Human-Machine Communication
Volume 8, 2024
https://doi.org/10.30658/hmc.8.4

Chatbot vs. Human: The Impact of Responsive Conversational Features on Users' Responses to Chat Advisors

Stefanie H. Klein[1] **and Sonja Utz**[1,2]

1 Everyday Media Lab, Leibniz-Institut für Wissensmedien, Tübingen, Germany
2 Department of Psychology, Eberhard Karls Universität Tübingen, Germany

Abstract

As educational organizations increasingly consider supporting or replacing human chat advisors with chatbots, it is crucial to examine if users perceive a chatbot differently from a human. Chatbots' conversational features may signal responsiveness and thus improve user responses. To explore this, we conducted three online experiments (N_{total} = 1,005) using a study advising setting. We computed pooled data analyses because the individual study results did not provide clear support for our hypotheses. Results indicate that users prefer human agents regarding competence and intention to use but not perceived enjoyment. Responsiveness increased likability, warmth, and satisfaction. Perceptions of the interaction mediated the responsiveness effects. Our findings suggest that educational organizations can support their study advising departments with well-functioning chatbots without eliciting negative user responses.

Keywords: agent type, responsiveness, chatbot, user response, human-machine communication

CONTACT Stefanie H. Klein • s.klein@iwm-tuebingen.de • Everyday Media Lab • Leibniz-Institut für Wissensmedien • Schleichstrasse 6 • 72076 Tübingen, Germany

ISSN 2638-602X (print)/ISSN 2638-6038 (online)
www.hmcjournal.com

Introduction

Chatbots are text-based automated agents interacting with users through natural language (Shawar & Atwell, 2007). They increasingly rely on artificial intelligence (AI-)based technologies, including large language models (LLMs) like GPT-3 (Kasneci et al., 2023). Already a part of many organizations' communication policies, AI-based chatbots are also increasingly used in educational contexts (e.g., for learning support or academic advising; Karrenbauer et al., 2021; Kasneci et al., 2023). Academic advising can be prescriptive, providing information about administrative issues, or developmental, aiming at defining and exploring study or career goals (Gordon, 1994; Mottarella et al., 2004). Developmental advising holds great potential for using AI-based chatbots but has received little attention in previous research (Meyer von Wolff et al., 2020). However, because developmental advising entails more than simply answering information requests, we deem it crucial to systematically investigate whether potential students prefer to be advised by a human or a chatbot (Sundar, 2020). Although developers have made significant technological progress in developing conversational agents, increasing chatbot acceptance among users remains a challenge (Neururer et al., 2018). Responsiveness, in the form of backchanneling cues, is a promising conversational feature that has been positively linked to organizational and relational outcomes in prior research on human-human (Davis & Perkowitz, 1979; De Ruyter & Wetzels, 2000) and human-robot interactions (Birnbaum et al., 2016) but has hardly been studied in chatbots, as shown in a recent systematic review by Van Pinxteren et al. (2020). We thus consider it important to examine the effects of chat agents' responsiveness on user responses. In three online vignette experiments, we aim to answer the research question: "To what extent do agent type (chatbot vs. human) and responsiveness influence users' responses to chat advisors?" Additionally, we look at the underlying mechanisms from three levels relating to different aspects of the interaction (i.e., the interaction in general, the dialogic nature, and the content of the conversation).

Related Research and Theoretical Background

User Responses to Chat Agents

Before turning to prior work on agent type and responsiveness, we want to introduce the user responses that form our dependent variables. Following previous work in human-agent interaction (e.g., Diers, 2020; S. Lee & Choi, 2017; Lou et al., 2021) and academic advising (e.g., Mottarella et al., 2004), we examine users' general attitude toward the way of communicating with an organization. To gain a deeper understanding, we look at specific cognitive components; for example, the extent to which the advisor is perceived as likable (b), intelligent (c), warm (d), and competent (e). These perceptions are basic dimensions in evaluating new actors (Bartneck et al., 2009; Fiske, 2018). Likability and perceived intelligence are concepts stemming from human-robot interaction research. Warmth (i.e., good social intentions) and competence (i.e., the ability or expertise of the advisor) play a crucial role in evaluating human study advisors (Lou et al., 2021; Mottarella et al., 2004). Likability and warmth cover the social aspect of agent perception, while intelligence and competence are rather task-related (Bartneck et al., 2009; Fiske, 2018). We also capture users' affective and behavioral ratings of chat advisors to provide a comprehensive picture. Perceived enjoyment (f) constitutes the affective component. It refers to the extent to which interacting with a system is perceived as pleasurable and fun (Diers, 2020). We conceptualize satisfaction (g) as participants' perceived performance of the advisor they see in the vignette. In line with the Technology Acceptance Model, a prominent model to explain users' acceptance and usage of emerging technologies (Venkatesh, 2000), attitude, perceived enjoyment, and satisfaction are considered to be antecedents of user acceptance in terms of intention to use the way of communication with the organization (Diers, 2020; S. Lee & Choi, 2017). We also include the intention to use the communication medium (h), which, in turn, is considered to predict actual usage behavior (Venkatesh, 2000), but we were unable to measure this in our vignette studies.

We use the Media Are Social Actors (MASA) paradigm as an overarching theoretical framework for examining the impact of social cues on user responses to media (Lombard & Xu, 2021). The MASA paradigm extends the Computers Are Social Actors (CASA) paradigm, which states that people apply social rules to their interactions with computers (Nass & Moon, 2000) by considering a medium's social cues and the social signals these elicit to users as crucial for activating user responses. Lombard and Xu adopt Fiore et al.'s definition of social cues as "features salient to observers because of their potential as channels of useful information" (2013, p. 2). In contrast, social signals refer to the interpretation of a sender medium's social cues by the receiver (Fiore et al., 2013). Examples of a medium's social cues include gestures, motion, and language use. These can send out social signals of social identity, interactivity, and responsiveness to users (Lombard & Xu, 2021). This research focuses on cues signaling identity (agent type: human vs. chatbot) and responsiveness (i.e., the use or nonuse of verbal backchanneling cues).

A helpful model for investigating the impact of such cues is the Modality-Agency-Interactivity-Navigability (MAIN) model (Sundar, 2008). The MAIN model postulates that cognitive heuristics about an interaction's character and content are triggered when visual and identity cues are used in the interface (Sundar, 2008). The features of an interface can thus shape users' interaction experience (Sundar, 2020). The model identifies four affordances, which are present in most media: Modality, Agency, Interactivity, and Navigability. In this research, we focus on agency, which refers to the information source whose identity is communicated by an interface, in our case, an agent, to the user (Sundar, 2008). Visual or verbal cues (e.g., a human-like picture or backchannel utterances) can communicate an interface's identity (Sundar, 2008). These cues can trigger cognitive heuristics, like the machine heuristic and the social presence heuristic, which likely affect users' responses to chat advisors (Sundar, 2008).

Agent Type: Chatbot vs. Human

According to the MAIN model, machines are expected to lack emotions and, thus, be objective, rule-governed, and invariant on the one hand (Sundar, 2008). Machine-like cues can positively influence credibility perceptions by triggering the machine heuristic (Sundar, 2008). For instance, Sundar and Nass (2001) found that news displayed by machine-like systems is perceived as more objective than news displayed by humans. On the other hand, machines are stereotyped as unemotional and cold (Sundar, 2020).

If the positive or negative impact of the machine heuristic prevails (i.e., whether people favor humans or machine agents) strongly depends on the task at hand (M. K. Lee, 2018). We argue that in developmental study advising, people might expect to talk to a human because they consider machines as "unfit for 'human tasks' that involve subjective judgments and emotional capabilities" (Sundar, 2020, p. 80). Advising prospective students about degree programs requires not only rational data processing but also intuitive judgments and interpersonal competencies (Mottarella et al., 2004). The outcome of developmental study advising can have decisive consequences for a person's life. Thus, in this context, people might prefer communicating with a human because they consider them more flexible, adaptable, and sympathetic than a machine—in our case, a chatbot. Previous studies found that machine-like cues hinder positive agent evaluations, while human-like cues promote positive assessments; for example, users rated human agents (vs. chatbots) higher regarding expected likability and social presence (Spence et al., 2014) and social attraction (Lew & Walther, 2023). Given theory and prior research, we expect participants to have a more positive attitude toward a human advisor. They should also consider a human more likable, intelligent, warm, and competent than the chatbot (Lou et al., 2021). Moreover, we expect perceived enjoyment, satisfaction, and use intention to be higher for the human advisor (Prahl & Van Swol, 2021):

> **H1:** Human identity cues have a positive effect on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) satisfaction, and (h) intention to use ($\mu_{\text{human identity cues}} > \mu_{\text{chatbot identity cues}}$).

Responsiveness

The general concept of responsiveness stems from the interpersonal relationship literature, where it refers to the likelihood of each partner responding to the other and the proportions of relevant and adequate responses (Davis, 1982). In close interpersonal relationships, responsiveness refers to "the processes through which relationship partners attend to and respond supportively to each other's needs, wishes, concerns, and goals, thereby promoting each other's welfare" (Reis & Clark, 2013, p. 400). It is considered pivotal for human attachment processes. Responsiveness depends on how partners perceive and respond to each other's needs (Reis & Clark, 2013). When a partner feels that their needs are being met, feelings of closeness and mutual sympathy emerge. Partners who are responsive (i.e., psychologically empathetic, attentive, and supportive of one another) benefit in terms of liking, well-being, and satisfaction (Birnbaum et al., 2016; Davis & Perkowitz, 1979; Reis & Clark, 2013). Responsive behavior manifests itself in asking questions, paralinguistic behavior in the form of backchannel utterances, summarizing and paraphrasing what has been said, and expressing understanding (Maisel et al., 2008). Backchannel cues signal "attention to, support or encouragement for, or even acceptance of the speaker's message" (Mulac et al., 1998, p. 647). In service encounters, responsiveness has been shown to increase outcomes such as customer satisfaction and trust (De Ruyter & Wetzels, 2000). In human-machine communication (HMC), positive responsiveness effects on perceived competence, sociability, and willingness to use have been found for social robots (Birnbaum et al., 2016). There is first evidence that backchanneling increases the intention to use a chatbot (S. Lee et al., 2020). Taken together, we expect that responsive (vs. not responsive) advisors increase cognitive, affective, and behavioral user responses (Birnbaum et al., 2016; De Ruyter & Wetzels, 2000):

> **H2:** Responsive verbal cues have a positive effect on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) satisfaction, and (h) intention to use ($\mu_{\text{responsive verbal cues}} > \mu_{\text{no responsive verbal cues}}$).

The Interplay Between Agent Type and Responsiveness

We do not know yet whether the effect of responsiveness is the same across both types of agents or not. According to expectation violation theory (Burgoon & Hale, 1988; Grimes et al., 2021), initial user expectations of an agent's performance and whether they are confirmed or violated matter. Users attribute more attention to violated expectations than confirmed ones (Burgoon & Hale, 1988). Expectation violations can be positive or negative: A positive violation is seen as beneficial (e.g., when the user perceives a conversational agent as better than expected). In contrast, a negative violation indicates that the user expected more of the agent than they received (Burgoon & Hale, 1988). By telling users they are about to chat with a chatbot, compared to a human, expectations of the agent are reduced (Grimes et al., 2021). When the chatbot employs responsive cues and thus strongly resembles the human agent in its conversational characteristics, a responsive chatbot might receive more positive user responses than a chatbot without responsive verbal cues. As overly anthropomorphic chatbots are often perceived as eerie (Mori et al., 2012), users might feel uncomfortable talking to a responsive chatbot. Therefore, one could also expect the responsive chatbot to receive more negative user responses than the chatbot without responsive verbal cues. Given these conflicting lines of reasoning and the lack of previous research (except, e.g., Beattie et al., 2020; Sundar et al., 2016), we formulate an exploratory research question:

> **RQ1:** Is there an interaction effect between agent type (human vs. chatbot identity cues) and responsiveness (responsive vs. no responsive verbal cues) on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) satisfaction, and (h) intention to use?

Underlying Processes of Agent Type and Responsiveness Effects

To contribute to a more comprehensive understanding of the effects of agent type and responsiveness on user responses, we consider the processes underlying this relationship from three levels of analysis that relate to different aspects of interaction believed to be important in virtual interactions with social actors (Go & Sundar, 2019; Van der Goot & Etzrodt, 2023): the interaction in general, the dialogic nature, and the conversation content.

When looking at the interaction in general, we examine the mediating role of social presence. Social presence is defined as the perception of "being with another" (Biocca et al., 2003, p. 468). In HMC, it refers to the user's perception of interacting with a social entity rather than a machine (Sundar, 2008). The concept has been shown to positively impact attitudinal and behavioral outcomes in virtual interactions (Gefen & Straub, 2004;

Oh et al., 2018). Social presence is a fleeting judgment of an interaction influenced by the medium (Biocca et al., 2003). For instance, agents that provide human-like visual and verbal cues lead to stronger perceptions of social presence than agents that do not (S. Lee et al., 2020; Sundar, 2008). We thus expect human and responsive advisors to elicit higher levels of social presence, resulting in higher cognitive, affective, and behavioral user-related outcomes (Biocca et al., 2003; Go & Sundar, 2019).

> **H3:** The effects of agent type on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) satisfaction, and (h) intention to use are mediated by social presence.

> **H4:** The effects of responsiveness on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) satisfaction, and (h) intention to use are mediated by social presence.

Next, we look at the dialogic nature of the conversation. Dialogue is a fundamental feature of human conversations and provides the interlocutors with a sense of reciprocity, cooperation, and support (Kent & Taylor, 2002), also attributed to responsiveness (Reis & Clark, 2013). Conversations with responsively communicating agents should be perceived more as a dialogue than ones with an agent not using responsive verbal cues. Similar effects have been found for verbal cues signaling message contingency (Go & Sundar, 2019; Sundar et al., 2016). Perceived dialogue has been shown to increase advisor perceptions and usage intention (Go & Sundar, 2019). We argue that responsive verbal cues positively affect user responses via perceived dialogue:

> **H5:** The effects of responsiveness on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) satisfaction, and (h) intention to use are mediated by perceived dialogue.

Closely related to the concept of perceived dialogue but focused more on the actual content of the conversation is the concept of feeling heard, defined as "the feeling that one's communication is received with attention, empathy, respect, and in a spirit of mutual understanding" (Roos et al., 2023, p. 5). Responsive verbal cues could reinforce the user's feeling of being heard, leading to more positive evaluations of the advisor and the interaction. As feeling heard is a new and under-researched concept, we want to answer the following research question:

> RQ2: Does feeling heard mediate the effects of responsiveness on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) satisfaction, and (h) intention to use?

All three potential mediating mechanisms are considered to play important roles in virtual interactions with social actors (Go & Sundar, 2019), which is why we believe they could operate in parallel (see Ischen et al., 2020 for a similar approach). Feelings of social presence and being heard as well as the perception of dialogue in an interaction all involve notions of reciprocity, responsiveness, and mutual understanding (Roos et al., 2023; Sundar et al., 2016). Therefore, we will simultaneously investigate whether social presence, perceived dialogue, and feeling heard mediate the responsiveness effects on our outcomes of interest. Entering all three mediators in the same model controls for shared variance and provides stronger evidence for conclusions about the underlying processes (Hayes, 2022).

Overview of the Current Studies

In three online vignette experiments conducted in 2021, we investigated the extent to which agent type and responsiveness influence users' responses to chat advisors. Vignette designs are common in HMC research (Greussing et al., 2022; e.g., Abendschein et al., 2021; Beattie et al., 2020). The local ethics committee of the Leibniz-Institut für Wissensmedien, Tübingen, approved the studies. Informed consent was obtained from participants before their participation. Preregistrations, materials, data, and additional results are freely accessible on OSF: https://osf.io/w8dzv.

Study 1

Method

We conducted a 2 (agent type: chatbot vs. human) × 2 (responsiveness: absence vs. presence of responsive verbal cues) between-subjects experiment. Participants were recruited via the online sampling platform Prolific. Of the 280 participants who completed the study, 253 passed the agent type manipulation check and were retained (n_{female} = 101, age: M = 28.18, SD = 9.25, range = 18–69) (power analysis in Appendix A). Participants were randomly assigned to one of four conditions. After providing informed consent, all participants saw a vignette in the form of a pre-recorded animated chat conversation between the study advisor, Sophie, and Marc, a prospective student at Sophie's university. The advisor asked the user several questions during the conversation to find out his interests. Based on Marc's answers, Sophie recommended suitable degree programs. Finally, we asked participants to complete a survey about their perceptions of the interaction and key demographics. The design of the chat interface resembled the design of contemporary messenger interfaces used in practice (Appendix B). A robot icon was used to represent the chatbot, and it introduced itself as "Sophie, the chatbot of the student advisory service." The human advisor was represented with the portrait of a businesswoman, and she introduced herself as "Sophie, a student advisor" (see Go & Sundar, 2019 for a similar approach). All conversations were equal in content and without disruptions. Responsiveness was manipulated using short backchanneling cues and tokens like questions, paraphrases, and expressions signaling thinking processes (Maisel et al., 2008). Specifically, the responsive agents responded to the user's inputs with utterances like "Mhm," "Got it," and "Hmm, let me think." The responsive agent also asked the user for his name and repeated it in the following input.

To measure the *attitude* toward the means of communication, we used a scale by Diers (2020). Specific cognitive user responses were assessed using the *likability* and *perceived intelligence* scales from the Godspeed Questionnaire (Bartneck et al., 2009) and the *warmth* and *competence* scales from Fiske (2018). Affective user response was assessed using a *perceived enjoyment* scale, and the behavioral aspect was reflected with an *intention to use* scale, both by Diers (2020). We asked participants to rate whether they would have found the advisor's behavior satisfactory if they had been in the student's position using Lagace et al.'s *satisfaction* scale (1991). We adopted the scales for *social presence, perceived dialogue*, and *feeling heard* from Gefen and Straub (2004), Sundar et al. (2016), and Roos et al. (2023), respectively. We included manipulation checks for agent type (adapted from Go & Sundar, 2019) and responsiveness (adapted from De Ruyter & Wetzels, 2000) to ensure effective manipulations. We assessed the same variables in all studies (Appendix C, descriptive statistics in Appendix D). Bivariate correlations were rather strong (Appendix E), ranging from $r = .39$ to $r = .89$ ($p < .001$). Attitude and intention to use, both drawn from the Technology Acceptance Model literature (Venkatesh, 2000) and relating to the communication with the organization, were strongly correlated ($r > .86$). Perceived dialogue and feeling heard also correlated strongly, potentially due to similarity in item content. Internal consistency of all constructs was satisfactory (Cronbach's $\alpha > .80$).

Results

As intended, participants in the responsive conditions perceived the agent as significantly more responsive ($M = 5.89$, $SD = 1.03$) than those in the conditions without responsive verbal cues ($M = 5.15$, $SD = 1.36$), as a Welch two-sample t-test showed ($t(237.59) = 4.88$, $p < .001$, $d = 0.61$). To test H1 and H2 and to answer RQ1, we carried out a two-way multivariate analysis of variance (MANOVA). Using Pillai's trace, there was a significant main effect of agent type on the outcomes ($V = 0.15$, $F(1, 248) = 5.40$, $p < .001$). Separate univariate analyses of variance (ANOVA) only revealed a significant agent type effect for satisfaction (g) ($F(1, 248) = 6.72$, $p = .010$): Participants in the chatbot conditions tended to be more satisfied ($M = 5.56$, $SD = 1.24$) than participants in the human conditions (Satisfaction: $M = 5.13$, $SD = 1.48$; $t(239.74) = -2.51$, $p = .013$, $d = -0.32$). Neither the responsiveness effect ($V = 0.04$, $F(1, 248) = 1.16$, $p = .324$) nor the interaction between agent type and responsiveness ($V = 0.02$, $F(1, 248) = 0.74$, $p = .657$) were significant. We rejected H1 and H2. We did not perform mediation analyses to test H3–H5 and to answer RQ2 because neither agent type nor responsiveness positively impacted the outcomes. We thus rejected H3–H5.

Discussion

Contrary to the hypotheses, neither human identity cues nor responsive behavior positively affected user responses. The zero effects of responsiveness are striking, given the significant responsiveness effects on the manipulation check. Moderation analyses did not yield significant results. The sample size was slightly below the target size due to our exclusion criterion. As we had based our power considerations on a small interaction effect that Go and Sundar (2019) found in a study where participants directly interacted with chat agents, the effect sizes in our vignette design could be even smaller. Hence, we decided to replicate our study with a larger sample.

Study 2

Method

The experimental design and measures were equal to those in Study 1. We aimed to recruit 403 participants via university mailing lists. A total of 520 participants completed the study. We excluded 118 participants because they failed the agent type manipulation check and one who admitted to not having answered the questionnaire reliably, which led to a final sample of $N = 401$ ($n_{\text{female}} = 287$, age: $M = 24.26$, $SD = 6.11$, range = 18–69).

Results

As intended, participants in the responsive conditions scored significantly higher on the responsiveness manipulation check ($M = 5.86$, $SD = 1.20$) than participants in the conditions without responsive verbal cues ($M = 4.67$, $SD = 1.54$) as a Welch two-sample t-test showed ($t(357.33) = 8.51$, $p < .001$, $d = 0.86$). We conducted a two-way MANOVA to test H1 and H2 and answer RQ1. Using Pillai's trace, there was a significant main effect of agent type ($V = 0.06$, $F(1, 394) = 3.36$, $p = .001$). The responsiveness effect ($V = 0.03$, $F(1, 394) = 1.58$, $p = .129$) and the interaction between agent type and responsiveness ($V = 0.02$, $F(1, 394) = 0.95$, $p = .388$) were not significant in the multivariate model. Separate ANOVAs revealed a significant agent type effect for likability (b) ($F(1, 394) = 4.92$, $p = .036$). Participants in the chatbot conditions rated the agent more likable ($M = 5.63$, $SD = 1.05$) than participants in the human conditions ($M = 5.41$, $SD = 1.07$; $t(397) = -2.21$, $p = .035$, $d = -0.21$). Although the responsiveness effect was not significant, we computed separate ANOVAs, revealing positive responsiveness effects on warmth (d) ($F(1, 394) = 6.32$, $p = .012$) and satisfaction (g) ($F(1, 394) = 6.53$, $p = .011$). We rejected H1 and accepted H2d), g). We computed parallel multiple mediator models (Hayes, 2022) predicting warmth and satisfaction using the R package *lavaan* (Rosseel et al., 2021) but did not find significant indirect effects (OSF).

Discussion

Like in Study 1, human identity cues did not significantly improve user responses in Study 2. However, we found significant effects of responsive conversational cues on warmth and satisfaction. We conducted a third study to clarify our findings from Studies 1 and 2.

Study 3

Method

We collected data from 418 participants via the crowdsourcing platform Clickworker. Three hundred fifty-one participants passed the agent type manipulation check and were retained ($n_{\text{female}} = 127$, age: $M = 38.53$, $SD = 12.28$, range = 18–73). The experimental design and measures were equal to those we used in Studies 1 and 2.[1]

1. Study 3 was designed to additionally explore the impact of agent response time on users' responses, so response time (immediate or dynamically delayed) was included as a third experimental factor. However, because the results were not vital to answering our research question, we decided to move them to OSF.

Results

Participants in the responsive conditions scored significantly higher on the responsiveness manipulation check ($M = 6.08$, $SD = 0.88$) than participants in the conditions without responsive verbal cues ($M = 5.54$, $SD = 0.96$), as shown in a two-sample t-test ($t(349) = 5.54$, $p < .001$, $d = 0.59$). To test H1–H2 and to answer RQ1, a two-way MANOVA was carried out. Using Pillai's trace, there was a significant main effect of agent type ($V = 0.08$, $F(1, 347) = 4.11$, $p < .001$) on the outcome variables. The responsiveness effect ($V = 0.04$, $F(1, 347) = 1.66$, $p = .108$) and the interaction between agent type and responsiveness ($V = 0.02$, $F(1, 347) = 1.04$, $p = .403$) were not significant in the multivariate model. Separate ANOVAs only revealed a significant agent type effect on competence (e) ($F(1, 347) = 12.49$, $p < .001$) and a marginally significant responsiveness effect on likability (b) ($F(1, 347) = 3.85$, $p = .051$). Follow-up tests yielded that participants in the human conditions rated the agent more competent ($M = 5.76$, $SD = 0.93$) than participants in the chatbot conditions ($M = 5.36$, $SD = 1.12$; $t(349) = 3.54$, $p < .001$, $d = 0.01$). We, therefore, accepted H2e) and rejected H2. As we did not find significant positive agent type or responsiveness effects (H2), we did not compute parallel multiple mediator models.

Discussion

Participants perceived the human agent as more competent than the chatbot, but no significant responsiveness effects were found. Still, a considerable proportion of participants did not recognize the alleged human, reducing the sample and resulting in a power loss for estimating interaction effects (posthoc power = 74.40%, $N = 351$, $\alpha = .05$, $f = .14$). To mitigate the potential power issues of Studies 1 and 3, we conducted analyses based on the pooled data from all studies.

Additional Analyses: Pooled Data

Using the pooled data ($N = 1{,}005$) and controlling for study number, we performed an exploratory MANCOVA to clarify the main effects on the outcomes. Using Pillai's trace, the inclusion of study number as a control variable indicated differences between studies ($V = 0.11$, $F(2, 995) = 7.19$, $p < .001$). Specifically, participants in Study 2 showed lower values on all outcomes than those in Studies 1 and 3, pointing to a generational effect. Study 2 comprised university students who were younger on average than participants in Studies 1 and 3 and, thus, may have had more experience with chat advisors. In contrast to the individual study results, using Pillai's trace, we found significant main effects of agent type (H1; $V = 0.07$, $F(1, 995) = 9.10$, $p < .001$) and responsiveness (H2; $V = 0.03$, $F(1, 995) = 3.25$, $p = .001$). The interaction between agent type and responsiveness (RQ1) was not significant ($V < .01$, $F(1, 995) = 0.59$, $p = .783$). Univariate ANCOVAs and pairwise comparisons yielded significant positive effects of human identity cues on competence (e) and intention to use (h). Participants perceived the interaction with the human agent as less enjoyable than the interaction with the chatbot. In addition, significant positive responsiveness effects emerged for likability (b), warmth (d), and satisfaction (g), strengthening the individual study findings (Table 1, with adjusted means in Table 2).

TABLE 1 Two-Way ANCOVA Statistics and Effect Sizes for Study Variables (Pooled Data)

Variable	Effect	*F* ratio	*p*	$\eta^2_{partial}$
Attitude	Study	15.80	< .001	.03
	AT	3.27	.071	.00
	R	0.64	.422	.00
	AT × R	0.44	.510	.00
Likability	Study	3.23	.040	.01
	AT	3.79	.052	.00
	R	6.12	.014	.01
	AT × R	0.10	.755	.00
Perceived intelligence	Study	16.43	< .001	.03
	AT	2.41	.121	.00
	R	1.97	.161	.00
	AT × R	0.34	.558	.00
Warmth	Study	10.45	< .001	.02
	AT	0.51	.475	.00
	R	8.93	.003	.01
	AT × R	0.71	.400	.00
Competence	Study	7.02	< .001	.01
	AT	7.32	.007	.01
	R	0.04	.847	.00
	AT × R	0.25	.616	.00
Perceived enjoyment	Study	26.27	< .001	.05
	AT	4.95	.026	.01
	R	2.67	.102	.00
	AT × R	0.01	.927	.00
Satisfaction	Study	14.29	< .001	.03
	AT	3.28	.070	.00
	R	6.09	.013	.01
	AT × R	1.11	.293	.00
Intention to use	Study	21.83	< .001	.04
	AT	6.90	.009	.01
	R	0.48	.487	.00
	AT × R	0.15	.694	.00

Note. N = 1,005. ANCOVA = analysis of covariance. Study = study number, AT = agent type, R = responsiveness. df = 1,995, except df_{Study} = 2,995.

TABLE 2 Adjusted Means and Effect Sizes for Study Variables (Pooled Data)

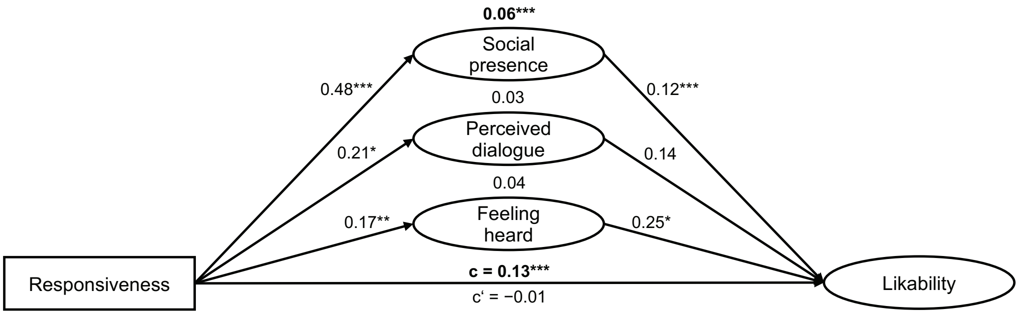
Variable	Agent Type				Responsiveness			
	Chatbot	Human	*p*	$\eta^2_{partial}$	Absent	Present	*p*	$\eta^2_{partial}$
Attitude	4.45	4.64	.068	.00	4.50	4.58	.435	.00
Likability	5.70	5.57	.050	.00	5.55	5.71	.013	.01
Perceived intelligence	5.58	5.68	.125	.00	5.58	5.67	.175	.00
Warmth	5.21	5.17	.481	.00	5.09	5.29	.003	.01
Competence	5.47	5.64	.006	.01	5.55	5.54	.844	.00
Perceived enjoyment	4.27	4.06	.029	.00	4.10	4.26	.089	.00
Satisfaction	5.20	5.04	.076	.00	5.01	5.24	.012	.01
Intention to use	4.35	4.65	.009	.01	4.45	4.52	.517	.00

Note. $N = 1{,}005$. One-way ANCOVAs controlled for study number.

To analyze the underlying relationship processes between agent type and responsiveness and the outcomes, we computed three parallel multiple mediator models predicting the outcomes significantly impacted by responsiveness; for example, likability (b), warmth (d), and satisfaction (g), using the R package *lavaan* (Rosseel et al., 2021). Including social presence, perceived dialogue, and feeling heard allowed us to model our three potential mediation levels simultaneously. We operationalized the latent constructs using reflective measurement models composed of the corresponding items.[2] All standardized factor loadings were sufficiently strong ($\lambda > .50$) and significant ($p < .001$). Knowing the mediators to be strongly correlated, we specified their covariances. We controlled the models for study number. Model fit was acceptable (Westland, 2015). As expected, we found high correlations between social presence and perceived dialogue ($r = .67$ in the likability, $r = .66$ in the warmth and satisfaction models), social presence and feeling heard ($r = .61$), and perceived dialogue and feeling heard ($r = .95$) throughout the models. Figures 1–3 display the results of the mediation models predicting likability, warmth, and satisfaction. Significant positive indirect effects emerged for likability via social presence, for warmth via social presence and feeling heard, and for satisfaction via social presence and perceived dialogue. The direct effects (i.e., the effects of responsiveness on the outcomes when the mediators were included in the model) were not significant, suggesting full mediations.

2. The standardized factor loadings of the two inversely coded feeling heard items were < .50. We thus followed Roos et al. (2023) in specifying the covariance between the residuals accounting for different response behaviors for inversely coded items. The error terms correlated moderately ($r = .42$, $p < .001$).

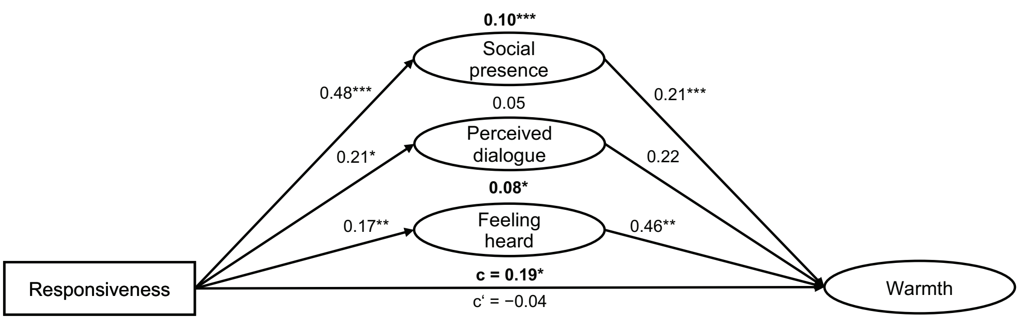
FIGURE 1 Parallel Multiple Mediator Model for the Effect of Responsiveness on Likability (Pooled Data)



Note. df = 261. X^2 = 1592.500, $p < .001$, CFI = .914, RMSEA = .071, CI_{RMSEA}(.068, .075), SRMR = .063. Unstandardized coefficients. Controlled for study number. $R^2_{\text{Likability}} = .473$, $R^2_{\text{Social presence}} = .054$, $R^2_{\text{Perceived dialogue}} = .014$, $R^2_{\text{Feeling heard}} = .009$.

* $p < .05$, ** $p < .01$, *** $p < .001$.

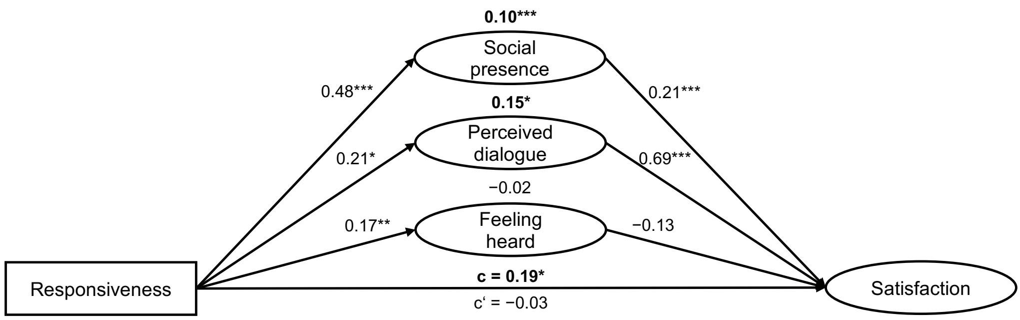
FIGURE 2 Parallel Multiple Mediator Model for the Effect of Responsiveness on Warmth (Pooled Data)



Note. *df* = 285. χ^2 = 2406.803, $p < .001$, CFI = .875, RMSEA = .086, CI_{RMSEA}(.083, .089), SRMR = .071. Unstandardized coefficients. Controlled for study number. $R^2_{\text{Warmth}} = .734$, $R^2_{\text{Social presence}} = .054$, $R^2_{\text{Perceived dialogue}} = .015$, $R^2_{\text{Feeling heard}} = .009$.

* $p < .05$, ** $p < .01$, *** $p < .001$.

FIGURE 3 Parallel Multiple Mediator Model for the Effect of Responsiveness on Satisfaction (Pooled Data)



Note. df = 238. χ^2 = 1552.062, $p < .001$, CFI = .918, RMSEA = .074, CI_{RMSEA}(.071, .078), SRMR = .063. Unstandardized coefficients. Controlled for study number. $R^2_{Satisfaction}$ = .592, $R^2_{Social\ presence}$ = .054, $R^2_{Perceived\ dialogue}$ = .015, $R^2_{Feeling\ heard}$ = .010.

* $p < .05$, ** $p < .01$, *** $p < .001$.

General Discussion

We conducted three vignette experiments to answer the question: To what extent do agent type and responsiveness influence users' responses to chat agents? Three key findings and subsequent implications emerged.

First, no study clearly supported our hypothesis on the positive impact of human identity cues on user responses (H1). The pooled analyses, however, suggested that participants considered the human agent more competent and indicated a higher intention to use it compared to the chatbot. This aligns with our hypothesis and suggests the machine heuristic does not hold for *human* tasks like developmental study advising (Sundar, 2020). Still, participants preferred the chatbot in terms of perceived enjoyment. There were no effects on the other outcomes, aligning with prior research suggesting little overall difference in the perception of humans and anthropomorphic chatbots (Beattie et al., 2020; Nass & Moon, 2000). Only the pooled analyses yielded small effects of human identity cues, although statistical power was high. The agents' error-free answers and suitable program recommendations to the user might be the reason for the small effects. Data were collected before ChatGPT was launched, so the high quality of the answers allegedly stemming from a chatbot might have been surprising for the participants. Considering the rapid improvement of generative AI, the question of whether and how small identity cues affect chatbot evaluations gets even more important. We can conclude that regardless of how well a chatbot performs and how much people enjoy it, human agents seem to be preferred as study advisors of a university.

Second, we found significant positive responsiveness effects on warmth and satisfaction in Study 2 (H2). The pooled analyses confirmed these findings and yielded an additional significant positive responsiveness effect on likability. We showed that a responsive communication style elicits positive responses in contexts where agents must provide support and understanding. Positive responses were elicited regarding the agent's social traits like likability, warmth, and satisfaction, corresponding to earlier findings from interpersonal communication research (Davis & Perkowitz, 1979). Responsive cues alone might be too subtle to tell whether people will want to be advised by an agent and not as decisive for perceptions relating to the successful completion of a task. Once high efficiency, the primary reason for using chatbots (Følstad & Skjuve, 2019), is reached, these softer cues might become more important.

Third, perceptions of the agent mediated the relationship between responsiveness and likability, warmth, and satisfaction (H4-H5, RQ2). Social presence consistently mediated the effects on likability, warmth, and satisfaction, whereas perceived dialogue and feeling heard only mediated the effect on satisfaction and warmth, respectively. Overall, this is consistent with our theoretical assumption that responsive verbal cues from an interlocutor signal understanding and support, thereby leading to more positive user responses (Reis & Clark, 2013). However, because perceived dialogue and feeling heard were highly correlated and had similarities in item content (Appendix C), controlling for one construct eliminated the respective other's effect. This raises the question of whether the variables represent different constructs. We showed that verbal cues have the potential to make users feel more socially present and heard (Lombard & Xu, 2021). How the interaction is perceived appears to be more critical to perceptions of the advisor's social attributes and satisfaction than individual aspects (e.g., its dialogic nature and conversation content).

Interestingly, responsiveness did not interact with agent type. Responsive cues seem equally important to peoples' perception of chatbots and humans. Future research could employ other interindividual moderators that could affect the effect of agent type on user outcomes (e.g., affinity for technological interaction; Franke et al., 2019). Context-specific differences could also be explored; for example, responsive cues might matter more when a chatbot serves as an emotional support tool (e.g., Birnbaum et al., 2016 for social robots).

Agent type influenced certain outcomes, while responsiveness influenced others. The machine heuristic suggests machine actors are viewed as more objective, rule-based, and competent than humans (Sundar, 2008). Our results challenge this, as competence and intention to use were higher in the human conditions. The task of study advising, which we consider a *human task* at its core, might be the reason (Sundar, 2020). Additionally, the items used to assess intention to use referred to the way of communication with an organization. Thus, even if chatbots perform just as well as humans, a preference for talking to humans and an aversion to the use of automation and algorithms in universities' communication remain (Dietvorst et al., 2015).

We aimed to investigate users' perceptions of the agent and the interaction as well as the classical technology acceptance variables attitude and intention to use (Venkatesh, 2000). Agent and interaction perceptions are established antecedents of attitude and use intention, whereas the latter can predict actual usage (i.e., adoption; Diers, 2020; S. Lee & Choi, 2017; Venkatesh, 2000). Responsiveness, in contrast, is more likely to impact social perceptions and seems relevant to users' satisfaction. For researchers more interested in the processes underlying chatbot adoption, the variables affected by responsiveness become relevant as they might mediate responsiveness effects on user satisfaction, which in turn might influence intention to use (Lou et al., 2021).

Our research contributes to the emerging research field of HMC regarding the impact of social cues on the perception and evaluation of machine agents (Gambino et al., 2020; Lombard & Xu, 2021). The different user responses to humans and chatbots suggest that the media equation does not apply to all social interactions with machines (i.e., not all machines are always perceived as social actors; Van der Goot & Etzrodt, 2023). There is reason to believe that users mindfully evaluate the source depending on situational factors (e.g., the interaction context) and dispositional factors (e.g., personality) (E.-J. Lee, 2023). Van der Goot and Etzrodt (2023) recommend conducting more qualitative research to unravel these processes and to understand "how users negotiate the blurring boundaries between humans and machines" (p. 27). This question will become more important as human-like chatbots based on generative AI continue to gain traction.

Previous research on study advising has shown that a warm and supportive advising style is an influential factor for satisfaction and is even more important than the advising approach (Mottarella et al., 2004). A warm communication style is often associated with developmental advising. Although prescriptive advising is task-oriented and focuses on explaining requirements and procedures, a more responsive style could improve student acceptance. We suggest that researchers investigating the differences between various advising approaches should pay more attention to the advisor's communication style, whether human or chatbot.

The results have implications for university practitioners considering using chatbots in developmental advising. While perceived competence and intention to use were higher for the human advisor, chatbot scores for these variables were above the scale means. So, chatbot support for student advising might be an efficient addition when financial resources or staff shortages are an issue. Leveraging the benefits of automated communication can thus be feasible without eliciting negative user responses. Yet, developers must ensure that the chatbots work well (e.g., by adequately exploiting the advantages of LLMs; Kasneci et al., 2023). But the way the chatbot presents the information is also critical. Study advising chatbots should be designed to evoke feelings of warmth and support, which have been shown to facilitate successful advising (Mottarella et al., 2004). Integrating responsive features into chat interactions may help universities and schools build and maintain warm and supportive relationships with their (potential) students. A well-thought-out dialogue design can help integrate responsive verbal cues without too much financial or human effort.

Limitations and Future Research

Although vignette designs have high internal validity and give participants a unique perspective (Abendschein et al., 2021), they cannot offer as much ecological validity as experiments where participants directly interact with an agent. In our studies, participants were mere observers of the interaction, which could have increased their distance from the interaction, decreasing their involvement and identification with the user. The high nonrandom dropout rates due to failed agent type manipulation checks could have been related to the study design. To ensure experimental control, we kept the layout and content across all conditions constant. We thus manipulated agent type only in terms of the agent's introduction and avatar, which may have led participants to perceive the human agent as an anthropomorphic chatbot. Future studies could examine participants' direct interactions with chat agents to increase ecological validity. To ensure the comparability of our results, we used the same stimulus materials in all studies, which might have affected the validity of our results in case the stimuli did not optimally manipulate our independent variables. Future studies could use stimulus sampling (i.e., employ a variety of user-agent conversations) to reduce the impact of the unique features of a particular stimulus on the results and strengthen the conclusions (Jackson & Jacobs, 1983).

Our studies focused on investigating the effects of agent type and responsiveness on a wide range of dependent variables, including cognitive, affective, and behavioral user responses, but not on the relationships between the outcomes. As there is a plethora of scales from different disciplines that measure similar constructs (e.g., human-likeness perceptions; Ischen et al., 2023), researchers call for common conceptualizations and measurement scales for key outcomes (Følstad et al., 2021; Greussing et al., 2022). A systematic assessment and confirmatory factor analysis of common scales used in HMC research could shed light on what makes each construct unique, how the constructs are empirically related, and how they contribute to chatbot adoption.

Conclusion

In three experiments, we investigated the impact of agent type and responsiveness on a wide range of user-related outcomes in the context of study advising. Our results suggest that human agents are favored in terms of competence and intention to use but not in terms of perceived enjoyment. Further, the results indicate that responsiveness positively impacts users' perceptions of agent likability, warmth, and satisfaction, mainly by increasing perceptions of the interaction. Our studies add novel insights to the literature on human-machine communication and offer two practical implications: First, our findings may encourage educational organizations to support their study advising departments with chatbots. Second, the use of responsive language by human agents and chatbots could help organizations build and maintain healthy and sustainable relationships with their (potential) students. Due to significant advances in generative AI, we can expect that people will increasingly be unable to distinguish whether they are interacting with a human or a chatbot in the future. Therefore, it will continue to be crucial to systematically investigate the role of relatively small social cues in the perception and evaluation of AI-based chatbots.

Author Biographies

Stefanie H. Klein (MA, University of Stuttgart) is a PhD student in the Everyday Media Lab at Leibniz-Institut für Wissensmedien in Tübingen, Germany. Her research focuses on human-machine communication, particularly the impact of chatbots' conversational characteristics on user acceptance.

https://orcid.org/0000-0002-7563-4548

Sonja Utz (PhD, Catholic University of Eichstätt) is the head of the Everyday Media lab at Leibniz-Institut für Wissensmedien in Tübingen and a full professor for communication via social media at the University of Tübingen. Her research focuses on the effects of social and mobile media use, especially in knowledge-related contexts, and on human-machine interaction.

https://orcid.org/0000-0002-7979-3554

Center for Open Science

OPEN DATA OPEN MATERIALS PREREGISTERED

This article has earned the Center for Open Science badges for Open Data and Open Materials through Open Practices Disclosure. The authors have made their data and materials freely accessible at https://osf.io/w8dzv/. The article also earned a Preregistered badge for having a preregistered design available at https://osf.io/w8dzv/.

References

Abendschein, B., Edwards, C., & Edwards, A. (2021). The influence of agent and message type on perceptions of social support in human-machine communication. *Communication Research Reports, 38*(5), 304–314. https://doi.org/10.1080/08824096.2021.1966405

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Beattie, A., Edwards, A. P., & Edwards, C. (2020). A bot and a smile: Interpersonal impressions of chatbots and humans using emoji in computer-mediated communication. *Communication Studies*, *71*(3), 409–427. https://doi.org/10.1080/10510974.2020.1725082

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators and Virtual Environments*, *12*(5), 456–480. https://doi.org/10.1162/105474603322761270

Birnbaum, G. E., Mizrahi, M., Hoffman, G., Reis, H. T., Finkel, E. J., & Sass, O. (2016). What robots can teach us about intimacy: The reassuring effects of robot responsiveness to human disclosure. *Computers in Human Behavior*, *63*, 416–423. https://doi.org/10.1016/j.chb.2016.05.064

Burgoon, J. K., & Hale, J. L. (1988). Nonverbal expectancy violations: Model elaboration and application to immediacy behaviors. *Communication Monographs*, *55*(1), 58–79. https://doi.org/10.1080/03637758809376158

Davis, D. (1982). Determinants of responsiveness in dyadic interaction. In W. Ickes & E. S. Knowles (Eds.), *Personality, Roles, and Social Behavior* (pp. 85–139). Springer. https://doi.org/10.1007/978-1-4613-9469-3_4

Davis, D., & Perkowitz, W. T. (1979). Consequences of responsiveness in dyadic interaction: Effects of probability of response and proportion of content-related responses on interpersonal attraction. *Journal of Personality and Social Psychology*, *37*(4), 534–550. https://doi.org/10.1037/0022-3514.37.4.534

De Ruyter, K., & Wetzels, M. G. M. (2000). The impact of perceived listening behavior in voice-to-voice service encounters. *Journal of Service Research*, *2*(3), 276–284. https://doi.org/10.1177/109467050023005

Diers, T. (2020). *Akzeptanz von Chatbots im Consumer-Marketing: Erfolgsfaktoren zwischen Konsumenten und künstlicher Intelligenz* [Acceptance of chatbots in consumer marketing: Success factors between consumers and artificial intelligence]. Springer. https://doi.org/10.1007/978-3-658-29317-8

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

Fiore, S. M., Wiltshire, T. J., Lobato, E. J. C., Jentsch, F. G., Huang, W. H., & Axelrod, B. (2013). Toward understanding social cues and signals in human-robot interaction: Effects of robot gaze and proxemic behavior. *Frontiers in Psychology*, *4*(859), 2–15. https://doi.org/10.3389/fpsyg.2013.00859

Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science*, *27*(2), 67–73. https://doi.org/10.1177/0963721417738825

Følstad, A., Araujo, T., Law, E. L.-C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., Ischen, C., Wald, R., Catania, F., Meyer von Wolff, R., Hobert, S., & Luger, E. (2021). Future directions for chatbot research: An interdisciplinary research agenda. *Computing*, *103*, 2915–2942. https://doi.org/10.1007/s00607-021-01016-7

Følstad, A., & Skjuve, M. (2019). Chatbots for customer service: User experience and motivation. In B. R. Cowan & L. Clark (Eds.), *Proceedings of the 1st International Conference on Conversational User Interfaces* (pp. 1–9). Association for Computing Machinery. https://doi.org/10.1145/3342775.3342784

Franke, T., Attig, C., & Wessel, D. (2019). A personal resource for technology interaction: Development and validation of the affinity for technology interaction (ATI) Scale. *International Journal of Human–Computer Interaction*, *35*(6), 456–467. https://doi.org/10.1080/10447318.2018.1456150

Gambino, A., Fox, J., & Ratan, R. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, *1,* 71–86. https://doi.org/10.30658/hmc.1.5

Gefen, D., & Straub, D. W. (2004). Consumer trust in B2C e-Commerce and the importance of social presence: Experiments in e-Products and e-Services. *Omega*, *32*(6), 407–424. https://doi.org/10.1016/j.omega.2004.01.006

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, *97*, 304–316. https://doi.org/10.1016/j.chb.2019.01.020

Gordon, V. N. (1994). Developmental advising: The Elusive ideal. *NACADA Journal, 14*(2), 72–76. https://doi.org/10.12930/NACADA-19-201

Greussing, E., Gaiser, F., Klein, S. H., Straßmann, C., Ischen, C., Eimler, S., Frehmann, K., Gieselmann, M., Knorr, C., Lermann Henestrosa, A., Räder, A., & Utz, S. (2022). Researching interactions between humans and machines: Methodological challenges. *Publizistik*, *67*(4), 531–554. https://doi.org/10.1007/s11616-022-00759-3

Grimes, G. M., Schuetzler, R. M., & Giboney, J. S. (2021). Mental models and expectation violations in conversational AI interactions. *Decision Support Systems*, *144*, 113515. https://doi.org/10.1016/j.dss.2021.113515

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd ed.). Guilford Press.

Ischen, C., Araujo, T., van Noort, G., Voorveld, H., & Smit, E. (2020). "I am here to assist you today": The role of entity, interactivity and experiential perceptions in chatbot persuasion. *Journal of Broadcasting & Electronic Media*, *64*(4), 615–639. https://doi.org/10.1080/08838151.2020.1834297

Ischen, C., Smit, E., & Wang, E. (2023, November 22–23). *Assessing human-likeness perceptions: Measurement scales of conversational agents* [Paper presentation]. Conversations 2023 –7th International Workshop on Chatbot Research and Design, Oslo, Norway. Retrieved from https://web.archive.org/web/20240418110620/https://2023.conversations.ws/wp-content/uploads/2023/11/conversations_2023_positionpaper_13_ischen.pdf

Jackson, S., & Jacobs, S. (1983). Generalizing about messages: Suggestions for design and analysis of experiments. *Human Communication Research*, *9*(2), 169–191. https://doi.org/10.1111/j.1468-2958.1983.tb00691.x

Karrenbauer, C., König, C. M., & Breitner, M. H. (2021). Individual digital study assistant for higher education institutions: Status quo analysis and further research agenda. In F. Ahlemann, R. Schütte, & S. Stieglitz (Eds.), *Innovation Through Information Systems. WI 2021. Lecture Notes in Information Systems and Organisation* (vol. 48, pp. 108–124). Springer. https://doi.org/10.1007/978-3-030-86800-0_8

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kent, M. L., & Taylor, M. (2002). Toward a dialogic theory of public relations. *Public Relations Review*, *28*(1), 21–37. https://doi.org/10.1016/S0363-8111(02)00108-X

Lagace, R. R., Dahlstrom, R., & Gassenheimer, J. B. (1991). The relevance of ethical salesperson behavior on relationship quality: The pharmaceutical industry. *Journal of Personal Selling & Sales Management*, *11*(4), 39–47. https://doi.org/10.1080/08853134.1991.10753888

Lee, E.-J. (2023). Minding the source: Toward an integrative theory of human-machine communication. *Human Communication Research*, hqad034. https://doi.org/10.1093/hcr/hqad034

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1), 205395171875668. https://doi.org/10.1177/2053951718756684

Lee, S., & Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies*, *103*, 95–105. https://doi.org/10.1016/j.ijhcs.2017.02.005

Lee, S., Lee, N., & Sah, Y. J. (2020). Perceiving a mind in a chatbot: Effect of mind perception and social cues on co-presence, closeness, and intention to use. *International Journal of Human–Computer Interaction*, *36*(10), 930–940. https://doi.org/10.1080/10447318.2019.1699748

Lew, Z., & Walther, J. B. (2023). Social scripts and expectancy violations: Evaluating communication with human or AI chatbot interactants. *Media Psychology*, *26*(1), 1–16. https://doi.org/10.1080/15213269.2022.2084111

Lombard, M., & Xu, K. (2021). Social responses to media technologies in the 21st century: The media are social actors paradigm. *Human-Machine Communication*, *2,* 29–55. https://doi.org/10.30658/hmc.2.2

Lou, C., Kang, H., & Tse, C. H. (2021). Bots vs. humans: How schema congruity, contingency-based interactivity, and sympathy influence consumer perceptions and patronage intentions. *International Journal of Advertising, 41*(4), 1–30. https://doi.org/10.1080/02650487.2021.1951510

Maisel, N., Gable, S. L., & Strachman, A. (2008). Responsive behaviors in good times and in bad. *Personal Relationships*, *15*(3), 317–338. https://doi.org/10.1111/j.1475-6811.2008.00201.x

Meyer von Wolff, R., Nörtemann, J., Hobert, S., & Schumann, M. (2020). Chatbots for the information acquisition at universities—A student's view on the application area. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, O.-C. Granmo, E. Luger, & P. B. Brandtzaeg (Chairs), *Chatbot Research and Design. CONVERSATIONS 2019, Lecture Notes in Computer Science* (vol. 11970, pp. 231–244). Springer. https://doi.org/10.1007/978-3-030-39540-7_16

Mori, M., MacDorman, K., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Mottarella, K. E., Fritzsche, B. A., & Cerabino, K. C. (2004). What do students want in advising? A policy capturing study. *NACADA Journal, 24*(1 & 2), 48–61. https://doi.org/10.12930/0271-9517-24.1-2.48

Mulac, A., Erlandson, K. T., Farrar, W. J., Hallett, J. S., Molloy, J. L., & Prescott, M. E. (1998). "Uh-huh. What's that all about?" *Communication Research*, *25*(6), 641–668. https://doi.org/10.1177/009365098025006004

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Neururer, M., Schlögl, S., Brinkschulte, L., & Groth, A. (2018). Perceptions on authenticity in chat bots. *Multimodal Technologies and Interaction*, *2*(3), 60. https://doi.org/10.3390/mti2030060

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*, 114. https://doi.org/10.3389/frobt.2018.00114

Prahl, A., & Van Swol, L. (2021). Out with the humans, in with the machines? Investigating the behavioral and psychological effects of replacing human advisors with a machine. *Human-Machine Communication*, *2,* 209–234. https://doi.org/10.30658/hmc.2.11

Reis, H. T., & Clark, M. S. (2013). Responsiveness. In J. A. Simpson & L. Campbell (Eds.), *Oxford library of psychology. The Oxford handbook of close relationships* (pp. 400–423). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195398694.013.0018

Roos, C. A., Postmes, T., & Koudenburg, N. (2023). Feeling heard: Operationalizing a key concept for social relations. *PLOS ONE*, *18*(11), e0292865. https://doi.org/10.1371/journal.pone.0292865

Rosseel, Y., Jorgensen, Terrence, D., & Rockwood, N. (2021). *Package "lavaan": Latent variable analysis* [computer software]. https://doi.org/10.18637/jss.v048.i02

Shawar, B., & Atwell, E. (2007). Chatbots: Are they really useful? *Journal for Language Technology and Computational Linguistics*, *22*(1), 29–49. https://doi.org/10.21248/jlcl.22.2007.88

Spence, P. R., Westerman, D., Edwards, C., & Edwards, A. (2014). Welcoming our robot overlords: Initial expectations about interaction with a robot. *Communication Research Reports*, *31*(3), 272–280. https://doi.org/10.1080/08824096.2014.924337

Sundar, S. S. (2008). The MAIN Model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73–100). MIT Press. https://betterlegalinfo.ca/wp-content/uploads/2019/12/Sundar-paper.pdf

Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human-AI interaction (HAII). *Journal of Computer-Mediated Communication*, *25*(1), 74–88. https://doi.org/10.1093/jcmc/zmz026

Sundar, S. S., Bellur, S., Oh, J., Jia, H., & Kim, H.-S. (2016). Theoretical importance of contingency in human-computer interaction. *Communication Research*, *43*(5), 595–625. https://doi.org/10.1177/0093650214534962

Sundar, S. S., & Nass, C. (2001). Conceptualizing sources in online news. *Journal of Communication*, *51*(1), 52–72. https://doi.org/10.1111/j.1460-2466.2001.tb02872.x

Van der Goot, M., & Etzrodt, K. (2023). Disentangling two fundamental paradigms in human-machine communication research: Media equation and media evocation. *Human-Machine Communication*, *6*, 17–30. https://doi.org/10.30658/hmc.6.2

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*(2), 203–225. https://doi.org/10.1108/JOSM-06-2019-0175

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, *11*(4), 342–365. https://doi.org/10.1287/isre.11.4.342.11872

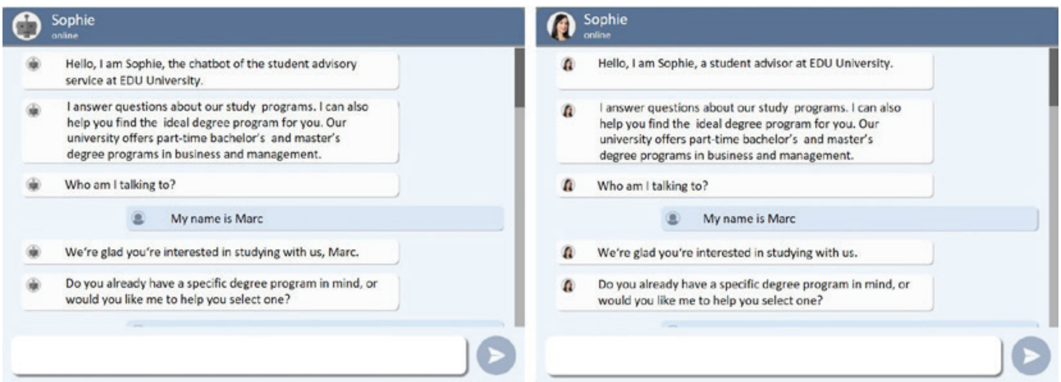
Westland, J. C. (2015). *Structural equation models: From paths to networks*. Springer. https://doi.org/10.1007/978-3-319-16507-3

Appendices

Appendix A—Power Analyses

The sample size of Study 1 ($N = 256$) was determined by an a priori power analysis ($f = .18$, $\alpha = .05$, $1-\beta = .80$) for an interaction effect resulting from an ANOVA. The power analysis was based on the small interaction effect ($\eta^2_{partial} = .03$) found by Go & Sundar (2019) in an experiment on the effects of identity and conversational cues on attitude toward the website where the chatbot is placed, using real interactions with chat agents. As we expected effects in Studies 2 and 3 to be smaller than Go and Sundar's interaction effect, we determined a sample size of $N = 403$ a priori for a significant interaction effect of $f = .14$ ($\alpha = .05$, $1-\beta = .80$).

Appendix B—Screenshots of Exemplary, Translated Chat Conversations in Chatbot and Human Conditions



Appendix C—English Translations of Items and Alpha Values for Variables Across Studies

Variable and Items	Cronbach's α in Study		
	1	2	3
Attitude 1. I find it attractive to communicate with an organization in this way. 2. I find it useful to communicate with an organization in this way. 3. I find this way of communicating with an organization interesting. 4. I find it helpful to communicate in this way with an organization. 5. Communicating in this way with an organization helps me to meet my needs.	.95	.95	.91
Likability Please rate your impression of Sophie on these scales: [dislike–like, unfriendly–friendly, unkind–kind, unpleasant–pleasant, awful–nice].	.92	.87	.90
Perceived intelligence Please rate your impression of Sophie on these scales: [incompetent–competent, ignorant–knowledgeable, irresponsible–responsible, unintelligent–intelligent, foolish–sensible].	.88	.90	.92
Warmth How [warm, trustworthy, friendly, honest, likable, sincere] do you think Sophie was?	.87	.88	.91
Competence How [competent, intelligent, skilled, efficient, assertive, confident] do you think Sophie was?	.88	.86	.91
Perceived enjoyment 1. The conversation evokes positive feelings in me. 2. I found the conversation entertaining. 3. I enjoyed reading the conversation.	.90	.89	.91
Satisfaction 1. I would be happy with Sophie's recommendations for courses of study. 2. I would be satisfied with the way Sophie spoke to Marc. 3. I would be satisfied with the information Sophie gave Marc. 4. I would be satisfied with the conversation Marc had with Sophie.	.92	.91	.94
Intention to use 1. If an organization offers this possibility of communication, I will use it. 2. If I have the opportunity to communicate with an organization in this way, I will. 3. I am very likely to use this way of communicating with an organization. 4. Once this way of communicating with an organization is established, it will be my preferred method.	.96	.96	.97

Variable and Items	Cronbach's α in Study		
	1	2	3
Social presence	.94	.94	.96
1. There was a sense of human contact in the interaction.			
2. There was a sense of personalness in the interaction.			
3. There was a feeling of sociability in the interaction.			
4. There was a feeling of human warmth in the interaction.			
5. There was a feeling of human sensitivity in the interaction.			
Perceived dialogue	.85	.82	.86
1. I had the feeling that Sophie was in an active dialogue with Marc.			
2. Marc's interactions with Sophie felt like a back-and-forth conversation.			
3. I felt that Sophie and Marc were involved in a joint task when choosing a program.			
4. Sophie was quick to respond to Marc's input and requests.			
5. I felt that Sophie took Marc's individual wishes into account.			
Feeling heard	.84	.82	.86
1. Marc felt heard.			
2. Marc was able to say what he really wanted to say.			
3. Sophie seemed to care more about something else than what Marc said.			
4. Sophie listened to Marc.			
5. Sophie tried to put herself in Marc's shoes.			
6. Sophie seemed insensitive to Marc's thoughts and feelings.			
7. Sophie treated Marc with respect.			
8. Sophie and Marc understood each other.			
Agent type manipulation check			
If you think back to the chat interaction you just saw: Who was Marc talking to?	—	—	—
1 = the study advisor Sophie, 2 = the professor Sophie, 3 = the chatbot Sophie, 4 = the doctor Sophie, 5 = don't know			
Responsiveness manipulation check	.77	.65	.59
1. Study 1: Sophie used affirmative expressions to indicate that she was really listening to Marc.			
2. Studies 2, 3: Sophie used affirmative expressions to indicate that she was listening to Marc.			
3. Study 1: Sophie appropriately picked up on what Marc said in her response.			
4. Studies 2, 3: Sophie picked up on what Marc said in her response.			

Note. 7-point Likert-type rating scales (1 = do not agree at all, 7 = fully agree), except likability, perceived intelligence (7-point semantic differentials) and the agent type manipulation check.

Appendix D—Means and Standard Deviations for Variables Across Studies

Variable	Study 1		Study 2		Study 3		Pooled Data	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Dependent Variables								
Attitude	4.79	1.58	4.18	1.67	4.76	1.61	4.53	1.65
Likability	5.68	1.08	5.53	1.06	5.72	0.96	5.63	1.04
Perceived intelligence	5.90	0.88	5.43	1.13	5.66	1.00	5.63	1.04
Warmth	5.35	0.99	5.01	1.11	5.29	1.11	5.19	1.09
Competence	5.73	0.90	5.43	0.98	5.54	1.06	5.54	1.00
Perceived enjoyment	4.29	1.42	3.78	1.51	4.55	1.50	4.18	1.52
Satisfaction	5.34	1.38	4.83	1.52	5.30	1.39	5.12	1.46
Intention to use	4.74	1.76	4.04	1.84	4.80	1.67	4.49	1.80
Mediators								
Social presence	3.80	1.52	3.58	1.51	4.35	1.52	3.91	1.55
Perceived dialogue	5.28	1.16	5.13	1.19	5.49	1.10	5.29	1.16
Feeling heard	5.43	0.92	5.30	0.95	5.48	0.97	5.39	0.95

Note. $N_1 = 253$, $N_2 = 401$, $N_3 = 351$, $N_{total} = 1{,}005$.

Appendix E—Bivariate Correlations Between Variables Across Studies

Variable	1	2	3	4	5	6	7	8	9	10	11
1. Attitude	—										
2. Likability	.39 .43 .49	—									
3. Perceived intelligence	.47 .38 .54	.70 .59 .73	—								
4. Warmth	.57 .56 .60	.69 .68 .74	.63 .52 .71	—							
5. Competence	.48 .50 .59	.44 .49 .64	.67 .54 .76	.66 .67 .80	—						
6. Perceived enjoyment	.61 .57 .65	.57 .46 .59	.49 .31 .56	.65 .61 .73	.45 .49 .60	—					
7. Satisfaction	.61 .65 .70	.58 .58 .63	.66 .52 .64	.72 .72 .76	.68 .68 .73	.61 .63 .71	—				
8. Intention to use	.89 .86 .91	.36 .38 .48	.41 .32 .50	.48 .49 .57	.44 .42 .54	.56 .54 .59	.53 .57 .63	—			
9. Social presence	.54 .51 .59	.51 .53 .52	.48 .42 .50	.68 .63 .67	.50 .50 .55	.71 .66 .70	.49 .46 .57	.57 .61 .59	—		
10. Perceived dialogue	.52 .49 .51	.56 .56 .61	.51 .46 .64	.73 .69 .72	.61 .59 .73	.55 .54 .59	.50 .43 .48	.71 .64 .68	.57 .63 .57	—	
11. Feeling heard	.52 .43 .46	.56 .55 .66	.53 .42 .62	.71 .67 .74	.59 .58 .72	.46 .47 .53	.45 .36 .42	.67 .59 .66	.53 .41 .51	.74 .79 .79	—

Note. Pearson's correlations *r*. Grey shaded cells: 1st line = Study 1 ($N = 253$), 2nd line = Study 2 ($N = 401$), 3rd line = Study 3 ($N = 351$). All correlations are significant at $p < .001$.

HUMAN-MACHINE
COMMUNICATION