (Buhrmester et al., 2016). We followed the same structure as in Studies 2 and 3: a product experience and a review generation phase. We implemented two different anthropomorphism conditions (low vs. moderate), as shown in Appendix G. For the review generation phase, we followed prior literature on online reviews and selected a survey that AMT workers completed as the experimental stimulus, or the product to be reviewed later (Burtch et al., 2018). The survey was titled “Demographics and Worker Activity on AMT.” As their task, the AMT worker had to fill out the survey and then review it. In the review generation phase, the participants provided their ratings on the survey—namely, on the survey design, the questions asked, and the response format alongside a written review.<sup>10</sup> These reviews served as the foundation for the next step: when a new sample of AMT workers would be asked to assess their helpfulness. Bearing this in mind, the choice of an AMT worker survey was deliberate because we wanted to collect ratings and reviews that would be relevant to AMT workers.

We presented these reviews to a new group of 300 US AMT workers whom we had recruited to evaluate them for a compensation of \$0.50 (\$0.13 per minute). We informed them of our interest in determining whether adding reviews from a specific AMT task completed by other workers would be a promising feature to support AMT worker in selecting tasks on the platform. We explained that the reviews originated from a previous task regarding “Demographics and Worker Activity on AMT,” where we asked AMT workers to review it after completion. We randomly allocated five reviews to each participant and asked them to evaluate review quality in terms of the reviews’ helpfulness. We displayed each review and respective questions on

<sup>9</sup> The results were robust (a) to the number of bootstraps and when we introduced mediators independently and (b) to models 8, 15, 58, and 59.

<sup>10</sup> Consistent with the previous studies, we found that the ratings in the low anthropomorphism condition were lower than in the moderate anthropomorphism condition. This effect was particularly pronounced for a low-quality product that had been manipulated (see Appendix G.).

separate pages. Multiple participants independently evaluated the reviews, allowing us to control for respondent fixed effects and eliminate any reader idiosyncrasies that otherwise would have biased our results.

We took the measures of anthropomorphism in the original review solicitation, as well as the rating valence and the review length (measured as the log-transformed number of words), from the first step of this study. We measured the reviews' helpfulness using three items established in prior literature ("The review is helpful / useful / informative,"  $\alpha=0.93$ ; Yin et al., 2014).

## Results

A total of 303 workers completed the study. Each participant evaluated multiple reviews, generated via low or moderate anthropomorphism, constructing a total panel dataset of 1,515 review impressions. Due to the panel structure of the data, we chose a worker-level fixed effects model to analyze the effect of anthropomorphism on review helpfulness. We present the results of separate regressions in Table 4. In Column 1, we found that reviews generated via moderately anthropomorphic media were 42% ( $(e^{-0.54} - 1) * 100 = -41.7$ ) shorter than reviews generated via media with low anthropomorphism. Next, in Column 2, we could see those reviews from the moderate anthropomorphism condition were less helpful than those from the low anthropomorphism condition. Finally, in Column 3 we regressed review helpfulness on review length and found a

significant positive relationship. These results provide initial evidence in support of H6 and H7.

The results were robust when we applied a random effects model specification, when we included the interaction between the solicitation medium's anthropomorphism and product quality, and when we controlled for the order of the displayed review and the helpfulness measure of the previous review. (To account for learning effects, see Appendix H.)<sup>11</sup>

## Mediation

In Table 4, we found that reviews solicited via more anthropomorphic media were shorter and perceived as less helpful than those generated via less anthropomorphic means. Next, we employed mediation analyses to investigate whether the shorter length of chatbot reviews (as we found previously) was responsible for their decreased helpfulness. We implemented mediation analyses using a Generalized Structural Equation (GSEM) model to accommodate for the panel nature of the data (Palmer & Sterne, 2015).<sup>12</sup> The results are shown in Fig. 4. We found a significant negative effect of moderate anthropomorphism on review length and a positive effect of review length on helpfulness. The negative direct effect of anthropomorphism on helpfulness was non-significant, suggesting a mediation effect of review length. In summary, we found a negative and significant indirect effect of moderate anthropomorphism (compared to low) on review helpfulness through review length ( $b = -0.28$ ,  $SE=0.08$ , 95%  $LLCI = -0.51$ , and 95%  $ULCI = -0.21$ ). This supports H7.

**Table 4** The effect of anthropomorphism on review helpfulness

Dependent Variable (DV)	DV		
	(1)	(2)	(3)
	(ln) Review Length (in words)	Review Helpfulness	
Anthropomorphism (Moderate)	-0.54*** (0.05)	-0.16* (0.07)	0.03 (0.07)
Product Quality (low)	0.09 (0.05)	-0.11 (0.07)	-0.14** (0.10)
(ln) Review Length (in words)			0.36*** (0.05)
Constant	3.09*** (0.04)	5.20** (0.06)	5.23** (0.07)
Participant FE	Yes	Yes	Yes
No. of Impressions	1,515	1,515	1,515
No. of Participants	303	303	303
R <sup>2</sup>	0.10	0.01	0.03

Note (a) Robust standard errors are in parentheses. \* $p < .05$ ; \*\* $p < .01$ . (b) A Breusch–Pagan test rejects the homoscedasticity of the error terms. (c) Among the 1,515 impressions, 1,193 were impressions of reviews, and the remainder were those of additional feedback. Our results are robust when removing the additional feedback data points

## Discussion

In Study 4, we complemented the characterization of the anthropomorphism effect on reviews by examining review helpfulness. Review helpfulness is of vital importance to firms and review platforms because it supports customer decision-making (Hong et al., 2017; Forman et al., 2008; Mudambi & Schuff, 2010). We showed that review readers considered reviews collected via chatbots less helpful. Our results suggest that the decrease in helpfulness is driven by chatbot-generated reviews being shorter. In summary, using chatbots to collect reviews may backfire for marketers because they decrease the resulting review helpfulness.

<sup>11</sup> We also collected evaluations of review credibility, persuasiveness, reviewer trustworthiness, and their intentions regarding choosing this task based on the reviews. The results were qualitatively consistent when we used them as dependent variables (see Appendix H).

<sup>12</sup> We confirmed these results when using 5,000 bootstrapped samples (Hayes, 2012), including user fixed effects.

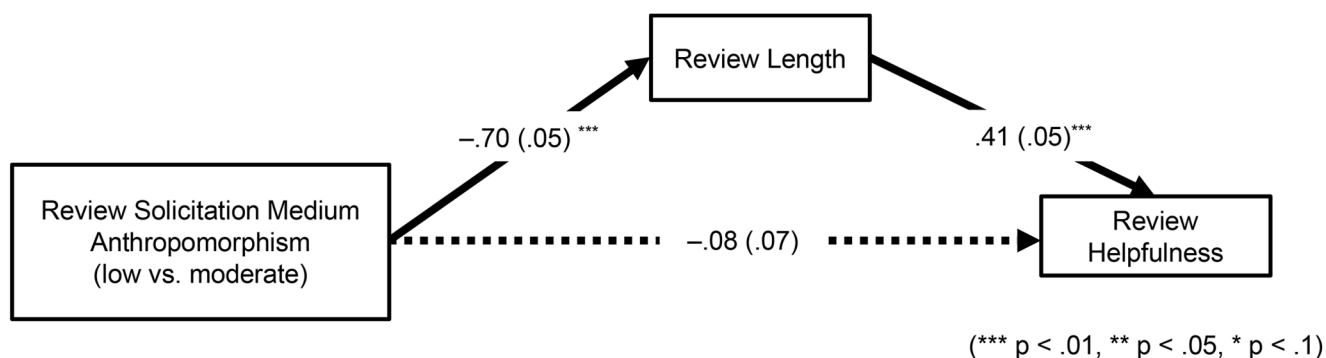


Fig. 4 Study 4 mediation analyses

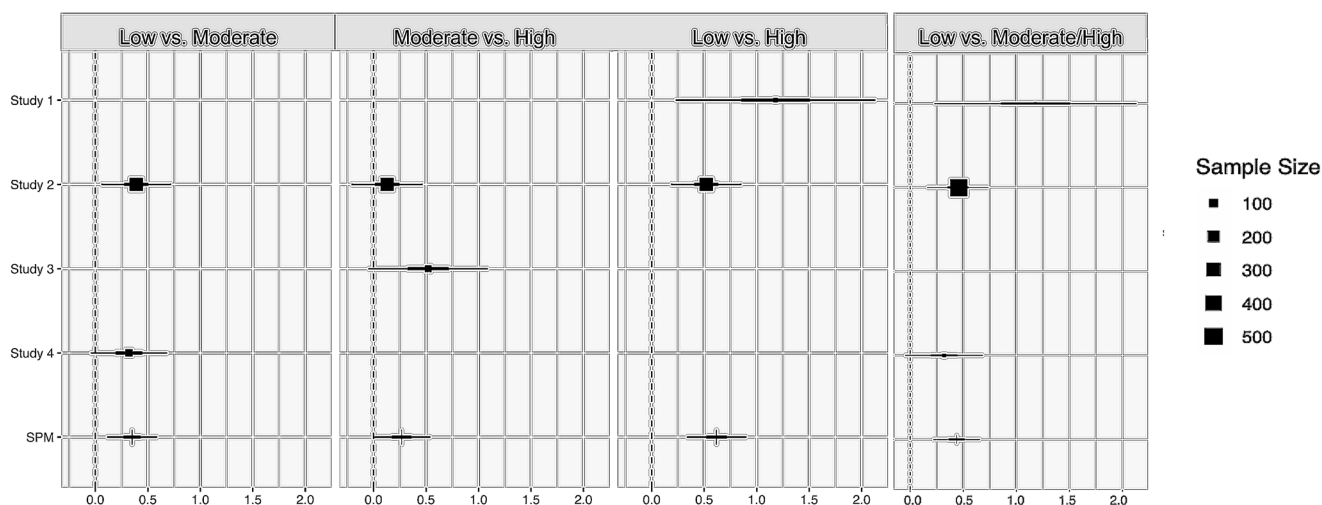


Fig. 5 Single paper meta-analysis graphical summary

### Single-paper meta-analysis

To test the overall validity of the effect of review solicitation medium anthropomorphism on rating valence, we conducted a single paper meta-analysis (SPM; McShane & Böckenholt, 2017). We standardized the dependent variables to account for variation in the rating scales. The SPM showed that, over four studies, the rating valence increases as the medium becomes more anthropomorphic. More precisely, we show that, compared to low anthropomorphism, moderate (estimate=0.35,  $s=0.13$ ,  $z=2.87$ ,  $p=.00$ ) and high (estimate=0.62,  $s=0.15$ ,  $z=4.26$ ,  $p=.00$ ) anthropomorphism in the review solicitation medium significantly increases the rating valence. High anthropomorphism also increases rating valence compared to moderate anthropomorphism (estimate=0.27,  $s=0.14$ ,  $z=1.93$ ,  $p=.05$ )<sup>13</sup>. Figure 5 presents the graphical summary of the contrast estimates across the conditions.

<sup>13</sup> The SPM when contrasting low versus moderate or high anthropomorphic conditions (i.e., no chatbot vs. chatbot) shows a significant increase in rating valence (estimate=0.44,  $s=0.11$ ,  $z=3.91$ ,  $p=.00$ ).

### General discussion

Chatbots have garnered notable industry interest as a new AI-enabled tool at the company-customer interface. They promise a low-cost, highly automated solution tailored to business and customer needs and represent an appealing new application for businesses, as evidenced by their rising use (Nirale, 2018).

Notwithstanding the promising potential of robotic automation, we should not overlook their potential side effects. Gathering and soliciting positive and informative online reviews is one of the most significant challenges that businesses face (Gutt et al., 2019). With the rise of chatbots, the environment in which reviews are solicited from customers awaits fundamental changes toward a more anthropomorphic, conversational direction. In this study, we sought to assess whether such changes affect the rating valence and review helpfulness. Not only did we isolate the effect of anthropomorphism on rating valence, but we uncovered two theoretical mechanisms governing the interaction between humans and (chat) robots.

Using four studies, we found that the anthropomorphism in the review solicitation medium increases rating valence. Our results suggest that, at a moderate level of anthropomorphism, the rating valence increase is driven by customers' interaction enjoyment. At high levels of anthropomorphism, however, the increased social presence is the prevalent driver (next to enjoyment) of the increase in rating valence. In both cases, the effects prove difficult to mute. Reducing the emotionality of chatbots cannot mute the interaction enjoyment mechanism, and decreasing the solicitor salience cannot mute the social presence mechanism. Finally, our results suggest that the anthropomorphism bias also affects review helpfulness. Increasing the solicitation medium anthropomorphism from low to medium decreases the helpfulness of the reviews. This effect is fully mediated by review length, which decreases when using anthropomorphic solicitation media.

The rise of chatbots warrants prudent elaboration of their pros and cons. Our paper is one step in that direction and primarily draws attention to the possible downsides of the anthropomorphism of such technologies. To the best of our knowledge, we are among the first to examine the consequences of anthropomorphic technology use for product perception by customers.

### Theoretical contributions

Four theoretical contributions emerge from our findings. First, to the best of our knowledge, we are the first to investigate how anthropomorphism affects consumers' online reviewing behavior. Thereby, we extend the scope of the literature on the effects of anthropomorphism on consumer behavior. Whereas previous literature investigated how anthropomorphism in chatbots affects consumers' perception of the firm (Crollic et al., 2022; Fotheringham & Wiles, 2022; Hildebrand & Bergner, 2021), we focused on how anthropomorphism affects the perception of a firm's products.

Second, although previous studies investigated the effects of anthropomorphism on consumer behavior (Crollic et al., 2022; Kim et al., 2016; Kwak et al., 2015; Schanke et al., 2021), whether the mechanisms responsible for the effects of anthropomorphism are consistent across the anthropomorphism continuum remained an open question. Our results demonstrate that this is not necessarily the case, documenting the distinct effects on consumer behavior at moderate (interaction enjoyment) and high (social presence) levels of anthropomorphism. Thereby, we offer a more fine-grained characterization of anthropomorphism effects and provide a conceptualization of two channels: interaction enjoyment and social presence, including a description of when we would expect one mechanism to dominate the

other, that future research can draw upon. Importantly, the mechanism of interaction enjoyment highlights a potential conflict of aims in a firm's use of anthropomorphic technology. Although it may seem trite, firms naturally try to make interactions with the firm pleasant for the consumers (Ran & Wan, 2023). That is especially true for new technology (Tonietto & Barasch, 2021) to encourage consumers to adopt and embrace it (Cai et al., 2022). In the context of online reviews, this strategy has a negative tradeoff because it biases rating valence and decreases review helpfulness.

Third, we contribute to the literature that has focused on online rating biases induced through the design of the online review environment (Gutt et al., 2019). Previous literature focused on the dimensionality of scales (Chen et al., 2018; Schneider et al., 2021) or online review templates (Poniatowski et al., 2019). These studies offer converging evidence that the static features provided in the interface for rating products can greatly affect consumer evaluations. We extend this stream of literature with our study on the effects of dynamic, AI-enabled, anthropomorphic review environment features. Thereby, we also respond to the call for research of AI-enabled technology on customer-firm interactions (Luo et al., 2019).

Finally, we contribute to the literature on unethical review collection (Luca & Zervas, 2016). Recent studies have focused on purchasing fake reviews (He et al., 2022) and the conditions that make firms more or less likely to fake reviews (Luca & Zervas, 2016; Mayzlin et al., 2014). We complement this literature by studying how deploying anthropomorphic chatbot technology positively biases online ratings and may hence constitute an unethical way to collect reviews.

### Practical implications

Our results present several practical implications to managers, consumers, and policymakers. To managers, our results imply that employing anthropomorphic technology to solicit reviews positively biases ratings. Managers who continue this practice knowing our results deliberately misrepresent the online ratings of their products raise ethical concerns because these ratings cannot be trusted. This contradicts the standard principles of ethical marketing that hold trust as an essential foundation (Gundlach & Murphy, 1993). In the long term, collecting reviews and ratings that misrepresent a product may cause serious damage to a brand's positioning and image (Bazaarvoice, 2022). Even if anthropomorphic technology becomes more widely known as an effective way of acquiring positive reviews, all sellers will want to use it, and no one will gain a competitive advantage with its



use, the buyers still stand to lose when all reviews become more positive and less discriminating.<sup>14</sup>

The incentive to misrepresent or disguise product quality might be particularly tempting for companies whose products will or have received low ratings.<sup>15</sup> Anthropomorphic technology would arguably be more effective for those companies at increasing product ratings than for companies whose products will obtain high ratings anyway. Nevertheless, our study also demonstrates some natural deterrents to such unethical practice. First, reviews obtained through chatbot solicitation are less helpful, which limits their value to the firm (Forman et al., 2008). Second, online reviews are an important source for product improvement and innovation. For example, hotels improve their quality based on online reviews (Ananthakrishnan et al., 2023), and app developers innovate their apps on that basis (Karanam et al., 2021). Ratings and reviews misrepresenting the underlying product threaten sustained improvement and innovation.

Consumers may be harmed in their decision-making when they face positively biased reviews. They may buy products based on positive ratings and end up disappointed with the product quality. This can extend beyond the context of business-to-consumer (B2C) commerce. As our field experiment shows, students choosing courses based on evaluations may also be misguided by positive evaluations using chatbots. This would be especially problematic if some product managers or course instructors used chatbots, but others refrained from their use. In this situation, the products or courses using chatbots would stand to improve their ratings over those that did not.

Taken together, our results call for the attention of policymakers to prevent managers from nefarious deployment of chatbots and consumers from harm. Recently, government bodies around the world (FTC, German Federal Cartel Office) have published guidelines on how to keep online reviews accurate and unbiased. As of today, they offer no advice on how to deal with anthropomorphized chatbots, but our study could be a force for positive change.

Our results also extend to platform owners. Akin to governments, they need to put platform governance mechanisms in place that enhance the welfare of their micro-economy (Foerderer et al., 2018). In particular, online review platforms like Yelp (Yelp, 2024) and e-commerce platforms like Amazon (Amazon, 2024) invest a lot of effort toward keeping their reviews unbiased and genuine. To ward off potential biases through chatbots, they would need to either ban chatbots or offer instructions on how to eliminate rating

biases via chatbot solicitation. The latter seems difficult based on our results.

## Limitations and future research

This research presents some limitations that provide fruitful avenues for future research. First, the generalizability of our findings is restricted by the boundaries of our research setting. These boundaries arise naturally from (1) certain limitations in sample representativeness emerging from the data collection platforms (Paolacci et al., 2010); (2) the product categories of the context of our study, including videos and the survey task; and (3) the design of solicitation media. Possible avenues to increase our study's generalizability involve the following: (a) including a more diverse and heterogeneous sample of participants with varying levels of experience and technology savviness; (b) applying different product categories for review solicitation, which may depend on the nature of the product; and (c) systematically varying numerous design aspects of the implemented solicitation media.

Second, consumers' behavioral response can change over time. Considering firms' high chatbot adoption and the popularization of ChatGPT, consumers might increasingly get used to chatbots. This can have repercussions for the mechanisms we identify, such as interaction enjoyment. Future research is needed.

Third, chatbot-mediated review length was significantly shorter only in Studies 2 and 4. We offer three possible explanations which can be further explored in future research. First, in line with recent evidence (Sachdeva et al., 2024), review length may depend on the structure of the review text solicitation (i.e., number of questions asked in the review process). Although we kept the review structure consistent within our studies, this finding suggests that even minor variations in the review environment may affect the length of the provided reviews. Further, Study 1 used German participants and Study 3 used a UK sample, whereas Studies 2 and 4 used US samples. This points to cultural differences being a potential reason for the inconsistencies. Indeed, there are indications for such effects in writing product review texts (Wang et al., 2019). Finally, the results may be attributed to small statistical power as the effect was insignificant in studies with the smallest sample. In summary, following McShane et al. (2024), we concluded that the cumulative evidence of the studies substantiates that anthropomorphism can decrease review length, and when it does it lowers review helpfulness. Future studies could further investigate the boundary conditions of the negative effect of anthropomorphism on review length.

Fourth, we conducted our study on a micro level to accurately elucidate the mechanisms and behavioral responses

<sup>14</sup> We thank an anonymous reviewer for drawing our attention toward the discussion of possible effects on demand and supply sides.

<sup>15</sup> To substantiate this, further analyses in Appendix G show that the anthropomorphism bias is particularly strong for low-quality products.

of humans to anthropomorphic chatbots. Although some studies (Fotheringham & Wiles, 2022) have focused on the market-level effects of chatbot adoption, additional future research is needed to investigate how the effects we identified play out on the macro level. Future studies might examine market outcomes and equilibria when all firms start implementing chatbots to solicit positive customer reviews and consumers become aware of the practices and their purposes. In this case, one can expect customer reviews to become less informative. Furthermore, consumers will adjust their reliance on reviews when making purchase decisions.

Fifth, we embedded our research into the CASA paradigm, showing that consumers respond to anthropomorphic technology in the same way they do to human-to-human interactions. For instance, when interacting with anthropomorphic technology, humans perceive social presence and care about the impression they leave. Despite its widespread application in research, the status of CASA as a theory, paradigm, or framework is subject to a current debate in the literature (Gambino et al., 2020). Because classifying the exact status of CASA is beyond the scope of our work, future research might develop a critical evaluation of this question.<sup>16</sup>

Sixth, even though our studies could not uncover ways of mitigating the biases induced through anthropomorphic chatbots, future studies should consider this. For instance, studies could impose minimum-length requirements or review templates (Poniatowski et al., 2019) to prevent a decrease in review length.

Finally, recent evidence has revealed that chatbot deployment can also have desirable consequences: it can increase the response rate and information completeness (Beam, 2023), increase the review volume (Buyerminds.com, 2021), and provide valid answers in medical contexts (Schick et al., 2022). Our studies focused on the biases on rating valence and length. However, analyzing whether positive effects of chatbots on customer feedback can countervail undesirable biases may prove useful.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11747-024-01027-8>.

**Acknowledgements** The authors thank Sandra Slotboom, Theresa Goeken, Theresa Kloos, Rowan Doesburg, and Sonja Lück for their research support. They benefited from discussions with Markus Weinmann and Rodrigo Belo and are grateful for feedback from research seminar participants at the University of Cologne, University of Münster, Weizenbaum Institute Berlin, WHU Otto Beisheim School of Management, Telecom ParisTec, Tel Aviv University, TU Munich, and University of St. Gallen as well as from conference participants at the

Conference on Information System and Technology (CIST) 2020, the Conference on Digital Experimentation @MIT (CODE@MIT) 2020, the Symposium on Statistical Challenges in E-Commerce Research (SCECR) 2020, and the European Conference on Information Systems (ECIS) 2020. Financial support from the Erasmus Research Institute for Management (ERIM) is gratefully acknowledged.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adam, M., Toutaoui, J., Pfeuffer, N., & Hinz, O. (2019). Investment decisions with robo-advisors: The role of anthropomorphism and personalized anchors in recommendations. In Proceedings of the European Conference on Information Systems, Stockholm, Sweden (ECIS 2019).
- Amazon (2024). Retrieved February 23, 2024 from amazon.<https://amazon.com/gp/help/customer/display.html?nodeId=G3UA5WC5S5UUKB5G>.
- Ananthakrishnan, U. M., Proserpio, D., & Sharma, S. (2023). I hear you: Does quality improve with customer voice? <https://ssrn.com/abstract=3467236>.
- Anderson, E. T., & Simester, D. I. (2014). Reviews without a purchase: Low ratings, loyal customers, and deception. *Journal of Marketing Research*, 51, 249–269.
- Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, 85, 183–189.
- Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, 57, 1485–1509.
- Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999). The role of emotions in marketing. *Journal of the Academy of Marketing Science*, 27(2), 184–206.
- Bazaarvoice (2022). Fake Reviews Report reveals consumers want brands and retailers fined for fake reviews. <https://www.bazaarvoice.com/press/fake-reviews-2022-report/>.
- Beam, E. A. (2023). Social media as a recruitment and data collection tool: Experimental evidence on the relative effectiveness of web surveys and chatbots. *Journal of Development Economics*, 162, 1–11.
- Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology*, 24, 586–607.

<sup>16</sup> We thank an anonymous reviewer for pointing out this issue and for fruitful discussions around the topic.

- Blut, M., Wang, C., Wunderlich, N. V., & Brock, C. (2020). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, 49, 1–27.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2016). Amazon's mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science*, 6, 3–5.
- Burtch, G., Hong, Y., Bapna, R., & Griskevicius, V. (2018). Stimulating online reviews by combining financial incentives and social norms. *Management Science*, 64, 2065–2082.
- Buyerminds.com (2021). Retrieved September 7, 2021 from <https://www.buyerminds.com/case-bol-com>.
- Cai, D., Li, H., & Law, R. (2022). Anthropomorphism and OTA chatbot adoption: A mixed methods study. *Journal of Travel & Tourism Marketing*, 39, 228–255.
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6), 893.
- Chen, P. Y., Hong, Y., & Liu, Y. (2018). The value of multidimensional rating systems: Evidence from a natural experiment and randomized experiments. *Management Science*, 64, 4629–4647.
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review Psychology*, 55, 591–621.
- Coursaris, C. K., & Sung, J. (2012). Antecedents and consequents of a mobile website's interactivity. *New Media & Society*, 14, 1128–1146.
- Crolic, C., Thomaz, F., Hadi, R., & Stephen, A. T. (2022). Blame the bot: Anthropomorphism and anger in customer-chatbot interactions. *Journal of Marketing*, 86, 132–148.
- Cui, G., Lui, H. K., & Guo, X. (2012). The effect of online consumer reviews on new product sales. *International Journal of Electronic Commerce*, 17, 39–58.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of Applied Social Psychology*, 22, 1111–1132.
- De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, 127, 853–869.
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114, 864–886.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G\* power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160.
- Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies*, 132, 138–161.
- Foerderer, J., Kude, T., Mithas, M., & Heinzl, A. (2018). Does platform owner's entry crowd out innovation? Evidence from Google photos. *Information Systems Research*, 29, 444–460.
- Fogg, B. J. (2002). Persuasive technology: Using computers to change what we think and do. *Ubiquity*, 2002, 89–120.
- Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, 19, 291–313.
- Fotheringham, D., & Wiles, M. A. (2022). The effect of implementing chatbot customer service on stock returns: An event study analysis. *Journal of the Academy of Marketing Science*, 50, 1–21.
- FTC (2024). Retrieved February 23, 2024 from [www.ftc.gov/business-guidance/resources/soliciting-paying-online-reviews-guide-marketers](http://www.ftc.gov/business-guidance/resources/soliciting-paying-online-reviews-guide-marketers).
- Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, 1, 71–85.
- Gartner (2018). Gartner Says 25% of customer service operations will use virtual customer assistants by 2020. Retrieved December 11, 2019 from <https://www.gartner.com/en/newsroom/press-releases/2018-02-19-gartner-says-25-percent-of-customer-service-operations-will-use-virtual-customer-assistants-by-2020>.
- Gefen, D., & Straub, D. (2003). Managing user trust in B2C e-services. *E-Service*, 2, 7–24.
- Gnewuch, U., Morana, S., Adam, M. T., & Maedche, A. (2022). Opposing effects of response time in human-chatbot interaction: The moderating role of prior experience. *Business & Information Systems Engineering*, 64, 773–791.
- Goldberg, M. E., & Gorn, G. J. (1987). Happy and sad TV programs: How they affect reactions to commercials. *Journal of Consumer Research*, 14, 387–403.
- Gonzales, A. L., Finley, T., & Duncan, S. P. (2019). (Perceived) interactivity: Does interactivity increase enjoyment and creative identity in artistic spaces? In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, April, 415–418.
- Gremler, D. D., & Gwinner, K. P. (2000). Customer-employee rapport in service relationships. *Journal of Service Research*, 3, 82–104.
- Gundlach, G. T., & Murphy, P. E. (1993). Ethical and legal foundations of relational marketing exchanges. *Journal of Marketing*, 57, 35–46.
- Gutt, D., Neumann, J., Zimmermann, S., Kundisch, D., & Chen, J. (2019). Design of review systems—A strategic instrument to shape online reviewing behavior and economic outcomes. *The Journal of Strategic Information Systems*, 28, 104–117.
- Han, E., Yin, D., & Zhang, H. (2022). Bots with feelings: Should AI agents express positive emotion in customer service? *Information Systems Research*, forthcoming.
- Haptik, T. (2018). Introducing feedback bot: Customer experience powered by conversational AI. Retrieved February 12, 2022 from [https://www.haptik.ai/blog/introducing-feedback-bot?utm\\_source=pocket\\_mylist](https://www.haptik.ai/blog/introducing-feedback-bot?utm_source=pocket_mylist).
- Haslam, N. (2006). Dehumanization: An integrative review. *Personality Social Psychology Review*, 10, 252–264.
- Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling. Retrieved March 30, 2023 from <http://www.afhayes.com/public/process2012.pdf>.
- He, S., Hollenbeck, B., & Proserpio, D. (2022). The market for fake reviews. *Marketing Science*, 41, 896–921.
- Health Europa (2019). New chatbot therapist to help combat depression unveiled. Retrieved November 27, 2019 from <https://www.health-europa.eu/chatbot-therapist-to-combat-depression-un/93609/>.
- Heath, C. (1996). Do people prefer to pass along good or bad news? Valence and relevance of news as predictors of transmission propensity. *Organizational Behavior and Human Decision Processes*, 68, 79–94.
- Hildebrand, C., & Bergner, A. (2021). Conversational robo advisors as surrogates of trust: Onboarding experience, firm perception, and consumer financial decision making. *Journal of the Academy of Marketing Science*, 49, 659–676.
- Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, 136, 390–421.
- Hong, H., Xu, D., Wang, A., G., & Fan, W. (2017). Understanding the determinants of online review helpfulness: A meta-analytic investigation. *Decision Support Systems*, 102, 1–11.
- Howard, D. J., & Gengler, C. (2001). Emotional contagion effects on product attitudes. *Journal of Consumer Research*, 28, 189–201.
- Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology*, 60, 653–670.



- Jang, Y., & Park, E. (2019). An adoption model for virtual reality games: The roles of presence and enjoyment. *Telematics and Informatics*, 42, 101–239.
- Jin, S. A. A. (2010). The effects of incorporating a virtual agent in a computer-aided test designed for stress management education: The mediating role of enjoyment. *Computers in Human Behavior*, 26, 443–451.
- Karanam, A., Agarwal, A., & Barua, A. (2021). ML-based product design: The case of mobile apps. In Proceedings of International Conference on Information Systems (ICIS 2021).
- Khern-am-nuai, W., Kannan, K., & Ghasemkhani, H. (2018). Extrinsic versus intrinsic rewards for contributing reviews in an online platform. *Information Systems Research*, 29, 871–892.
- Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior*, 28, 241–250.
- Kim, S., Chen, R. P., & Zhang, K. (2016). Anthropomorphized helpers undermine autonomy and enjoyment in computer games. *Journal of Consumer Research*, 43, 282–302.
- Kim, T., Lee, H., Kim, M. Y., Kim, S., & Duhachek, A. (2022). AI increases unethical consumer behavior due to reduced anticipatory guilt. *Journal of the Academy of Marketing Science*, 1–17.
- Krauss, R. M., & Fussell, S. R. (1996). Social psychological models of interpersonal communication. In E. T. Higgins, & A. Kruglanski (Eds.), *Social psychology: Handbook of Basic principles* (pp. 655–701). Guilford Press.
- Kuan, K. K., Hui, K. L., Prasarnphanich, P., & Lai, H. Y. (2015). What makes a review voted? An empirical investigation of review voting in online review systems. *Journal of the Association for Information Systems*, 16, 48–71.
- Kwak, H., Puzakova, M., & Rocereto, J. F. (2015). Better not smile at the price: The differential role of brand anthropomorphism on perceived price fairness. *Journal of Marketing*, 79, 56–76.
- LaFrance, M. (1979). Nonverbal synchrony and rapport: Analysis by the cross-lag panel technique. *Social Psychology Quarterly*, 42, 66–70.
- Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior*, 27, 145–162.
- Langer, E. J. (1992). Matters of mind: Mindfulness/mindlessness in perspective. *Consciousness and Cognition*, 1, 289–305.
- Lappas, T., Sabnis, G., & Valkanas, G. (2016). The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. *Information Systems Research*, 27, 940–961.
- Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, 107, 34–47.
- Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science*, 62, 3412–3427.
- 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, 937–947.
- Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104, 2421–2455.
- McShane, B. B., & Böckenholt, U. (2017). Single-paper meta-analysis: Benefits for study summary, theory testing, and replicability. *Journal of Consumer Research*, 43, 1048–1063.
- McShane, B. B., Bradlow, E. T., Lynch Jr, J. G., & Meyer, R. J. (2024). Statistical significance and statistical reporting: Moving Beyond Binary. *Journal of Marketing*.
- Miao, F., Kozlenkova, I. V., Wang, H., Xie, T., & Palmatier, R. W. (2022). An emerging theory of avatar marketing. *Journal of Marketing*, 86, 67–90.
- Mimoun, M. S. B., & Poncin, I. (2015). A valued agent: How ECAs affect website customers' satisfaction and behaviors. *Journal of Retailing and Consumer Services*, 26, 70–82.
- Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly*, 34, 185–200.
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56, 81–103.
- Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. Paper presented at the Human Factors in Computer Science, Boston.
- Nass, C., Fogg, B., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, 45, 669–678.
- Neumann, J., & Gutt, D. (2019). Money makes the reviewer go round—Ambivalent effects of online review elicitation in B2B markets. In *Proceedings of the Americas Conference on Information Systems, Cancun (AMCIS 2019)*.
- Neumann, R., & Strack, F. (2000). Mood contagion: The automatic transfer of mood between persons. *Journal of Personality and Social Psychology*, 79, 211–223.
- Nirale, S. (2018). What makes emerging technologies the future of customer experience? Retrieved March 30, 2023 from <https://servion.com/blog/what-emerging-technologies-future-customer-experience/>.
- Noble, S. M., Mende, M., Grewal, D., & Parasuraman, A. (2022). The Fifth Industrial Revolution: How harmonious human-machine collaboration is triggering a retail and service [r]evolution. *Journal of Retailing*, 98, 199–208.
- Novak, T. P., & Hoffman, D. L. (2019). Relationship journeys in the internet of things: A new framework for understanding interactions between consumers and smart objects. *Journal of the Academy of Marketing Science*, 47, 216–237.
- Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, 5, 1–35.
- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral Experimental Finance*, 17, 22–27.
- Palmatier, R. W., Dant, R. P., Grewal, D., & Evans, K. R. (2006). Factors influencing the effectiveness of relationship marketing: A meta-analysis. *Journal of Marketing*, 70, 136–153.
- Palmer, T. M., & Sterne, J. A. (2015). Fitting fixed-and random-effects meta-analysis models using structural equation modeling with the sem and gsem commands. *The Stata Journal*, 15, 645–671.
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5, 411–419.
- Poniatowski, M., Neumann, J., Goerzen, T., & Kundisch, D. (2019). Organizing their thoughts—How online review templates affect the review text. In Proceedings of the European Conference on Information Systems, Stockholm & Uppsala, Sweden (ECIS 2019).
- Puntoni, S., De Langhe, B., & Van Osselaer, S. M. (2009). Bilingualism and the emotional intensity of advertising language. *Journal of Consumer Research*, 35, 1012–1025.
- Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems*, 25, 145–181.
- Radziwill, N. M., & Benton, M. C. (2017). *Evaluating quality of chatbots and intelligent conversational agents*. *arXiv Preprint:170404579*.
- Ran, Y., & Wan, E. W. (2023). Enjoyment or autonomy? The interactive effect of brand ritual and brand personality on consumer purchase. *Psychology & Marketing*, 40, 89–106.

- Ransbotham, S., Lurie, N. H., & Liu, H. (2019). Creation and consumption of mobile word of mouth: How are mobile reviews different? *Marketing Science*, 38, 773–792.
- Reddy, T. (2017). How chatbots can help reduce customer service costs by 30%. Retrieved November 25, 2019 from <https://www.ibm.com/blogs/watson/2017/10/how-chatbots-reduce-customer-service-costs-by-30-percent/>.
- Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review Neuroscience*, 27, 169–192.
- Rosen, S., & Tesser, A. (1972). Fear of negative evaluation and the reluctance to transmit bad news. *Journal of Communication*, 22, 124–141.
- Sachdeva, A., Kim, A., & Dennis, A. R. (2024). Taking the Chat out of Chatbot? Collecting user reviews with Chatbots and web forms. *Journal of Management Information Systems*, 41, 146–177.
- Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the impact of ‘humanizing’ customer service chatbots. *Information Systems Research*, 32, 736–751.
- Schick, A., Feine, J., Morana, S., Maedche, A., & Reininghaus, U. (2022). Validity of chatbot use for mental health assessment: Experimental study. *JMIR mHealth and uHealth*, 10, 1–33.
- Schneider, C., Weinmann, M., Mohr, P., & vom Brocke, J. (2021). When the stars shine too bright: The influence of multidimensional ratings on online consumer ratings. *Management Science*, 67, 3871–3898.
- Schuetzler, R. M., Giboney, J. S., Grimes, G. M., & Nunamaker Jr, J. F. (2018). The influence of conversational agent embodiment and conversational relevance on socially desirable responding. *Decision Support Systems*, 114, 94–102.
- Seeger, A. M., Pfeiffer, J., & Heinzl, A. (2021). Texting with human-like conversational agents: Designing for anthropomorphism. *Journal of the Association for Information Systems*, 22, 931–967.
- Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Wiley.
- Tedeschi, J. T. (2013). *Impression management theory and social psychological research*. Academic.
- Thomaz, F., Salge, C., Karahanna, E., & Hulland, J. (2020). Learning from the Dark web: Leveraging conversational agents in the era of hyper-privacy to enhance marketing. *Journal of the Academy of Marketing Science*, 48, 43–63.
- Tonietto, G. N., & Barasch, A. (2021). Generating content increases enjoyment by immersing consumers and accelerating perceived time. *Journal of Marketing*, 85, 83–100.
- Tourangeau, R., Couper, M. P., & Steiger, D. M. (2003). Humanizing self-administered surveys: Experiments on social presence in web and IVR surveys. *Computers in Human Behavior*, 19, 1–24.
- Tsekouras, D. (2017). The effect of rating scale design on extreme response tendency in consumer product ratings. *International Journal of Electronic Commerce*, 21, 270–296.
- Van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo arigato Mr. Roboto: Emergence of automated social presence in organizational frontlines and customers’ service experiences. *Journal of Service Research*, 20, 43–58.
- Van Pinxteren, M. M., Pluymaekers, M., & Lemmink, J. G. (2020). Human-like communication in conversational agents: A literature review and research agenda. *Journal of Service Management*, 31, 203–225.
- Verhagen, T., Van Nes, J., Feldberg, F., & Van Dolen, W. (2014). Virtual customer service agents: Using social presence and personalization to shape online service encounters. *Journal of Computer-Mediated Communication*, 19, 529–545.
- Von der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S. H. (2010). It doesn’t matter what you are! Explaining social effects of agents and avatars. *Computers in Human Behavior*, 26, 1641–1650.
- Wang, Y., Wang, Z., Zhang, D., & Zhang, R. (2019). Discovering cultural differences in online consumer product reviews. *Journal of Electronic Commerce Research*, 20, 169–183.
- Woolley, K., & Sharif, M. A. (2021). Incentives increase relative positivity of review content and enjoyment of review writing. *Journal of Marketing Research*, 58, 539–558.
- Yelp (2024). Retrieved February 23, 2024 from [yelp.com/guidelines](https://www.yelp.com/guidelines).
- Yin, D., Bond, S. D., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly*, 38, 539–560.
- You, Y., Vadakkepatt, G. G., & Joshi, A. M. (2015). A meta-analysis of electronic word-of-mouth elasticity. *Journal of Marketing*, 79, 19–39.
- Zhang, M. (2018). Proven strategy to get reviews on Amazon using ManyChat! Retrieved September 7, 2021 from <https://www.amazonstep.com/blog/amazon-reviews-manychat/>.
- Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, 74, 133–148.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.