Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

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

I just wanna blame somebody, not something! Reactions to a computer agent giving negative feedback based on the instructions of a person

Aike C. Horstmann [a,*], Jonathan Gratch [b], Nicole C. Krämer [a]

[a] *Social Psychology: Media and Communication, University of Duisburg-Essen, Duisburg, Germany*
[b] *Institute for Creative Technologies, University of Southern California, Playa Vista, CA, USA*

ARTICLE INFO

Keywords:
Human-agent interaction
Automated agents
Agent representatives
Agency
Blame
Attribution theory

ABSTRACT

Previous research focused on differences between interacting with a person-controlled avatar and a computer-controlled virtual agent. This study however examines an aspiring form of technology called agent representative which constitutes a mix of the former two interaction partner types since it is a computer agent which was previously instructed by a person to take over a task on the person's behalf. In an experimental lab study with a 2 × 3 between-subjects-design (N = 195), people believed to study either together with an agent representative, avatar, or virtual agent. The interaction partner was described to either possess high or low expertise, while always giving negative feedback regarding the participant's performance. Results show small but interesting differences regarding the type of agency. People attributed the most agency and blame to the person(s) behind the software and reported the most negative affect when interacting with an avatar, which was less the case for a person's agent representative and the least for a virtual agent. Level of expertise had no significant effect and other evaluation measures were not affected.

1. Introduction

Software systems are growing in sophistication and autonomy (Hancock, 2017) allowing users to delegate complex tasks that heretofore would have required human oversight (Gogoll and Uhl, 2018; Mosier et al., 1997). Against this background, a new field of application receives growing attention – computer agents that act autonomously on the behalf of their users, representing their user's interests in interactions with other people (de Melo et al., 2018). These so-called agent representatives constitute a new combination of computer- and person-controlled interaction partner. Their actions are controlled by a computer program, not a person, but follow previously given instructions to represent a person in a certain situation. An example would be Google Duplex, a phone call system which schedules appointments on behalf of a person using a natural human-sounding voice (Leviathan and Matias, 2018). Agent representatives potentially lead to a blurring of borders between the categories person and computer, which this study aims to shed further light on.

Previous research in this field has predominantly focused on the different effects of person-controlled avatars compared to computer-controlled virtual agents (Blascovich et al., 2002; Gallagher et al., 2002; Gazzola et al., 2007; Lucas et al., 2014; von der Pütten et al., 2010). With agent representatives, which constitute a mix of computer and human agency, new research questions arise. For example, whether people instruct their representatives to act differently than they themselves might act (de Melo et al., 2018; Mell et al., 2018). One question that has not been sufficiently addressed yet is how people assign blame when the feel mistreated by another's agent representative, and particularly whether they differentiate between the instructing person and the agent following the person's instructions.

Attributions of blame are shaped by a variety of factors including whether the actor has full agency over a potentially blameworthy act (Kelley and Michela, 1980). In case of an avatar, which is controlled in real time by a person who possesses full agency and acts intention-driven, the attribution of blame should clearly be targeted at the person in control of the avatar. Agent representatives, however, function as third party-representatives acting on a certain person's previously given instructions which may lead to a deflection of blame (Bivins, 2006; Royzman and Baron, 2002). In contrast to other virtual agents, a specific person with agency and intentions gave instructions to

This manuscript features online supplementary material (OSM), which includes the data of the study, descriptive values of the dependent variables, descriptions of the experimental manipulations, and the interaction script. The OSM can be accessed here: https://osf.io/uc6s3/

[*] Corresponding author.
E-mail address: aike.horstmann@uni-due.de (A.C. Horstmann).

https://doi.org/10.1016/j.ijhcs.2021.102683
Received 10 November 2020; Received in revised form 31 May 2021; Accepted 7 June 2021
Available online 17 June 2021

an agent representative so that it may represent this person and their intentions. Thus, the behavior should be perceived as more intentional, even though transferred, compared to that of a classic virtual agent which is solely controlled by a computer program. The aim of this study is therefore to examine whether and to what extend blame is attributed differently when interacting with an impolitely behaving avatar, virtual agent, or agent representative.

In addition to the agency of the interaction partner, also their expertise may substantially influence how the content is perceived and how people react to it (Bannister, 1986; Graefe et al., 2018). Someone with a similar level of expertise might be perceived as an ally (Nass et al., 1996; Wilson et al., 1965) from whom impolite, negative feedback might be perceived as hostile. Since people generally appear to be more convinced by and receptive to a source with high expertise, negative feedback by an expert in a tutoring role is probably perceived more appropriate and thus more positive than coming from a source with low expertise (Berlo et al., 1969; Dijks et al., 2018; Fogg, 2002; Hovland et al., 1953; Krämer et al., 2017). In general, these insights are adapted by attributing high education or the role of a teacher to computers to make them more influential (Fogg, 2002). Besides this conscious attempts to enhance perceived expertise, artificial entities are generally presumed to be highly expert, credible, and authoritative (Burgoon et al., 2000; Fogg and Tseng, 1999; Graefe et al., 2018; Horstmann and Krämer, 2019; Nourani et al., 2020). Therefore, the interaction partner's type of agency (avatar, virtual agent, or agent representative) may interact with its level of expertise, which is why expertise is further included as an influencing factor in this study.

1.1. Research aim

Examining the circumstances of attribution processes in human-computer interaction will lead us to a deepened understanding of how people perceive and react to machines which increasingly blur the borders between humans and machines. This ultimately holds valuable insights for the design and potential future applications of these kinds of technologies. Against this background, the overarching research aim of this study is to examine whether impolite, negative feedback is perceived differently when presented by a computer-controlled but person-instructed agent representative compared to a person-controlled avatar and a computer-controlled virtual agent. Particularly, the question is how blame is attributed in interactions with agent representatives which may cause a greater blurring of boundaries between the categories human and computer compared to the other two types of agency. Furthermore, level of expertise is considered as a potential influencing factor given that computers are typically perceived as experts.

2. Theoretical background

2.1. Type of agency: Person-controlled avatar vs. computer-controlled virtual agent

An extensive body of research is concerned with the differences and similarities of interacting with a human versus a computer interaction partner (Krämer et al., 2012). This is underlined by a bill that has gone into effect in California in 2019 stating that people need to be informed about whether they are communicating with an artificial or human identity (S.B.-1001 Bots: disclosure, 2019). On the one hand, several studies showed that people react in a fundamentally social way towards interactive media (Nass and Moon, 2000), e.g., by showing politeness (Hoffmann et al., 2009; Nass et al., 1999) or responding to flattery (Fogg and Nass, 1997). This phenomenon of people applying social norms when interacting with media is described by the media equation theory (Reeves and Nass, 1996) and was tested with interactive computers (Nass and Moon, 2000), smartphones (Carolus et al., 2019), robots (Eyssel and Hegel, 2012; Horstmann et al., 2018; Lee et al., 2005), and virtual agents (Hoffmann et al., 2009; Kang et al., 2008). Research further showed that telling people that they will be interacting with an avatar (defined as virtual representation of a human) or agent (described as autonomous computer-controlled agent) leads to few or no significant evaluation or behavioral differences (Krämer et al., 2017; von der Pütten et al., 2010). Generally, interactions with humans and computers share many similarities. Accordingly, theories from human-human interactions may work as valuable framework when examining interactions with artificial entities (Krämer et al., 2012).

On the other hand, there are also differences between people's interactions with machines and humans (Krämer et al., 2012), which several studies show (Bartneck et al., 2005; Blascovich et al., 2002; Gallagher et al., 2002; Gazzola et al., 2007; Lucas et al., 2014; Rosenthal-von der Pütten et al., 2014). According to Blascovich (2002), people only respond socially to a person or an avatar controlled by a person. A virtual agent controlled by a computer would not elicit social responses unless the agent is not distinguishable from an avatar (Blascovich, 2002).

According to the attribution theory, people interpret others' behavior in terms of its causes, which then affects people's reactions towards it (Kelley and Michela, 1980). Causes can be divided into internal causes which are located "inside" the person (e.g., attitudes and dispositions) and external causes which are located "outside" the person (e.g., peer pressure and situational forces; Kelley, 1973; Kelley and Michela, 1980). Responsibility for a behavior is further connected to agency – when a person possesses adequate agency to do something, then this person is responsible for the outcomes of their actions (Coeckelbergh, 2020). However, other factors such as voluntariness, foreknowledge, and intention need to be considered as well (Coeckelbergh, 2020; Fischer and Ravizza, 1998; Mao and Gratch, 2003; Weiner, 1995). Against this background, the question arises whether people react differently to their interaction partner's actions when they are aware that they are interacting with a computer program compared to a person. To examine those differences is pivotal since artificial interaction partners are becoming increasingly prevalent. Due to the broad range of design variations of artificial entities' appearance and, even more so, behavior, research is needed to better understand users' reactions to different forms of artificial entities in order to provide valuable guidelines for designers and developers.

For instance, results of a meta-analysis by Fox et al. (2015) lead to the conclusion that in social situations avatars are more influential than agents. In this vein, people performed significantly worse on a task in the presence of human-controlled avatars compared with computer-controlled agents (Blascovich et al., 2002). In a different study, people reported more comfort negotiating with a computer program which was evaluated more cooperative and punished less than a human opponent (Gratch et al., 2016). This research suggests that challenging situations are perceived more negatively with a person-controlled avatar compared to a computer-controlled agent present (Blascovich et al., 2002; Gratch et al., 2016). Since computer agents usually act in a very polite way (Sayin and Krishna, 2019), there is not much knowledge about how people would perceive and react to an impolitely acting one. And more particularly, how the perception and reaction would differ from or resemble how people react to an impolite person.

In general, it is a fundamental human need to form interpersonal attachments and not to be rejected or excluded by others (Baumeister and Leary, 1995). This leads to the assumption that receiving harsh, negative feedback from another person versus a computer agent has more detrimental effects on a person's state of mood. In line with that, feelings of embarrassment are also enhanced when interacting with a person via their avatar and diminished when interacting with a computer agent (Bartneck et al., 2010; Choi et al., 2014). Furthermore, volition and intentions are rather attributed to humans than machines which are perceived as being restricted to their programming (Banks, 2019; Bigman and Gray, 2018; Malle and Knobe, 1997). According to the attribution theory, aggressive negative actions need to be perceived as being executed with intention in order for the actors to be blamed for them (Kelley, 1973; Kelley and Michela, 1980; Malle et al., 2001; Malle and Knobe, 1997). As a consequence, we assume that people rather attribute blame to a human-controlled than to a computer-controlled interaction partner. Against this background, the following hypothesis is formulated:

H1. *An avatar leads to a) a greater sense of agency, b) a greater amount of blame attribution, and c) a more negative affect compared to a virtual agent when giving negative feedback.*

2.2. *Type of agency: Agent representative vs. avatar and virtual agent*

With the immense advancement of virtual agent technologies becoming increasingly reliable as well as autonomous (Hancock, 2017), a new hybrid form of virtual interaction partner is emerging. An agent representative is an autonomous computer agent which can be instructed to represent a certain person in a certain situation. On several occasions people may need representation. This could be a lawyer, a real estate broker, or just another person who conveys their interests to others when they cannot be present themselves (Mell et al., 2018). Agent representatives are supposed to take over these tasks instead of humans. They act autonomously in the situation in a sense that they are not controlled in real-time by the person they are representing. However, they follow previously given instructions by this person on how to act on this person's behalf, i.e., in accordance with the person's intentions, motivations, beliefs, and attitudes. Unlike an avatar, which can be defined as real-time representation of a person, an agent representative would not reproduce the actual behavior of a person in that moment. Instead it acts autonomously in the interest of a person based on their previously given instructions. Of course, every virtual agent was programmed by a person or a group of persons at some point and here research showed that people do not think of the programmer when interacting with a computer (Nass and Moon, 2000; Sundar and Nass, 2000). However, with an agent representative the clear emphasis lays on representing a certain person and their intentions. This is different to the presence of some abstract, distant programmer. An agent representative will be introduced as representing a specific person, which should put more focus on this person compared to the usual programmer.

Popular examples for agent representatives are automated phone call or bidding systems (well-known from the Internet auction platform eBay) as well as automated negotiators and self-driving cars (de Melo et al., 2018). An agent representative can help people save time (de Melo et al., 2016) and, unlike a human, does not get tired or bored (Fox et al., 2015). No personal beliefs, attitudes, intentions, or norms would interfere with the agent's behavior, which should result in a controllable and reliable representative.

When people start to interact through an agent representative software with others, the borders between human and computer interaction partner become even blurrier. Against this background, the question arises whether the agent representative is perceived to possess agency and whether it or the person who instructed it is perceived responsible for its behavior. To tackle this question before introducing this technology on a large scale is crucial since it entails relevant questions of accountability and liability. This gains even greater relevance in case of unpleasant or undesirable behavior. Virtual agents often occupy service positions and thus are expected to comply to social norms like politeness (Sayin and Krishna, 2019). However, in case of the representation situation, a person could also instruct the agent to be tough, rude, or even mean. In accordance with the attribution theory (Kelley, 1973; Kelley and Michela, 1980), the instructing person should be blamed for any negative interactional behavior of the agent since the agent representative only acts upon this person's behalf and naturally has no own will.

Here, it needs to be considered that the agent representative functions as intermediary and thus attribution of agency and intentionality might not be as clear as with an avatar. An avatar immediately replicates the person's intentions in real-time. In contrast, agent representatives only transfer a person's intentions as far as they were instructed. Thus, there is a time delay and possibly not all situational aspects are considered. As a consequence, blame for negative behavior may be deflected (Bivins, 2006; Royzman and Baron, 2002).

Nevertheless, an agent representative should still be perceived to act with more agency and to deserve more attribution of blame in case of negative behavior than a virtual agent. With an agent representative the feedback would indirectly come from a person and people dislike hostility by others (Baumeister and Leary, 1995). Thus, when people take into account that someone had time to think about what to say and how to act through the agent representative, they probably attribute more intention followed by blame to it compared to when their interaction partner's behavior is determined by some abstract, autonomous computer program. Based on these theoretical deliberations, the following is hypothesized:

H2. *An avatar compared to an agent representative and an agent representative compared to a virtual agent leads to a) a greater sense of agency, b) a greater amount of blame attribution, and c) a more negative affect when giving negative feedback.*

2.3. *Level of expertise: High vs. low expertise*

Negative feedback is perceived differently depending on the source's characteristics, such as expertness, credibility (Hovland et al., 1953), authority (Fogg, 2002), and qualifications (Berlo et al., 1969). Whether a person is perceived as qualified, credible source depends on whether they are portrayed, for instance, as trained, experienced, important, educated, or expert (Berlo et al., 1969). In general, expertness is seen as a very important determent for source credibility (Fisher et al., 1979; Graefe et al., 2018; Hovland et al., 1953). A study by Bannister (1986) further showed that people are more satisfied with negative feedback provided by an experienced and knowledgeable person than a young and unexperienced one (Bannister, 1986). This goes in line with findings by Hovland et al. (1953), which show that feedback from a source with high expertness and trustworthiness is evaluated more favorable and also has a stronger persuasive effect on people. Nass et al. (1996) were able to show that even electronic devices labeled as specialists are perceived more credible than devices labeled as generalists. Thus, to make computers more influential, names are chosen that suggests higher education or an authoritarian role (e.g., "WinDoctor"; Fogg, 2002).

Other research has focused on using computer programs in form of peer learning agents with similarly low expertise to promote collaborative learning (Chan and Baskin, 1988; Kersey et al., 2010; Vizcaíno, 2005). A peer study partner with a low level of expertise should be expected to be less authoritarian (Fogg, 2002) and to rather support their partner in acquiring more knowledge together (Cha et al., 2014; Nass et al., 1996; Wilson et al., 1965). Consequently, people will likely expect cooperative and not degrading behavior from a study partner with a low level of expertise. Solely harsh, negative feedback should then negatively violate these expectations, causing detrimental communication and relationship outcomes (Burgoon and Hale, 1988).

Summing up, negative feedback provided by a tutor with a high level of expertise should be more convincing as well as acceptable than provided by a peer study partner with a low level of expertise (Bannister, 1986; Hovland et al., 1953). Consequently, less blame should be attributed to the expert since the expert's feedback is more convincible and thus rather accepted as being true and appropriate than the study partner's feedback (Fisher et al., 1979; Hovland et al., 1953). Moreover, in contrast to the expert tutor, the peer study partner could be expected to help and support the participant like a team member (Nass et al., 1996; Wilson et al., 1965). When this expectation is violated with mean and non-supportive comments regarding the participant's performance (Burgoon and Hale, 1988), this should negatively affect people's mood. This should be less the case with an expert interaction partner.

Consequently, the following hypothesis is formulated:

H3. *A low expert peer study partner leads to a) a greater amount of blame attribution and b) more negative affect compared to a high expert tutor when giving negative feedback.*

2.4. Interaction effects of type of agency and level of expertise

The level of expertise might play a greater role when a person compared to a computer is involved in the interaction. Modern interaction technologies are generally expected to be performance-oriented (e.g., efficient, reliable, and precise; Arras and Cerqui, 2005; Ezer et al., 2009) and perceived as more credible, expert (Graefe et al., 2018), and influential (Burgoon et al., 2000) than human partners. People appear to attribute credibility to media by conferring impressions of authoritativeness and expertise on them (Burgoon et al., 2000). In other words, people presume that artificial entities have expertise and are trustworthy (Fogg and Tseng, 1999; Graefe et al., 2018; Horstmann and Krämer, 2019; Nourani et al., 2020). Moreover, the human peer study partner probably rather resembles an ally than the expert or the computer interaction partners. A human expert or computer program might be perceived to just react to poor performance on a professional level. Harsh feedback by a human peer study partner might feel more like a personal attack, maybe even betrayal. Thus, when a person which is expected to have a low level of expertise and to help as an equal study partner gives negative feedback, this might lead to a greater attribution of blame and more negative affect than a computer-controlled or expert interaction partner. Therefore, to examine possible interaction effects of the type of agency and level of expertise, the following hypothesis is postulated:

H4. *An avatar controlled by a person with a low level of expertise leads to a) a greater attribution of blame and b) more negative affect compared to expert or computer-controlled interaction partners when giving negative feedback.*

In addition to state of mood and blame attribution, the evaluation of the interaction partner and the interaction in general as well as further contact intentions should also be affected by the interaction partner's type of agency and level of expertise. Negative feedback from a person, in contrast to a computer program, should rather be perceived as social and/or relational hostility (Baumeister and Leary, 1995). The feedback should further be more accepted coming from a tutor with high expertise than a peer study partner with low expertise (Bannister, 1986; Hovland et al., 1953). Since computers are generally perceived as credible and expert (Fogg and Tseng, 1999; Graefe et al., 2018; Nourani et al., 2020), type of agency and level of expertise should interact with each other. A person with low expertise providing feedback through an avatar should be evaluated worse than a) an expert perceived to act merely professionally, b) an agent representative with psychological distance to the person who instructed the agent (de Melo et al., 2018), and c) a virtual agent which acts upon its programming. Consequently, the following is hypothesized:

H5. *An avatar controlled by a person with a low level of expertise leads to a more negative evaluation of a) the interaction partner and b) the interaction as well as to c) less future contact intentions compared to expert or computer-controlled interaction partners.*

3. Method

An experimental 3 (type of agency: avatar vs. virtual agent vs. agent representative) x 2 (level of expertise: high vs. low) between-subjects design was applied for this laboratory study. Participants were randomly assigned to one of the six conditions. The local ethics committee approved the study and written informed consent was obtained.

3.1. Sample

Results of an a priori power analysis using G*power 3.1 software (based on 95% power and a medium effect size of $f^2 = 0.15$; Cohen, 1988; Wullenkord et al., 2016) recommended a sample size of 194 participants. In total, 201 individuals participated in the study, 98 in the USA and 103 in Germany. Six of the data sets had to be excluded from all calculations because the respective subjects failed both manipulation checks and generally showed severe signs of inattention. Of the remaining 195 participants, 94 participated in the study in the USA and 101 in Germany. Comparing the US and the German sample, there were no significant differences regarding the dependent variables (except the interaction partner's social attractiveness, $F(1, 193) = 4.80$, $p = .030$, $\eta_p^{2=}.02$). Consequently, the two subsamples are combined to one and the country was not included as additional factor for the following analyses.

102 participants were male and 93 were female with an average age of 34.80 years ($SD = 13.16$; range: 18 to 70). Regarding education, most participants hold a high school or equivalent degree (76; 39%), a bachelor's degree (70; 35.9%), or a master's degree (27; 13.8%). Most participants further stated either to be a student (72; 36.9%), employed full-time (52; 26.7%) or part-time (32; 16.4%), unemployed (14; 7.2%), or self-employed (10; 5.1%).

3.2. Experimental procedure

First, a cover story was presented to the participants which explained that the study's purpose was to examine and compare the effectiveness of two different learning methods - one would be tutorial YouTube videos, the other one classic school textbooks. All participants were further told that they were randomly assigned to the tutorial video condition and that they will be given time to study, either with a high or low expertise study partner, before their knowledge will be tested in a final exam. As an incentive, participants were told that the person with the highest exam score will receive a monetary price. The study time with a study partner was justified by the study's aim to replicate an actual learning situation as closely as possible. It was argued that people often study with a tutor (high expertise) or peer (low expertise) to prepare for an important exam. It was further explained that to have a person present at the lab for all the experiments would however cost a lot of time and money, which is why an avatar/virtual agent/agent representative software was used. Please see the online supplementary material for more details regarding the experimental manipulations of the level of expertise and type of agency.

After the cover story was explained, written informed consent was obtained and participants were asked to read another description of the study's aim and procedure. Thus, the cover story was presented to the participants orally as well as in written form. Participants in the agent representative condition were further asked to complete one small task before starting with the actual experiment to reinforce the manipulation. Allegedly, since the agent representative software is still in development and data is always needed, it would help if all participants would enter a few instructions for an agent representative. They were asked to imagine a study situation, where they would send an agent representative in their place. Particularly, they were asked to enter three exemplary sentences for the agent representative to say to their student/study partner along with instructions when to say it so that they would feel adequately represented.

All participants then started with some pre-questionnaires assessing their demographical and technological background, followed by the tutorial video about the processes of the water cycle. The experimenter left the room to avoid distractions and participants were asked to ring a bell after they watched the video. The experimenter then came back and switched on a TV, via which the study session with their interaction partner Brad (Fig. 1) would take place. At this point it was made sure that participants were aware with whom or what they were going to be interacting with (avatar, virtual agent, or agent representative as well as

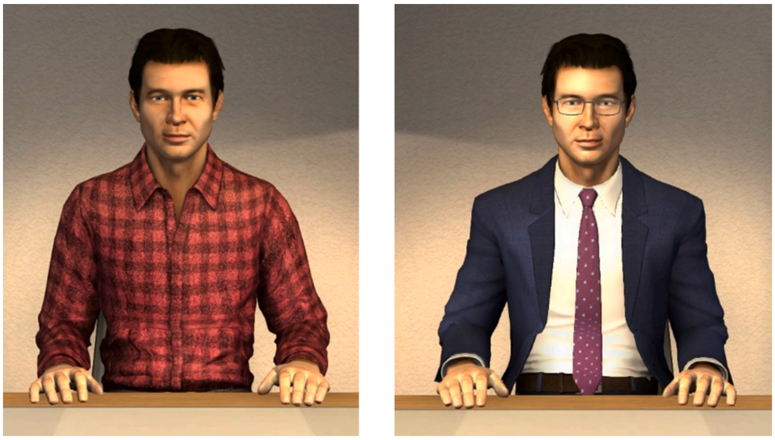

Fig. 1. Left: low expertise peer study partner; right: high expertise tutor.

high or low expertise study partner). For this purpose, they were asked to explain to the experimenter in their own words what they were told about their interaction partner. If necessary, participants were corrected so that all of them had a clear understanding. After the manipulations were reinforced this way, it was explained that the interaction partner would initiate the conversation. During the interaction, the virtual agent Brad was controlled by the experimenter. By using a webcam installed above the TV, the experimenter was able to see and hear the participant in order to let Brad respond accordingly (wizard of oz design; see Dahlbäck et al., 1993).

The interaction started with Brad introducing himself and explaining that this would be like a practice quiz. Brad then asked 30 questions regarding the content of the tutorial video. Based on the results of a pretest, questions were chosen that were purposely very hard to answer as well as open-ended and vague. This way, participants would have a hard time being certain whether their answer is completely right. For instance, it was asked about minor details or a listing of aspects mentioned in the video (e.g., “What is the name of the aquifer mentioned in the video?”; “How many and what kinds of movements of water are mentioned in the video?”). After every two or three questions, Brad gave some harsh negative feedback regarding the participant’s general performance, regardless of how the participant answered (e.g., “You still need a lot of practice. Right now, you have no chance compared to the others.”). Whenever a participant took a very long time to answer or just stopped answering, Brad would say “Just answer the question”. After Brad was done with all questions, he stated “Okay. Well, I’m done with my preparation questions. This was a torture. Good luck with the final exams. You will need it. Bye.”. The complete interaction script can be viewed in the online supplementary material. After the interaction with Brad, a different computer voice told the participant to go back to the laptop to continue with the final exam and afterwards some questionnaires rating Brad and their interaction with Brad. After participants rang the bell at the end, the experimenter returned, debriefed the participants, and compensated their time either with course credits or money.

3.3. *Measurements*

All self-constructed as well as adapted scales and items can be found in the online supplementary material.

3.3.1. *Sense of agency*

The interaction partner’s perceived agency was assessed with the Sense of Agency Scale (Tapal et al., 2017; 11 items; e.g., “Brad is in full control of what Brad does.”; 1 = “strongly disagree” to 5 = “strongly agree”; α = 0.79)

3.3.2. *Attribution of blame*

It was assessed to what extent (1 = “strongly disagree” to 5 = “strongly agree”) participants attribute blame to themselves (2 items; e.g., “I am responsible for the mistakes I made during the practice test.”; $\rho = 0.41$), to the software (2 items; e.g., “The avatar/virtual agent/agent representative software is to blame for the mistakes I made during the practice test.”; $\rho = 0.54$), or to the person(s) behind the software (2 items; e.g., “I would not have made so many mistakes during the practice test if it was not for the person(s) behind the avatar/virtual agent/agent representative software.”; $\rho = 0.72$). Person(s) behind the software was not further defined for the participants so that they could apply this to the person(s) they may have suspected to have an influence on their interaction partner’s behavior.

3.3.3. *State of mood*

Participants reported on their emotional state right after the interaction using the Positive Affect Negative Affect Scale (PANAS; Watson et al., 1988; 1 = “very slightly or not at all” to 5 = “extremely”), which is divided into measuring positive affect (10 items; e.g., “excited”; α = 0.89) and negative affect (10 items; e.g., “distressed”; α = 0.85).

3.3.4. *Evaluation of the interaction partner Brad*

Participants answered the task attractiveness subscale (5 items; e.g., “Brad would be a poor problem solver.”; α = 0.80) and the social attractiveness subscale (5 items; e.g., “I think Brad could be a friend of mine.”; α = 0.81) of the Interpersonal Attractiveness Scale (McCroskey and McCain, 1974; 1 = „strongly disagree“ to 5 = „strongly agree“). Moreover, 27 items subtracted from several person, robot, and agent evaluation scales (Bartneck et al., 2009; Bente et al., 1996; Carpinella et al., 2017; Fogg and Tseng, 1999; Lea and Spears, 1992; McCroskey and Young, 1981; von der Pütten et al., 2010) were used to have participants evaluate their interaction partner’s likeability (e.g., “cold – warm”; α = 0.94), competence (e.g., “incapable – capable”; α = 0.89), and human-likeness (e.g., “unemotional – emotional”; α = 0.51) on a 5-point semantical differential. The theoretical constructs were verified via factor analysis.

3.3.5. *Interaction evaluation*

Adapted versions of the Evaluation (4 items; e.g., “I was enjoying the interaction with Brad.”; α = 0.84) and Expectedness (4 items; e.g., “The behavior of Brad was as I expected it to be.”; α = 0.79) subscales were used based on Burgoon and Walther (1990; 1 = „strongly disagree“ to 5 = „strongly agree“). Two self-constructed items were added to assess the perceived appropriateness of behavior (e.g., “Brad is behaving in a way that fits the situation”; α = 0.70).

3.3.6. *Contact intentions*

Participants were asked to what extent they would like to interact with their interaction partner again in the future (Eyssel et al., 2011; e. g., “I would like to talk to Brad more.”; 1 = “strongly disagree” to 5=“strongly agree”; α = 0.95).

3.3.7. *Further assessments*

Participants' age, sex, educational level, current employment or training status, and race(s) they identify with were assessed. Furthermore, people's locus of control when using technology (Beier, 1999), their technical affinity (Karrer et al., 2009), perceived psychological safety (Baer and Frese, 2003; Edmondson, 1999), how helpful they perceived the feedback (Krämer et al., 2017), who they would prefer as future study partner, and feedback for their interaction partner were assessed but not analyzed for this paper.

3.3.8. *Manipulation checks*

At the end of the questionnaire, participants were asked to state via free text input with what kind of interlocutor they were told to be interacting with and what kind of expertise they were told that their interlocutor has. Then, participants were asked again via forced-choice in two steps: first, whether they were told to be interacting with another person, visually represented by an avatar, or a computer program. In case of a computer program, they were further asked whether an agent representative or virtual agent software was described to them. For level of expertise, participants chose between tutor with high expertise or peer study partner with a low expertise. Participants always had the option to choose “I would have to guess”.

4. Results

Post-hoc power analyses were computed using G*Power 3.1 with the actual sample size of 195 subjects. The statistical power for this study was 0.41 for detecting a small effect ($f^2 = .02$) and exceeded 0.99 for a medium effect ($f^2 = .15$) as well as for a large effect ($f^2 = .35$; Cohen, 1988). Consequently, more than adequate statistical power was reached for medium to large effect sizes, but less than adequate power for a small effect size. Extensive descriptive statistics can be found in the online supplementary material.

4.1. *Manipulation checks*

Regarding the type of agency, in a first step, 185 participants were able to recall correctly whether it was a person-controlled avatar or computer-controlled agent they were told to be interacting with, 15 recalled wrong. In the second step, another 11 participants were not able to name the exact correct type of computer-controlled agent, i.e., whether it was a virtual agent or an agent representative. Regarding expertise, 188 participants reported the right level of expertise (high or low), while 12 failed to name the right level.

4.2. *Sense of agency, attribution of blame and negative affect*

To test hypotheses H1 to H4, a multivariate analysis of variance (MANOVA) was calculated with type of agency and level of expertise as factors and agency, blame attribution, and affect as dependent variables. The first hypothesis postulates that a virtual agent giving negative feedback leads to a) lower perceived agency, b) lower amount of blame attribution, and c) a less negative affect compared to an avatar. According to hypothesis H2, an avatar compared to an agent representative and an agent representative compared to a virtual agent causes a) lower perceived agency, b) lower amount of blame attribution, and c) less negative affect when giving negative feedback. Using Pillai's trace, there was a significant effect of *type of agency* on sense of agency, blame attribution, and affect, $V = 0.13$, $F(12, 370) = 2.08$, $p = .017$.

For hypothesis H1, multiple comparisons with Turkey's HSD test revealed on a marginally significant level that interacting with an avatar compared to a virtual agent leads to a greater sense of agency, $p = .058$, and a greater amount of blame attributed to the person(s) behind the software, $p = .076$ (see Table 1). However, pervading all conditions, the descriptive values for attributing blame to oneself are notably higher than for attributing blame to the software or to the person(s) behind the software (see Table 1). Furthermore, the avatar causes significantly more negative affect than the virtual agent, $p = .019$. Please see Fig. 2 for an overview of the relevant results regarding the interaction partner's type of agency. Based on the results, hypothesis H1a, H1b, and H1c are partly supported.

For hypothesis H2, the multiple comparisons with Turkey's HSD test showed that an avatar compared to a person's agent representative leads to a significantly greater sense of agency, $p = .017$, no significant difference regarding the attribution of blame to the person(s) behind the software, $p = .329$, and on a marginally significant level to a more negative affect, $p = .080$ (see Table 1 and Fig. 2). Between agent representative and virtual agent, no significant differences were found regarding agency, $p = .898$, attribution of blame, $p = .726$, and negative affect, $p = .835$. Consequently, H2a is partly supported, H2b is not supported, and H2c is partly supported.

Hypothesis H3 considers an effect of level of expertise whereby a high expertise is assumed to lead to a) a lower amount of blame attributed to the interaction partner and b) less negative affect compared to low expertise. According to Pillai's trace, there was no significant effect of level of expertise on attribution of blame and affect, $V = 0.04$, $F(6, 184) = 1.41$, $p = .214$. Therefore, H3 a) and b) are not supported.

According to hypothesis H4, a higher type of agency combined with a lower level of expertise causes a) a higher amount of blame attributed to the interaction partner and b) more negative affect compared to a low type of agency and low level of expertise. Pillai's trace also shows no significant effect of the interaction of type of agency and level of expertise on attribution of blame and affect, $V = 0.08$, $F(12, 370) = 1.30$, $p = .216$. This leads to the conclusion that H4 a) and b) need to be rejected.

4.3. *Interaction partner and interaction evaluation as well as future contact intentions*

The fifth and last hypothesis (H5) assumes an influence of the interaction partner's type of agency and level of expertise on participant's evaluation of a) the interaction partner and b) the interaction as well as on c) their intentions to have further contact. To test this hypothesis, another MANOVA was calculated. Using Pillai's trace, there was no significant main effect of the type of agency, $V = 0.09$, $F(16, 366) = 1.08$, $p = .375$, and level of expertise, $V = 0.02$, $F(8, 182) = 0.54$, $p = .829$, as well as no interaction effect of type of agency and level of expertise, $V = 0.08$, $F(16, 366) = 0.96$, $p = .505$, on the evaluation of the interaction partner Brad, the general evaluation of the interaction with Brad, and future contact intentions. Consequently, H5a, H5b and H5c

Table 1
Descriptive statistics of the main dependent variables by type of agency.

	Type of agency					
	Avatar (*N* = 66)		Virtual agent (*N* = 64)		Agent representative (*N* = 65)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Agency						
Sense of agency	2.73	0.76	2.44	0.77	2.38	0.61
Attribution of blame						
Self	3.53	0.98	3.73	0.93	3.82	0.94
Software	2.35	1.14	2.10	0.96	2.05	0.91
Person(s) behind software	2.44	1.21	2.02	0.95	2.17	1.05
Affect						
Positive affect	2.58	0.84	2.79	0.89	2.86	0.82
Negative affect	3.06	0.87	2.65	0.77	2.74	0.89

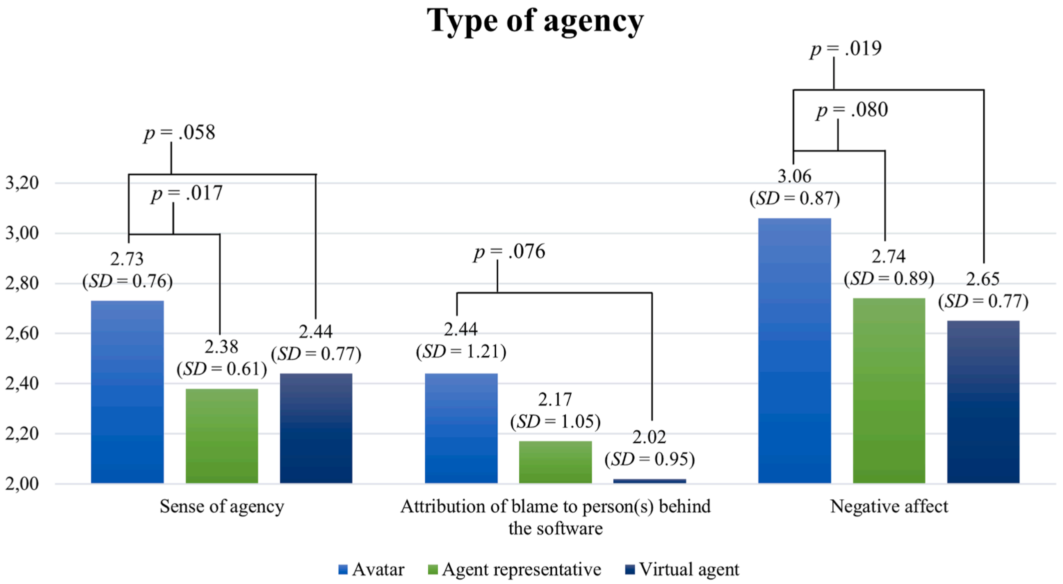

Fig. 2. Sense of agency, blame attribution to person(s) behind the software, and negative affect divided by type of agency.

are not supported.

4.4. *Additional analyses of facial expressivity*

For a broader understanding, videos of the US subsample were analyzed for facial expressivity using the computer-based video classification algorithm FACET. Means for overall amount of expressivity were compared between conditions. There was no significant main effect of type of agency, $F(2, 87) = 0.19, p = 830, \eta_p^2 = .00$, nor of the level of expertise, $F(1, 87) = 0.06$, $p = .807$, $\eta_p^2 = .00$. There was also no interaction effect of type of agency and level of expertise, $F(2, 87) = 1.60$, $p = .208$, $\eta_p^2 = .04$.

5. Discussion

More and more tasks are delegated to machines to support and relieve humans (Gogoll and Uhl, 2018; Mosier et al., 1997), e.g., in form of automated negotiator (de Melo et al., 2018), phone call, and bidding systems. These autonomous technologies act on behalf of a human and thus constitute a mix of a computer-controlled and person-instructed agent. Therefore, research on how people perceive and react to this form of technology is highly relevant. This study explored whether and how people's perceptions and reactions differ when they are supposedly interacting with an avatar, virtual agent, or agent representative with an either high or low level of expertise. This was examined by exploring differences in reactions towards negative feedback in form of perceived agency, attribution of blame, and negative affect.

5.1. *Type of agency: Avatar vs. virtual agent*

Findings of various studies indicate differences between interactions with person-controlled avatars and computer-controlled virtual agents (e.g., Bailenson et al., 2003; Blascovich et al., 2002; Gratch et al., 2016; Lucas et al., 2014). Results of this study indicated on a marginally significant level that an avatar controlled in real-time by a person is perceived to possess more agency and elicits people to attribute more blame to the person(s) behind the software than a virtual agent fully controlled by an autonomous computer program. Although differences are rather small, an avatar is perceived to have more control over its actions and consequently more responsibility for the actions is ascribed to it.

Previous literature confirms that responsibility is connected to agency – a person decides to act in a certain way which has an effect on the person's environment and for which the person is then responsible (Coeckelbergh, 2020). However, in addition to causality, other factors contribute as well, such as knowledge or foreseeability, intention as well as coercion or rather voluntariness (Coeckelbergh, 2020; Mao and Gratch, 2003). In accordance with Weiner (1995), causality and coercion are determinants of responsibility, however, foreseeability and intention decide over the intensity of blame assigned to an actor (in case of negative behavior). Likewise, following the so-called Aristoteles approach, two conditions need to be met to assign full responsibility: the control-condition (having control over the action/causing the action) and the epistemic-condition (knowing/being aware of actions; Fischer and Ravizza, 1998). Against this background, it could be argued that responsibility and blameworthiness are only fully attributed to a real human since a computer program is perceived to be more constrained and to have less of a free choice over how to act. Humans are generally inclined to attribute failures to others rather than to themselves to protect their self-esteem (Blaine and Crocker, 1993; Bradley, 1978; Miller, 1976). This may further explain why people attribute blame to their interaction partner if there is a real person behind it who they can ascribe responsibility to. Since technologies lack necessary preconditions for responsibility such as freedom, consciousness, and foreknowledge, humans need to be held responsible for what the technology does (Neri et al., 2020). This may explain why more agency and blame are attributed to an avatar than a virtual agent. That the virtual agent's behavior was developed and/or programmed by (a) person(s) at some point appears to be neglected. Previous research shows that people do not think about the programmer(s) and their intentions when interacting with a computer program (Sundar and Nass, 2000). However, effects were only marginally significant and thus need to be interpreted with caution.

Furthermore, people who thought they interacted with an avatar compared to a virtual agent reported more negative affect. Thus, interacting with a rude person affects people's mood in a more negative way compared to interacting with a rude virtual agent. An explanation could be that the impoliteness is perceived as hostility which is worse coming from a person than from a computer. According to Baumeister and Leary (1995), to be accepted, liked, and included by others is a fundamental human need. Even stronger than this need to belong is people's aversion to rejection and the distress it is accompanied by (Leary, 2001). The sociometer theory by Leary and Baumeister (2000) describes how people monitor other people's reactions towards them and how negative affect is triggered when relational deficiencies or signs of hostility are detected. Hostility can be seen as interpersonal devaluation which is the case when the interaction partner regards the relationship as worthless or even assigns a negative value to it (Leary, 2001). Brad's behavior in the current study, which involved exclusively harsh, negative feedback and degrading comments, likely conveyed hostility and relational devaluation. This in turn would explain participants' negative emotional reactions which did not occur in such a strong manner with a computer interaction partner. The reason may be that hostility by a person conveys social and relational meaning which a hostile computer program does not.

Another explanation could be that negative feedback from a computer program is simply perceived as objective evaluation of the performance. In general, people do not expect a computer to have or express personal feelings (Ezer et al., 2009; Horstmann and Krämer, 2019). Consequently, the way a computer program gives negative feedback may be perceived as programmed or instructed and not as hostility and personal devaluation which would be the case with another person controlling an avatar.

5.2. *Type of agency: Virtual agent vs. agent representative*

In this study, the focus lays on a special form of computer agent called agent representative. Agent representatives are supposed to represent a person by acting on instructions which the person specified beforehand to be adequately represented (de Melo et al., 2018). Since agent representatives are autonomous computer programs acting on the instructions of a real person, the categories person and computer increasingly blur. This makes it so interesting to examine how reactions to agent representatives compare to how people react to another person or to a computer agent. Previous research addressed the question how people would instruct an agent representative to behave on their behalf (de Melo et al., 2018; Mell et al., 2018). Closing a gap, this study focuses on the other side: how people who interact with another person's agent representative perceive and react to this agent.

Results of this study show that people attribute more agency to an avatar than to an agent representative. This can be simply explained by the fact that the agent representative acts upon fixed instructions and thus is constrained in its scope of actions (Fischer and Ravizza, 1998; Weiner, 1995). The avatar is controlled in real-time and thus less constrained and more flexible to react to the current situation. No significant difference regarding agency was found between agent representative and virtual agent. This indicates that both forms are perceived as constrained, to a person's instructions or its programming.

Regarding blame attribution, there was no difference between the agent representative and the avatar as well as the virtual agent. Descriptive values show that the least amount of blame was attributed to the person(s) behind the software in case of a virtual agent and most with an avatar. The agent representative was found in the middle with no significant difference to either one of the other two forms. These results indicate that with the agent representative, people were either divided or undecided regarding the question who or what is responsible for the agent representative's behavior. An explanation could be that by using a third-party representative people neither attribute blame to the agent representative nor to the person who instructed the agent to act this way, but rather that blame is deflected (Bivins, 2006; Royzman and Baron, 2002). Certain aspects such as agency, followed by responsibility and blame, may only be attributed to real persons and may not be representable by a computer agent. Furthermore, it is noteworthy that in general blame appears to be predominantly attributed to oneself and to a lesser extent to the software or person(s) behind the software. Therefore, people seem to blame themselves most for their performance regardless of the type of agency controlling their interaction partner.

In line with the other results, no significant difference regarding participant's negative affect was found between agent representative and virtual agent and only a marginally significant difference between agent representative and avatar. The avatar elicited the most negative affect which may be because impolite criticism coming from a person is perceived more personal than coming from a computer program (Baumeister and Leary, 1995; Ezer et al., 2009; Horstmann and Krämer, 2019). This is followed by the agent representative (with a marginally significant difference), probably since the impoliteness is still coming to some extent from a person, although only transferred via instructions and not in real-time. The virtual agent elicits the least amount of negative affect, likely because here the negative feedback is not perceived as coming from a person at all (neither in real-time nor in form of instructions; programmers are neglected; Sundar and Nass, 2000). However, since there were only small differences between the conditions, they need to be interpreted with great caution.

5.3. *Level of expertise: High vs. low expertise*

Another dimension considered in this work is the expertise of the source since this may affect the perception of the feedback and people's reactions to it (Bannister, 1986). In the past, feedback from a source with high expertise and rich experience was found to be more acceptable and convincing (Berlo et al., 1969; Fogg, 2002; Hovland et al., 1953). Therefore, negative feedback should be evaluated as more appropriate and thus as more positive when an expert tutor gives it compared to a peer study partner with low expertise.

In contrast to our assumptions, expertise by itself and in interaction with the type of agency had no significant effect on participant's attribution of blame and their affective state. A contextual explanation could be the one-sided structure of the interaction with Brad. Brad was the one questioning and criticizing the participants which may have conveyed the impression of a hierarchic as well as knowledge-based gap between the participant and Brad. This may have transferred more power to Brad regardless of Brad's advertised level of expertise (Emerson, 1962). Furthermore, responsibility is rather attributed internally with a person of high status and externally with a person of low status (Thibaut and Riecken, 1955). Thus, in case of unpleasant behavior, more blame might have been attributed to an interaction partner with high expertise compared to low expertise. However, since negative feedback by a high expertise tutor might be accepted more, this may have counteracted the attribution of blame leading to non-significant differences between the interaction partners with low and high expertise.

Another explanation could be that in all conditions, Brad only asked questions and made mean comments, but never explained what exactly was wrong about the participant's answers and what the right answers would be. Thus, Brad did not act like an expert tutor as which he was portrayed in the high expertise conditions, which may have caused participants to question Brad's expertise. This goes in line with Rickenberg and Reeves (2000) who emphasize the great significance of an agent's behavior during an interaction for the subsequent evaluation of this agent. For instance, a study showed that a robot's interaction skills predominantly determine how this robot is evaluated, while a previous description of this robot only plays a subordinate role (Horstmann and Krämer, 2020). Against this background, it is possible that the expert tutor was not perceived to have significantly more expertise than the study partner. This reasoning is supported by the task attractiveness and competence measurements, which also show no significant differences between the two expertise levels.

5.4. *Interaction partner and interaction evaluation, contact intentions, and facial expressivity*

Neither type of agency nor level of expertise nor the interaction of the two factors had an influence on the evaluation of the interaction partner and the interaction as well as on future contact intentions. An explanation may be that the explicitly negative behavior of the interaction partner was so overwhelming that participants evaluated the partner as well as the interaction negatively not taking the partner's agency and expertise into account. This goes in line with previous research emphasizing the strong effects that the behavior of a non-human interaction partner has on people which may overshadow previously given descriptions (Horstmann and Krämer, 2020; Rickenberg and Reeves, 2000). Most likely, this also explains why additional analyses of the facial expressivity in the US participant videos revealed no significant differences between the different types of agency. Overall, these are valuable findings since it suggests that people evaluate their interaction partner based on what they are experiencing with them. Who or what is controlling the interaction partner's behavior and the partner's level of expertise appear to be neglected.

5.5. Limitations and future research

One limitation of the study is that the main dependent variables of the study were self-reported. Considering also behavioral measurements could bring further insights, for example regarding verbal and non-verbal expressions of anger and frustration as well as reciprocity and retaliation behaviors. Another aspect which needs to be mentioned is that about 26 participants failed the manipulation check regarding the type of agency indicating that they were not able to recall with whom or what they were told to be interacting with at the beginning. However, the manipulation check occurred at the end of the experiment which might have contributed to these recall failures. Manipulations were reinforced several times (written and orally) which gives us the confidence that during the interaction people were aware of with whom or what they supposedly interacted with. In future studies, the manipulations may also be reinforced during the interaction and manipulation checks should occur earlier. In addition to the manipulation checks after the manipulations, in future studies people's impression of their interaction partner should also be assessed after the interaction took place since the behavior may have an altering effect. Besides that, to reinforce the manipulation of the interaction partner's expertise, visual cues were used (glasses and suit for high expertise; button-down shirt and no glasses for low expertise). Together with the assigned roles, this may as well have conveyed other influencing aspects than expertise, such as status or formality.

Furthermore, comparing the effects of negative feedback, as examined in this study, with neutral and positive feedback as well as the consideration of different contexts and environments and long-term effects would deliver further insights. A learning situation may particularly constitute a situation where people tend to blame themselves the most, as our results suggest as well. Therefore, it would be interesting to shift the experimental setting to a situation with a higher focus on attributing blame to an interaction partner. Nevertheless, we would like to emphasize that this study employed an elaborate study design using the wizard of oz technique (Dahlbäck et al., 1993) to ensure an implementation of a convincing interaction and consequently a more realistic evaluation of the (artificial) interaction partner. We further would like to highlight the study's sample composition with mixed backgrounds regarding education and employment status, but also regarding culture with a half German, half US-American sample, which enhances the generalizability of the results.

6. Conclusion

The current study presents novel insights in the field of autonomous, computer-controlled agents, which follow the instructions of a specific person in a certain situation and thus represent a mix of a computer-controlled but person-instructed interaction partner. Results of this study demonstrate some small but noteworthy differences between an agent representative, an avatar, and a virtual agent interaction partner. In the avatar-conditions, people rather attributed agency and blame to the person(s) behind the software, which was less the case with an agent representative and the least with a virtual agent. The effects on people's affective state reflect this pattern: the most negative affect was reported when people interacted with an avatar, followed by an agent representative. The least negative affect was reported with a virtual agent. In sum, the results of this study partly confirm, but also extend previous insights regarding the perception of agent representatives as mix form by showing that there are small but plausible differences regarding attribution of agency and blame as well as people's affect, but no differences in the evaluation of the interaction partner and the interaction in general. Consequently, when designing computer agents to represent persons in certain situations it needs to be kept in mind that on a cognitive level, blame is neither clearly attributed to the person who gave the instructions on how to behave nor to the agent representative performing the behavior. This poses the danger of people escaping their responsibility and acting in an unethical way through their agent representatives. However, people are still affected by rude, negative behavior by an agent representative. Accordingly, some restrictions should be implemented by HCI designers with regard to the instructions people can give to their agent representative so that certain norms of polite social interactions are always maintained. Furthermore, to keep people constantly aware of who or what is controlling their virtual interaction partner's behavior, this information should be repeated and emphasized by the virtual interaction partner throughout the interaction (e.g., “I was instructed by … to say…”).

CRediT authorship contribution statement

Aike C. Horstmann: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization, Project administration. **Jonathan Gratch:** Conceptualization, Methodology, Validation, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Nicole C. Krämer:** Conceptualization, Methodology, Validation, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Research was sponsored by the Air Force Office of Scientific Research under grant FA9550-18-1-0182 and the Army Research Office under Cooperative Agreement Number W911NF-20-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

Arras, K.O., Cerqui, D., 2005. Do We Want to Share Our Lives and Bodies with Robots? A 2000 People Survey. Autonomous Systems Lab, Swiss Federal Institute of Technology, EPFL. https://doi.org/10.3929/ETHZ-A-010113633. Technical Report Nr. 0605-0012005.

Baer, M., Frese, M., 2003. Innovation is not enough: climates for initiative and psychological safety, process innovations, and firm performance. J. Organ. Behav. 24 (1), 45–68. https://doi.org/10.1002/job.179.

Bailenson, J.N., Blascovich, J., Beall, A.C., Loomis, J.M., 2003. Interpersonal distance in immersive virtual environments. Pers. Soc. Psychol. Bull. 29 (7), 819–833. https://doi.org/10.1177/0146167203029007002.

Banks, J., 2019. A perceived moral agency scale: development and validation of a metric for humans and social machines. Comput. Human Behav. 90, 363–371. https://doi.org/10.1016/j.chb.2018.08.028.

Bannister, B.D., 1986. Performance outcome feedback and attributional feedback: interactive effects on recipient responses. J. Appl. Psychol. 71 (2), 203–210. https://doi.org/10.1037/0021-9010.71.2.203.

Bartneck, C., Bleeker, T., Bun, J., Fens, P., Riet, L., 2010. The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. Paladyn. J. Behav. Robot. 1 (2), 109–115. https://doi.org/10.2478/s13230-010-0011-3.

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

Bartneck, C., Rosalia, C., Menges, R., Deckers, I., 2005. Robot abuse – a limitation of the media equation. A. de Angeli, S. Brahnam, & P. Wallis (Chairs). In: Abuse: The Darker Side of Human-Computer Interaction: Proceedings of an INTERACT 2005 Workshop.

Baumeister, R.F., Leary, M.R., 1995. The need to belong: desire for interpersonal attachments as a fundamental human motivation. Psychol. Bull. 117 (3), 497–529.

Beier, G., 1999. Kontrollüberzeugungen im Umgang mit Technik [Locus of control when using technology]. Report Psychologie 9, 684–693.

Bente, G., Feist, A., Elder, S., 1996. Person perception effects of computer-simulated male and female head movement. J. Nonverbal Behav. 20 (4), 213–228. https://doi.org/10.1007/BF02248674.

Berlo, D.K., Lemert, J.B., Mertz, R.J., 1969. Dimensions for evaluating the acceptability of message sources. Public Opin. Q. 33 (4), 563–576. https://doi.org/10.1086/267745.

Bigman, Y.E., Gray, K., 2018. People are averse to machines making moral decisions. Cognition 181, 21–34. https://doi.org/10.1016/j.cognition.2018.08.003.

Bivins, T., 2006. Responsibility and accountability. In: Fitzpatrick, K.R., Bronstein, C. (Eds.), Ethics in Public Relations: Responsible Advocacy, pp. 19–38.

Blaine, B., Crocker, J., 1993. Self-Esteem and self-serving biases in reactions to positive and negative events: an integrative review. In: Baumeister, R.F. (Ed.), Self-Esteem, Self-Esteem, 108. Springer US, pp. 55–85. https://doi.org/10.1007/978-1-4684-8956-9_4.

Blascovich, J., 2002. A theoretical model of social influence for increasing the utility of collaborative virtual environments. In: Broll, W., Greenhalgh, C., Churchill, E.F. (Eds.), Proceedings of the 4th International Conference on Collaborative Virtual Environments - CVE '02. ACM Press, pp. 25–30. https://doi.org/10.1145/571878.571883.

Blascovich, J., Loomis, J.M., Beall, A.C., Swinth, K.R., Hoyt, C.L., Bailenson, J.N., 2002. Immersive virtual environment technology as a methodological tool for social psychology. Psychol. Inq. 13 (2), 103–124. https://doi.org/10.1207/S15327965PLI1302_01.

Bradley, G.W., 1978. Self-serving biases in the attribution process: a reexamination of the fact or fiction question. J. Pers. Soc. Psychol. 36 (1), 56–71. https://doi.org/10.1037/0022-3514.36.1.56.

Burgoon, J.K., Bonito, J.A., Bengtsson, B., Cederberg, C., Lundeberg, M., Allspach, L., 2000. Interactivity in human–computer interaction: a study of credibility, understanding, and influence. Comput. Human Behav. 16 (6), 553–574. https://doi.org/10.1016/S0747-5632(00)00029-7.

Burgoon, J.K., Hale, J.L., 1988. Nonverbal expectancy violations: model elaboration and application to immediacy behaviors. Commun. Monogr. 55 (1), 58–79. https://doi.org/10.1080/03637758809376158.

Burgoon, J.K., Walther, J.B., 1990. Nonverbal expectancies and the evaluative consequences of violations. Hum. Commun. Res. 17 (2), 232–265. https://doi.org/10.1111/j.1468-2958.1990.tb00232.x.

S.B.-1001 Bots: disclosure, Business and professions Code, Division 7 - General business regulations, Part 3 - Representations to the public, Chapter 6 - Bots (2019).

Carolus, A., Muench, R., Schmidt, C., Schneider, F., 2019. Impertinent mobiles - effects of politeness and impoliteness in human-smartphone interaction. Comput. Human Behav. 93, 290–300. https://doi.org/10.1016/j.chb.2018.12.030.

Carpinella, C.M., Wyman, A.B., Perez, M.A., Stroessner, S.J., 2017. The robotic social attributes scale (RoSAS): development and validation (Eds.). In: Mutlu, B., Tscheligi, M., Weiss, A., Young, J.E. (Eds.), Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction - HRI '17. IEEE, pp. 254–262. https://doi.org/10.1145/2909824.3020208.

Cha, M., Park, J.-G., Lee, J, 2014. Effects of team member psychological proximity on teamwork performance. Team Perform. Manage. 20 (1/2), 81–96. https://doi.org/10.1108/TPM-03-2013-0007.

Chan, T.-W., Baskin, A.B, 1988. "Studying with the prince" the computer as a learning companion. In: Winkels, R. (Ed.), Proceedings of the International Conference on Intelligent Tutoring Systems - ITS '88. ACM, pp. 194–200.

Choi, J.J., Kim, Y., Kwak, S.S., 2014. Are you embarrassed? The impact of robot types on emotional engagement with a robot (Eds.). In: Sagerer, G., Imai, M., Belpaeme, T., Thomaz, A. (Eds.), Proceedings of the 9th ACM/IEEE international conference on Human-robot interaction - HRI '14. ACM Press, pp. 138–139. https://doi.org/10.1145/2559636.2559798.

Coeckelbergh, M., 2020. Artificial intelligence, responsibility attribution, and a relational justification of explainability. Sci. Eng. Ethics 26 (4), 2051–2068. https://doi.org/10.1007/s11948-019-00146-8.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. Erlbaum.

Dahlbäck, N., Jönsson, A., Ahrenberg, L., 1993. Wizard of Oz studies: why and how. Knowl. Based Syst. 6 (4), 258–266. https://doi.org/10.1016/0950-7051(93)90017-N.

de Melo, C.M., Marsella, S., Gratch, J, 2016. Do as I say, not as I do" Challenges in delegating decisions to automated agents (Eds.),. In: Thangarajah, J., Tuyls, K., Jonker, C.M., Marsella, S. (Eds.), Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems - AAMAS '16. IFAAMAS, pp. 949–956.

de Melo, C.M., Marsella, S., Gratch, J., 2018. Social decisions and fairness change when people's interests are represented by autonomous agents. Auton. Agent. Multi Agent Syst. 32 (1), 163–187. https://doi.org/10.1007/s10458-017-9376-6.

Dijks, M.A., Brummer, L., Kostons, D., 2018. The anonymous reviewer: the relationship between perceived expertise and the perceptions of peer feedback in higher education. Assess. Evaluat. Higher Educ. 43 (8), 1258–1271. https://doi.org/10.1080/02602938.2018.1447645.

Edmondson, A., 1999. Psychological safety and learning behavior in work teams. Adm. Sci. Q. 44 (2), 350–383. https://doi.org/10.2307/2666999.

Emerson, R.M., 1962. Power-dependence relations. Am. Sociol. Rev. 27 (1), 31–41. https://doi.org/10.2307/2089716.

Eyssel, F., Hegel, F., 2012. (S)he's got the look: gender stereotyping of robots. J. Appl. Soc. Psychol. 42 (9), 2213–2230. https://doi.org/10.1111/j.1559-1816.2012.00937.x.

Eyssel, F., Kuchenbrandt, D., Bobinger, S., 2011. Effects of anticipated human-robot interaction and predictability of robot behavior on perceptions of anthropomorphism (Eds.),. In: Billard, A., Kahn, P.H., Adams, J.A., Trafton, G. (Eds.), Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction - HRI '11. ACM Press, pp. 61–68. https://doi.org/10.1145/1957656.1957673.

Ezer, N., Fisk, A.D., Rogers, W.A., 2009. Attitudinal and intentional acceptance of domestic robots by younger and older adults. In: Stephanidis, C. (Ed.), Proceedings of the 5th International Conference on Universal Access in Human-Computer Interaction - UAHCI '09, 5615. Springer Berlin, Heidelberg, pp. 39–48. https://doi.org/10.1007/978-3-642-02710-9_5.

Fischer, J.M., Ravizza, M., 1998. Responsibility and control: A Theory of Moral Responsibility. Cambridge University Press.

Fisher, C.D., Ilgen, D.R., Hoyer, W.D., 1979. Source credibility, information favorability, and job offer acceptance. Acad. Manage. J. 22 (1), 94–103. https://doi.org/10.5465/255481.

Fogg, B.J., 2002. Persuasive technology: Using Computers to Change What We Think and do. Morgan Kaufmann Publishers. https://doi.org/10.1145/764008.763957.

Fogg, B.J., Nass, C.I., 1997. Silicon sycophants: the effects of computers that flatter. Int. J. Hum. Comput. Stud. 46 (5), 551–561. https://doi.org/10.1006/ijhc.1996.0104.

Fogg, B.J., Tseng, H., 1999. The elements of computer credibility. In: Williams, M.G. (Ed.), Proceedings of the SIGCHI conference on Human Factors in Computing Systems. ACM, pp. 80–87. https://doi.org/10.1145/302979.303001.

Fox, J., Ahn, S.J., Janssen, J.H., Yeykelis, L., Segovia, K.Y., Bailenson, J.N., 2015. Avatars versus agents: a meta-analysis quantifying the effect of agency on social influence. Hum.–Comput. Interact. 30 (5), 401–432. https://doi.org/10.1080/07370024.2014.921494.

Gallagher, H.L., Jack, A.I., Roepstorff, A., Frith, C.D., 2002. Imaging the intentional stance in a competitive game. Neuroimage 16 (3 Pt 1), 814–821. https://doi.org/10.1006/nimg.2002.1117.

Gazzola, V., Rizzolatti, G., Wicker, B., Keysers, C., 2007. The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. Neuroimage 35 (4), 1674–1684. https://doi.org/10.1016/j.neuroimage.2007.02.003.

Gogoll, J., Uhl, M., 2018. Rage against the machine: automation in the moral domain. J. Behav. Exp. Econ. 74, 97–103. https://doi.org/10.1016/j.socec.2018.04.003.

Graefe, A., Haim, M., Haarmann, B., Brosius, H.-B, 2018. Readers' perception of computer-generated news: credibility, expertise, and readability. Journalism 19 (5), 595–610. https://doi.org/10.1177/1464884916641269.

Gratch, J., DeVault, D., Lucas, G.M., 2016. The benefits of virtual humans for teaching negotiation. In: Traum, D., Swartout, W., Khooshabeh, P., Kopp, S., Scherer, S., Leuski, A. (Eds.), Intelligent Virtual Agents: Proceedings of the 16th International Conference on Intelligent Virtual Agents - IVA '16. Springer, pp. 283–294. https://doi.org/10.1007/978-3-319-47665-0_25.

Hancock, P.A., 2017. Imposing limits on autonomous systems. Ergonomics 60 (2), 284–291. https://doi.org/10.1080/00140139.2016.1190035.

Hoffmann, L., Krämer, N.C., Lam-chi, A., Kopp, S., 2009. Media equation revisited: do users show polite reactions towards an embodied agent? In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsson, H.H. (Eds.), Intelligent Virtual Agents: Proceedings of the 9th International Conference on Intelligent Virtual Agents - IVA '09. Springer, pp. 159–165. https://doi.org/10.1007/978-3-642-04380-2_19.

Horstmann, A.C., Bock, N., Linhuber, E., Szczuka, J.M., Straßmann, C., Krämer, N.C., 2018. Do a robot's social skills and its objection discourage interactants from switching the robot off? PLoS ONE 13 (7), e0201581. https://doi.org/10.1371/journal.pone.0201581.

Horstmann, A.C., Krämer, N.C., 2019. Great expectations? Relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment. Front. Psychol. 10, 939. https://doi.org/10.3389/fpsyg.2019.00939.

Horstmann, A.C., Krämer, N.C., 2020. Expectations vs. actual behavior of a social robot: an experimental investigation of the effects of a social robot's interaction skill level and its expected future role on people's evaluations. PLoS ONE 15 (8), e0238133. https://doi.org/10.1371/journal.pone.0238133.

Hovland, C.I., Janis, I.L., Kelley, H.H., 1953. Communication and Persuasion. Yale University Press.

Kang, S.-H., Gratch, J., Wang, N., Watt, J.H, 2008. Agreeable people like agreeable virtual humans. In: Prendinger, H., Lester, J.C., Ishizuka, M. (Eds.), Intelligent Virtual Agents: Proceedings of the 8th International Conference on Intelligent Virtual Agents - IVA '08. Springer, pp. 253–261. https://doi.org/10.1007/978-3-540-85483-8_26.

Karrer, K., Glaser, C., Clemens, C., Bruder, C., 2009. Technikaffinität erfassen: der Fragebogen TA-EG [Measuring technical affinity - the questionnaire TA-EG]. Der Mensch Im Mittelpunkt Technischer Systeme 8, 196–201.

Kelley, H.H., 1973. The processes of causal attribution. Am. Psychol. 28 (2), 107–128. https://doi.org/10.1037/h0034225.

Kelley, H.H., Michela, J.L., 1980. Attribution theory and research. Annu. Rev. Psychol. 31, 457–501. https://doi.org/10.1146/annurev.ps.31.020180.002325.

Kersey, C., Eugenio, Di, B., Jordan, Katz, S, 2010. KSC-PaL: a peer learning agent. D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, V. Aleven, J. Kay, & J. Mostow (Eds.). In: Intelligent tutoring systems, 6095. Springer, pp. 72–81. https://doi.org/10.1007/978-3-642-13437-1_8.

Krämer, N.C., Leiße, L.-M., Hollingshead, A., Gratch, J, 2017. Evaluated by a machine: effects of negative feedback by a computer or human boss. J. Beskow, C. Peters, G. Castellano, C. O'Sullivan, I. Leite, & S. Kopp (Eds.). In: Intelligent Virtual Agents: Proceedings of the 17th International Conference on Intelligent Virtual Agents - IVA '17. Springer, pp. 235–238. https://doi.org/10.1007/978-3-319-67401-8_29.

Krämer, N.C., von der Pütten, A.M., Eimler, S.C., 2012. Human-agent and human-robot interaction theory: similarities to and differences from human-human interaction. In: Zacarias, M., de Oliveira, J.V. (Eds.), Studies in Computational Intelligence. Human-Computer Interaction: The Agency Perspective, 396. Springer, Berlin Heidelberg, pp. 215–240. https://doi.org/10.1007/978-3-642-25691-2_9.

Lea, M., Spears, R., 1992. Paralanguage and social perception in computer-mediated communication. J. Organ. Comput. 2 (3–4), 321–341. https://doi.org/10.1080/10919399209540190.

Leary, M.R., Leary, M.R., 2001. Toward a conceptualization of interpersonal rejection. Interpersonal Rejection. Oxford University Press, pp. 3–20.

Leary, M.R., Baumeister, R.F., 2000. The nature and function of self-esteem: sociometer theory. Adv. Exp. Soc. Psychol. 32, 1–62. https://doi.org/10.1016/S0065-2601(00)80003-9.

Lee, S., Lau, I.Y., Kiesler, S., Chiu, C.-Y, 2005. Human mental models of humanoid robots. In: Proceedings of the 2005 IEEE International Conference on Robotics and Automation - ICRA '05. IEEE, pp. 2767–2772. https://doi.org/10.1109/ROBOT.2005.1570532.

Leviathan, Y., & Matias, Y. (2018). Google Duplex: an AI system for accomplishing real-world tasks over the phone. Google AI Blog. https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html.

Lucas, G.M., Gratch, J., King, A., Morency, L.-P, 2014. It's only a computer: virtual humans increase willingness to disclose. Comput. Human Behav. 37, 94–100. https://doi.org/10.1016/j.chb.2014.04.043.

Malle, B.F., Knobe, J., 1997. The folk concept of intentionality. J. Exp. Soc. Psychol. 33 (2), 101–121. https://doi.org/10.1006/jesp.1996.1314.

Malle, B.F., Moses, L.J., Baldwin, D.A, 2001. Intentions and Intentionality: Foundations of Social Cognition. MIT Press.

Mao, W., Gratch, J., 2003. The social credit assignment problem. G. Goos, J. Hartmanis, J. van Leeuwen, T. Rist, R. S. Aylett, D. Ballin, & J. Rickel (Eds.). In: Intelligent Virtual Agents: Proceedings of the 3rd International Conference on Intelligent Virtual Agents - IVA '03. Springer, pp. 39–47. https://doi.org/10.1007/978-3-540-39396-2_8.

McCroskey, J.C., McCain, T.A., 1974. The measurement of interpersonal attraction. Speech Monogr. 41 (3), 261–266. https://doi.org/10.1080/03637757409375845.

McCroskey, J.C., Young, T.J., 1981. Ethos and credibility: the construct and its measurement after three decades. Cent. States Speech J. 32 (1), 24–34. https://doi.org/10.1080/10510978109368075.

Mell, J., Lucas, G.M., Gratch, J., 2018. Welcome to the real world: how agent strategy increases human willingness to deceive. In: André, E., Koenig, S. (Eds.), Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems - AAMAS '18. IFAAMAS, pp. 1250–1257.

Miller, D.T., 1976. Ego involvement and attributions for success and failure. J. Pers. Soc. Psychol. 34 (5), 901–906. https://doi.org/10.1037/0022-3514.34.5.901.

Mosier, K.L., Skitka, L.J., Heers, S., Burdick, M., 1997. Automation bias: decision making and performance in high-tech cockpits. Int. J. Aviat. Psychol. 8 (1), 47–63. https://doi.org/10.1207/s15327108ijap0801_3.

Nass, C.I., Fogg, B.J., Moon, Y., 1996a. Can computers be teammates? Int. J. Hum. Comput. Stud. 45 (6), 669–678. https://doi.org/10.1006/ijhc.1996.0073.

Nass, C.I., Moon, Y., 2000. Machines and mindlessness: social responses to computers. J. Soc. Issues 56 (1), 81–103. https://doi.org/10.1111/0022-4537.00153.

Nass, C.I., Moon, Y., Carney, P., 1999. Are people polite to computers? Responses to computer-based interviewing systems. J. Appl. Soc. Psychol. 29 (5), 1093–1109. https://doi.org/10.1111/j.1559-1816.1999.tb00142.x.

Nass, C.I., Reeves, B., Leshner, G., 1996b. Technology and roles: a tale of two TVs. J. Commun. 46 (2), 121–128. https://doi.org/10.1111/j.1460-2466.1996.tb01477.x.

Neri, E., Coppola, F., Miele, V., Bibbolino, C., Grassi, R., 2020. Artificial intelligence: who is responsible for the diagnosis? La Radiologia Medica 125 (6), 517–521. https://doi.org/10.1007/s11547-020-01135-9.

Nourani, M., King, J.T., Ragan, E.D., 2020. The role of domain expertise in user trust and the impact of first impressions with intelligent systems. In: Proceedings of the AAAI Conference on Human Computation and Crowdsourcing - HCOMP '20, 8(1), pp. 112–121. https://arxiv.org/pdf/2008.09100.

Reeves, B., Nass, C.I., 1996. The Media equation: How People Treat Computers, Television, and New Media Like Real People and Places. CSLI Publications.

Rickenberg, R., Reeves, B., 2000. The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. In: Turner, T. (Ed.), Proceedings of the SIGCHI conference on Human Factors in Computing Systems. ACM, pp. 49–56. https://doi.org/10.1145/332040.332406.

Rosenthal-von der Pütten, A.M., Schulte, F.P., Eimler, S.C., Sobieraj, S., Hoffmann, L., Maderwald, S., Brand, M., Krämer, N.C., 2014. Investigations on empathy towards humans and robots using fMRI. Comput. Human Behav. 33, 201–212. https://doi.org/10.1016/j.chb.2014.01.004.

Royzman, E.B., Baron, J., 2002. The preference for indirect harm. Soc. Justice Res. 15 (2), 165–184. https://doi.org/10.1023/A:1019923923537.

Sayin, E., Krishna, A., 2019. You can't be too polite, Alexa! Implied politeness of mechanized auditory feedback and its impact on perceived performance accuracy. In: Bagchi, R., Block, L., Lee, L. (Eds.), Advances in Consumer Research, 47. Association for Consumer Research, pp. 243–248.

Sundar, S.S., Nass, C.I., 2000. Source orientation in human-computer interaction. Communic. Res. 27 (6), 683–703. https://doi.org/10.1177/009365000027006001.

Tapal, A., Oren, E., Dar, R., Eitam, B., 2017. The Sense of Agency Scale: a measure of consciously perceived control over one's mind, body, and the immediate environment. Front. Psychol. 8, 1552. https://doi.org/10.3389/fpsyg.2017.01552.

Thibaut, J.W., Riecken, H., 1955. Some determinants and consequences of the perception of social causality. J. Pers. 24 (2), 113–133. https://doi.org/10.1111/j.1467-6494.1955.tb01178.x.

Vizcaíno, A., 2005. A simulated student can improve collaborative learning. Int. J. Artif. Intell. Educ. 15 (1), 3–40.

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. Comput. Human Behav. 26 (6), 1641–1650. https://doi.org/10.1016/j.chb.2010.06.012.

Watson, D., Clark, L.A., Tellegen, A., 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. J. Pers. Soc. Psychol. 54 (6), 1063–1070.

Weiner, B., 1995. Judgements of responsibility: A Foundation for a Theory of Social Conduct. Guilford Press.

Wilson, W., Chun, N., Kayatani, M., 1965. Projection, attraction, and strategy choices in intergroup competition. J. Pers. Soc. Psychol. 2 (3), 432–435. https://doi.org/10.1037/h0022287.

Wullenkord, R., Fraune, M.R., Eyssel, F., Sabanovic, S., 2016. Getting in Touch: how imagined, actual, and physical contact affect evaluations of robots. In: Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication - RO-MAN '16. IEEE, pp. 980–985. https://doi.org/10.1109/ROMAN.2016.7745228.

Aike C. Horstmann received her M.Sc. in Applied Cognitive and Media Science in 2017. In 2021 she completed her Ph.D. at the Faculty of Social Psychology: Media and Communication at the University of Duisburg-Essen, Germany. With her research she focuses on human-computer interaction, i.e., interactions with social robots and virtual agents, with a special interest on expectations, attributions, and behavior. She also led various student research projects and taught courses on human-computer interaction at the University of Duisburg-Essen.

Jonathan Gratch is a Research Full Professor of Computer Science, Psychology and Media Arts and Practice at the University of Southern California (USC) and Director for Virtual Human Research at USC's Institute for Creative Technologies. He completed his Ph.D. in Computer Science at the University of Illinois in Urbana-Champaign in 1995. Dr. Gratch's research focuses on computational models of human cognitive and social processes, especially emotion, and explores these models' role advancing psychological theory and in shaping human-machine interaction. He is the founding Editor-in-Chief of IEEE's Transactions on Affective Computing and founding Associate Editor of Affective Science.

Nicole Krämer is Full Professor of Social Psychology: Media and Communication at the University of Duisburg-Essen, Germany. She completed her PhD in Psychology at the University of Cologne, Germany, in 2001 and received the venia legendi for psychology in 2006. Dr. Krämer's research focuses on social psychological aspects of human-machine-interaction (especially social effects of robots and virtual agents) and computer-mediated-communication (CMC). She heads numerous projects that received third party funding. She served as Editor-in-Chief of the Journal of Media Psychology 2015–2017 and currently is Associate Editor of the Journal of Computer Mediated Communication (JCMC).