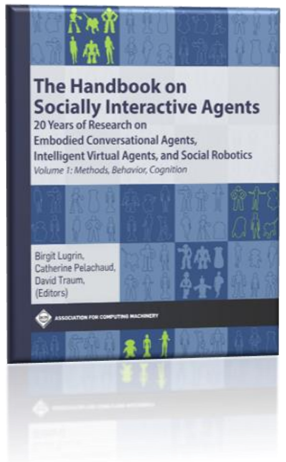
Social Reactions to Socially Interactive Agents and Their Ethical Implications

Nicole Krämer and Arne Manzeschke



Author note:

This is a preprint. The final article is published in "The Handbook on Socially Interactive Agents" by ACM books.

Citation information:

Krämer, N., and Manzeschke A., (2021). Social Reactions to Socially Interactive Agents and Their Ethical Implications. In B. Lugrin, C. Pelachaud, D. Traum (Eds.), *Handbook on Socially Interactive Agents – 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics*, Volume 1: Methods, Behavior, Cognition (pp. 77-103). ACM books.

DOI of the final chapter: 10.1145/3477322.3477326

DOI of volume 1 of the handbook: 10.1145/3477322

Correspondence concerning this chapter should be addressed to Nicole Krämer (nicole.kraemer@uni-due.de) or Arne Manzeschke (arne.manzeschke@evhn.de)

3 Social Reactions to Socially Interactive Agents and their Ethical Implications

Nicole Krämer and Arne Manzeschke

3.1 Motivation

Socially interactive agents have now been developed and discussed for more than 20 years. While they have aptly been described as a testbed to gain a better understanding of human communication skills, they have also always been developed with a view to supporting users by means of helpful applications and implementations. In this regard, they have from the beginning raised interest from a psychological and ethical point of view: Here, it soon became apparent that early applications – even when they were not yet very sophisticated regarding their interaction abilities – already proofed to be able to elicit social reactions from the users. The goal of the chapter therefore is to summarize research on these social effects. While this has already been done from a psychological perspective (Krämer, Rosenthal-von der Pütten & Hoffmann, 2015), the combination of a psychological and an ethical view is unique to this chapter.

Effects of socially interactive agents can be termed 'social' if a participant's emotional, cognitive, or behavioral reactions are similar to reactions shown during interactions with other human beings (for an early review see Krämer, 2005) or, in some respect, other animals (e. g. de Waal 2019). A number of studies demonstrate that those – actually inappropriate - reactions really do occur, sometimes even without the appearance of a human-like character. Studies (Fogg & Nass, 1997; Nass et al., 1997; Nass et al., 1994)indicate that in interactions with computers politeness phrases are employed, principles of person perception and gender stereotypes apply, and liking is triggered in a similar way as within human relationships (computer that 'flatter' and give positive feedback are evaluated more positive).
Although the studies suggest that a "rich human presentation" (Nass et al., 1994) is not necessary and specifically speech might be sufficient to trigger social reactions (Nass & Moon, 2000), studies suggest that intelligent virtual agents (IVA) might be especially prone to lead to social reactions. There is long-standing evidence that even subtle social phenomena such as impression management [see e.g. Leary, 1995] are prevalent in human agent interaction. When a human-like face is present, participants aim at leaving a favourable impression by e.g. choosing a socially desirable TV program (documentary about Albert Einstein compared to James Bond movie) (Krämer, Bente & Piesk, 2003) or by presenting themselves in a socially desirable way (Sproull et al., 1996). Krämer (2005) showed that IVAs affect the way in which users communicate with a TV-VCR system. When an IVA is visible instead of a graphical user interface or a user interface with speech output, users address the system significantly more often using natural speech rather than using a remote control. Additional qualitative analyses of the semantic content of all speech acts indicate that users seem to have a more human-like attitude and behavior towards the system when it is represented by an anthropomorphic agent.
Similarly, Hoffmann, Krämer, Lam-Chi and Kopp (2009) replicated the politeness experiment of Nass et al. (1999) with an IVA. Participants had to evaluate the IVA Max (at Bielefeld University) after a short interaction, whereby the questioning was either conducted by Max itself, a paper-and-pencil questionnaire in the same room, or a paper-and-pencil questionnaire in a separate room. The results showed that participants were more polite, i.e. provided better evaluations of Max's competence when Max itself asked for its evaluation, compared to the questionnaire in the same room. However, no significant difference was observed compared to the questionnaire in a separate room.

It should be noted that not all studies analyzing people´s reactions towards artificial entities found social effects in the manner and to the extent which was demonstrated in the CASA (Computers are Social Actors; Nass et al., 1994) studies and their follow-ups with embodied agents. Based on their criticism that the setting in the CASA studies focused too narrowly on the computer asking for evaluation rather than allowing for and assessing actual interaction, Shechtman and Horowitz (2003) conducted an experiment in which they analyzed the actual conversational discourse and – in order to address a second shortcoming of the CASA approach – added an experimental condition in which the interlocutor was announced to be human. Within the CASA approach, the authors frequently conclude that the behavior towards a computer is the same as that towards fellow humans, although this is not tested directly. In their study, Shechtman and Horowitz (2003) discovered as a result of the conversational analysis that human-human interaction is indeed different from human-computer interaction. When participants thought that their partner was human, they, for instance, used more words, spent more time and used more relationship statements.

Despite these findings, further studies more compellingly show that people indeed unconsciously react in social ways when confronted with an IVA: Krämer, Kopp, Becker-Asano, and Sommer (2013) demonstrate that people mimic the smiling behavior of a virtual agent. Mimicry is a widely cited phenomenon of human-human communication that has been shown to be especially indicative of the sociality of the situation. In a between-subjects design, participants conducted an eight-minute small-talk conversation with an agent that either did not smile, showed occasional smiles, or displayed frequent smiles. Results show that the human interaction partners themselves smiled for longer when the agent was smiling. Interestingly, the smiling activity did not have an impact on people´s evaluation of the agent, and nor were they able to reliably indicate whether the agent had smiled and whether this was occasional or frequent. Therefore, it can be concluded that the participants´ behavioral reactions were rather unconscious and automatic. Moreover, since mimicry has been shown to be indicative of the sociality of the situation, the fact that participants smiled at the agent can be seen as evidence that they – at least unconsciously – experienced the situation as social.

It can therefore be summarized that numerous studies yield evidence that artificial entities lead at least partly to social effects. From a psychological point of view these results are interesting because when we learn what triggers social reactions, we learn something about human nature. In line with this, Parise, Kiesler, Sproull and Waters (1999) suggested: *„As computer interfaces can display more life-like qualities such as speech output and personable characters or agents, it becomes important to understand and assess user´s interaction behavior within a social interaction framework rather than only a narrower machine interaction one"* (p. 123).

Given these social reactions and the necessity to understand these new forms of human-machine interactions within a social interaction framework, it also became necessary to reflect on these issues from a normative stance. Therefore, from the beginning, ethical questions have been raised. Kiesler and Sproull voiced first concerns about the tendency to use humanoid interfaces as early as 1996: "Many people want computers to be responsive to people. But do we also want people to be responsive to computers?" (Sproull et al., 1996; p. 119). With continuously more realistic and human-like machines, these ethical questions have neither been solved nor became less important: Human beings are those beings capable of speech, language, crying and laughter (Plessner 1970) and thus able to communicate in various modes with each other (i. e. rational, emotional, spiritual). It is this capability which characterizes the human species among all others and enables human beings to recognize each other as such. Encountering machines capable of human language blurs the fundamental distinction between the anthropos zoon logon echon (cf. Aristotle, Politics 1253a, meaning that human beings are political beings in an advanced sense due to their capability of speech. Aristotle discerns between language, which many animals are capable of, and language which only belongs to human beings.) and all the other beings. It poses the fundamental question who or what the 'speaking machine' should be recognized as. Furthermore, it should be acknowledged that speech does not only transport information from a sender to a receiver, but speech opens a social space accompanied by emotions such as expectations, worries, and - more generally spoken - social negotiations. But, who or what is the counterpart in these social negotiations?

Besides the problematic aspect that categories are blurred and the unambiguous assignment to the categories "human" or "thing" is no longer possible, another aspect is the question whether the human user is manipulated in an undue way. One very basic assumption to feature machines capable of speech is that it would enhance the human-machine interaction and thus the productivity. From an ethical point of view this seems to be at least questionable. Besides the fact that it is not clear whether efficiency actually increases, it should be questioned whether human emotionality should be exploited for purposes like that (cf. Manzeschke et al. 2016).

3.2 Models and Approaches

Based on the first findings on social reactions towards speaking machines and IVAs diverse models have been formulated which try to explain the origin of these social reactions (for an earlier summary see Krämer, Rosenthal-von der Pütten, & Hoffmann, 2015). The most prominent of them is the media equation approach: Reeves and Nass (1996) postulate that individuals treat computers and other artifacts as social actors, and term this phenomenon 'media equation' (media equals real life). To test this assumption, they conducted numerous experiments within the Computers Are Social Actors (CASA) paradigm, in which human subjects had to interact with computers. These experiments all followed a similar pattern: Search for a social science finding, replace 'human' with 'computer' in the theory statement and method, and observe whether the social rule is still observable (Nass et al., 1994). The media equation is considered to be verified when the results resemble the findings from interpersonal contexts.

Different approaches which try to explain social reactions towards non-social artifacts have been mentioned in the literature. These approaches can generally be divided into (a) approaches which assume that the reactions cannot be considered as truly social as they either result from a deficit on the part of the user or are conscious reactions due to demand characteristics of the situation, and (b) approaches which argue that social reactions towards artifacts occur unconsciously and are even denied by the human interlocutor (mindlessness, ethopoeia, computer as source). Additionally, (c) it has been suggested that social reactions depend on the level of assumed agency (Threshold Model of Social Influence).

Approaches which assume that reactions are not truly social

Early criticism with regard to the CASA study results stated that users who react in a social way must have deficits resulting from psychological dysfunctions, young age or lack of experience. However, this can be denied, as the participants in the CASA studies were mostly healthy undergraduate students who had extensive experience with computers (e.g. Nass et al., 1999; Nass, Fogg, & Moon, 1996). Similarly, the notion that the participants assume that they are interacting rather with the programmer than with the computer, i.e. the computer is not seen as a source but merely as a medium, has been refuted and rejected based on empirical data (Sundar, 1994; Sundar & Nass, 2000). A further explanation that continues to compete against the explanations below lies in the assumption that the observable reactions are not really "social" but are merely due to demand characteristics of the situation. It is argued that people show 'as if' reactions (in the sense that people consciously tell themselves that they will talk to the artificial entity as if it were a person) that merely occur because appropriate scripts are missing when humans interact with computers (Kiesler & Sproull, 1997). However, this assumption can at least be questioned based on some of the results on social reactions which obviously happen unconsciously (Krämer et al., 2013).

Approaches which assume that reactions are social and unconscious

Supporters of the media equation assumption see social reactions to artificial entities as truly social in the sense that "People respond socially and naturally to media even though they believe it is not reasonable to do so, and even though they don't think that these responses characterize themselves." (Reeves & Nass, 1996, p. 7). Nass and Moon (2000) suggest using the term 'ethopoeia' as an explanation for this unconscious and automatic behavior (social reaction) which is inconsistent with one's conscious opinion (computers do not need social treatment). According to this approach, minimal social cues like a human-sounding voice mindlessly (cf. Langer, 1989) trigger social responses because humans cannot avoid reacting automatically to social cues. The ethopoeia approach is supported by the fact that participants in the studies of Nass et al. obviously did not consciously recognize their social behaviors and when they were asked in the debriefing, they stated that they did not act socially (e.g. polite) towards the computers and that they believe such behavior to be inappropriate (Nass et al., 1999). Studies suggest that social responses to computers are indeed moderated by the extent of cognitive busyness (Lee, 2010). Also, it has been demonstrated that anthropomorphism of computers can happen mindlessly, leading to the social reactions which have been described (Kim & Sundar, 2012). In a revised version of the ethopoeia concept authors acknowledge that automatic and unconscious social reactions will be stronger if there are more social cues (Morkes, Kernal, & Nass, 1999; Nass & Yen, 2012) – speaking to the possibility that socially interactive agents might be more prone to trigger social reactions than speaking computers.

Approaches which assume that social reactions depend on the level of assumed agency

In the course of the development of anthropomorphic characters, Blascovich (2002) established the Threshold Model of Social Influence (TMSI), which predicts the social verification of a virtual other depending on the factors of agency and behavioral realism. Agency here means the degree to which the virtual entity is controlled by a real human (low agency is present in the case of a virtual agent that is controlled by an autonomous computer program; high agency is given when a human controls the virtual character – which is then termed avatar). The authors assume a Threshold of Social Influence, which has to be crossed to evoke social reactions by the user. This is only possible when the level of social verification is sufficiently high. When the factor of agency is high (i.e. when the user knows that the virtual character is a representation of a human being), then the factor of behavioral realism does not have to be high in order for social verification to take place and for social effects to occur. Conversely, when the factor of agency is low (i.e. when the user knows that the virtual character is a mere computer program), the factor of behavioral realism has to be very high to compensate for the lack of agency. In sum, it can be derived that according to the TMSI, the social influence of real persons will always be high, whereas the influence of an artificial entity depends on the realism of its behavior.

In our own research, we were especially interested in testing the ethopoiea model and the TMSI model against each other. The major difference is that the TMSI model assumes that there is a fundamental difference between agents and avatars in the sense that users react socially to avatars (i.e. mediated fellow humans) but will only react socially to agents when they show sufficient social cues. The ethopoiea model, on the other hand, assumes that agents will automatically evoke social reactions in the same way as fellow humans do. Therefore, Von der Pütten, Krämer, Gratch, and Kang (2010) empirically tested the TMSI against the ethopoeia approach. With the aim of testing the aforementioned assumptions, agency and behavioral realism of a virtual agent (the Rapport Agent; Gratch et al., 2006) were experimentally manipulated in a 2x2 between-subjects design. Participants were led to believe that they would be interacting either with another participant mediated by a virtual character or with an autonomous computer program. Moreover, the agent with higher behavioral realism featured responsive nonverbal behavior while participants were interacting with the agent, whereas the agent in the low behavioral realism condition showed only idle behavior (breathing, eye blinking), but no responsive behaviors. According to the TMSI, interaction effects between agency and behavioral realism should occur (in the sense that social reactions are observable in both avatar conditions but only in the agent condition with high behavioral realism). However, if the ethopoeia concept in its revised version (which acknowledges that automatic and unconscious social reactions will be stronger if there are more social cues; Morkes, Kernal, & Nass, 1999; Nass & Yen, 2012) is more accurate, social reactions should be reinforced when behavioral realism increases and should be independent of assumed agency. During the interaction, the Rapport Agent asked the participants intimate questions so that self-disclosure behavior of the participants could be used as dependent variable. Additionally, self-report scales to evaluate the virtual character as well as the situation were employed.
The data analyses revealed that the belief of interacting with either an avatar or an agent resulted in barely any differences with regard to the evaluation of the virtual character or behavioral reactions, whereas higher behavioral realism affected both (e.g., participants experienced more feelings of mutual awareness, and they used more words during the interaction when behavioral realism was high). However, no interaction effects of the factors of agency and behavioral realism emerged. Ultimately, since main effects of behavioral realism, but no interaction effects, were found, the results support the Revised Ethopoeia Concept but not the TMSI. However, it should be noted that a recent meta-analysis by Fox et al. (2015) rather provided evidence for the notion that the perception of agency is decisive when interacting with virtual characters. The analysis revealed that, overall, social reactions were stronger when people thought that they were interacting with another human compared to when they believed they were interacting with a computer program. Therefore, the role of agency in terms of the emergence of social reactions is still unclear.

Regarding ethical models, there are several theories and philosophical assumptions which are relevant here. There is, for example, a strong moral claim that human beings should have the control in any socio-technical arrangement (cf. Sturma 2004). Otherwise responsibility is hard to assign (cf. Christen 2004). There is a vivid debate on how to assign responsibility to a technical system. The EU Commission fostered the idea of a »electronic personhood« for „the most sophisticated autonomous robots" to cover claims for compensation when autonomous systems cause any damage. However, given the sometimes subtle effects (as described above) it might be difficult to directly detect damage. Therefore, it has been argued that something called "robot ethics 2.0" should be implemented (Lin, Jenkins & Abney, 2017; Misselhorn 2018).

3.3 History / Overview

From an ethical point of view, agency is not only behavior but also the ability of setting oneself goals and choosing means to achieve these goals. Furthermore, ethics deals with the justification of means and goals by

arguments and reasoning within the human society. Reducing agency to the ‚outer' behavior might lead to an impasse with regard to a richer understanding of agency. Expanding the circle of agents, specific animals, especially domestic animals, should be taken into consideration. Reacting to social cues does not necessarily mean to interact in a social way as is in human interaction. Humans interact with animals and they know quite well that this social interaction is different from those with human beings. It is a social habit which has been established within some thousand years. From this anthropological and ethical point of view it seems to be reasonable to contrast interaction of human-human relation, human-animal relation and human-technology relation. Thus, a specific difference appears between the first two relations and the latter one: There is not yet a practice of interaction between human beings and technological agents and, thus, a lack of experience and habitualization. Secondly, these technological agents are constructed by human beings and depend on human presets regarding design and interaction.
Besides the work focusing on social reactions summarized above, the last 20 years have seen a wealth of studies on factors which affect agent acceptance and ultimately also social reactions (for an earlier review see Krämer, Rosenthal-von der Pütten & Hoffmann, 2015). Influencing factors include attributes of the agents (agent behavior and appearance) as well as attributes of the user (gender, expertise, personality). In the following, first, agent attributes such as communicative behavior, nonverbal behavior and appearance will be discussed and, second, the role of user attributes will be reflected on.

Influence of agent attributes

Communicative behavior of the agent: As early as 2000, Rickenberg and Reeves showed that the (communicative) behavior of an agent is decisive. They demonstrate that whether a virtual character on a website monitored the user or ignored him/her had an impact on the user´s perceived anxiety and conclude that it is not sufficient "to focus on whether or not an animated character is present. Rather the ultimate evaluation is similar to those for real people, it depends on what the character does, what it says and how it presents itself" (p. 55). Indeed, further studies indicate that, for example, the quantity of the agent´s communicative utterances is influential. It was found that an interview agent´s self-disclosure (quality of utterances) lead only to minor effects, while the agent´s verboseness (quantity of utterances) affected both the participants' verbal behavior (with regard to word usage and intimacy of answers) and their perception of the interview (von der Pütten, Klatt, Hoffmann, & Krämer, 2011). Participants more often disclosed specific embarrassing situations, their biggest disappointment and what they feel guilty about to the agent regardless of its previous self-disclosure. Moreover, when the agent was more talkative it was generally evaluated more positively and the interview was perceived as being more pleasant. It can therefore be assumed that talkativeness led to a more favorable evaluation by the users and subsequently facilitated self-disclosure and thereby social reactions.

Nonverbal communicative features of agents: Traditionally, there has been more research on nonverbal compared to verbal behavior. For example, Krämer, Simons, and Kopp (2007) demonstrated that when the IVA Max showed self-touching gestures (e.g. touching its arm or face) this had positive effects on the experiences and evaluations of the user, whereas eyebrow-raising evoked less positive experiences and evaluations in contrast to no eyebrow-raising. Based on the notion that gestures influence the perception of competence (Maricchioloa, Gniscib, Bonaiutoa, & Ficcab, 2009), further research showed that when manipulating extensiveness of gesture usage and gender of a leader a positive impact of extensive nonverbal behavior is observable. Participants were more willing to hire the virtual person who used hand and arm gestures than the more rigid person. The virtual person using gestures was also perceived as exhibiting more leadership skills and general competence than the person in the non-gesture condition (Klatt, Haferkamp, Tetzlaff, & Krämer, 2012). The effects and efficiency of nonverbal behavior has especially been debated in the area of pedagogical agents. Here, researchers have discussed whether nonverbal behavior is decisive for learning experiences and that it therefore needs to be implemented in pedagogical agents. Baylor and Ryu (2003), for example, suggest that the key advantage of embodied pedagogical agents is that human-likeness creates more positive learning experiences and provides a strong motivating effect. However, Rajan et al. (2001) demonstrated that it is first and foremost the voice that is responsible for these effects. Moreno (2003) further summarized that – in line with results that especially the voice is decisive - there is no evidence for the social cue hypothesis as it has not been shown that the mere presence of social aspects such as a human-like body leads to distinct effects. However, the cognitive guiding functions provided by vocalizations and a program's didactic concept proved to be influential. Also, recent research (Carlotto & Jaques, 2016) as well as a recent meta-analysis (Schroeder & Adesope, 2014) have supported the notion that voice is more important than nonverbal expressiveness. Still, as Krämer (2017) argues, these results have to be considered with caution given the fact that the systems that had been evaluated did not (yet) include very sophisticated nonverbal behavior. It needs to be considered that nonverbal behavior is very complex: The dynamics of the movements are important, very subtle movements have distinct effects (e.g., head movements such as a head tilt) and the effects are context-dependent (e.g., a smile leads to a different effect when accompanied by a head tilt). This complexity, however, is mostly not implemented in pedagogical agents. So far, only very few pedagogical agent systems have achieved realistic and sufficiently subtle nonverbal behavior in order to administer a fair test. And indeed, when employing technology that provides realistic, dynamic nonverbal behavior, results show that nonverbal rapport behavior leads to an increase in effort and performance (Krämer et al., 2016). Therefore, the conclusion that embodiment and nonverbal behavior is less decisive compared to voice is premature.

More recently, specific nonverbal behavior of robots have been analysed (Hoffmann, 2017). In a series of studies it was investigated whether the positive effects of interpersonal touch are also observable with regard to robot touch. Based on media equation assumptions (Reeves & Nass, 1996) an experimental study in which a robot either touched or did not touch a human interaction partner revealed positive emotional reactions towards robot-initiated touch as well as increased compliance to the robot´s suggestions.

In conclusion, numerous nonverbal features have been demonstrated to influence the evaluation of the agent. As results largely mirror the results which have been found in human-human interaction the realm of nonverbal effects can also be seen as an area in which social effects have been corroborated.

Physical appearance: Several studies have shown that the appearance of a virtual character matters, for example for acceptance and evaluation of the character (van Vugt, Konijn, Hoorn, Keur, & Eliens, 2007; Domagk, 2010). In particular, Domagk (2010) shows that when the appearance (and voice) is likeable, a pedagogical agent has more positive effects. Building on this, we compared the impact of a virtual tutor depending on its appearance as either a cartoon-like rabbit character or a realistic anthropomorphic agent (Sträfling, Fleischer, Polzer, Leutner, & Krämer, 2010). Results show that the rabbit-like agent was not only preferred, but people exposed themselves to the tutoring session for longer when the rabbit provided feedback. However, this was not related to an increase in learning performance. Other studies, which focus more on credibility rather than learning and likeability, show that characters which are more anthropomorphic are perceived as more credible (Nowak & Rauh, 2005). More recent studies with more sophisticated appearances show that different appearances appeal to different groups of users: While students prefer non-human looking agents, specifically elderly users benefit from the social outcomes of a humanoid appearance (Straßmann, 2017). In addition, results demonstrate that attractive agents were found to be more likeable and were more persuasive. These effects, however, did not increase in a longterm relationship with an agent.

In sum, there is sufficient evidence to conclude that social effects also depend on aspects of physical appearance. However, given that different studies focus on different dimensions of appearance (e.g. realism, anthropomorphism, attractiveness/likeability), it is still difficult to conclude which physical features are decisive. A first attempt of systematizing the area is presented by Straßmann and Krämer (2017).

Influence of user attributes

Various user attributes have been considered as potential predictors of the reactions towards socially interactive agents, among them gender, age, computer literacy and personality. An overview is given in Krämer, Rosenthal-von der Pütten & Hoffmann (2015). Here, some examples will be summarized in order to demonstrate that these characteristics affect how users a) experience the interactions and b) perceive and evaluate artificial entities.
Gender. Krämer, Hoffmann, and Kopp (2010) revealed in their re-analysis of earlier studies that men and women have different preferences with regard to IVAs. In fact, compared to the effects of age and computer literacy, the influence of gender was more important. In one study, women were found to be more nervous during the interaction with the agent, which contradicts the vision that IVAs will facilitate human-computer interaction for these kinds of users. The data suggest further that female users´ interest and acceptance can be increased when nonverbal behaviors are implemented (here: self-touching gestures) and when the agent frequently smiles. Interestingly, Krämer et al. (2016) demonstrate with regard to pedagogical agents nonverbal behaviors communicating rapport were especially beneficial when displayed by agents of the opposite sex. In sum, it can be concluded that women especially benefit from an increased nonverbal behavior of the agent, in line with the finding that women are more sensitive to nonverbal behaviors (Hall, 1984).

Age. It is important to analyze older users´ reactions as more and more technology is developed to enable ambient assisted living for seniors. Part of this development are also multiple virtual agent applications (see, for example Kopp et al., 2018). Although the overall goal is that an IVA leads to a facilitation of human-technology-interactions studies suggest that older persons are more nervous when interacting with an IVA than younger ones (Krämer et al., 2007). Further results show that empathic nonverbal behavior can be helpful (Hosseinpanah, Krämer, & Straßmann, 2018). Interestingly, an agent that behaves in a dominant way leads to more persuasion when interacting with elderly users (Rosenthal-von der Pütten et al., 2019). With regard to appearance variables, older people seem to prefer more humanoid appearances (Straßmann & Krämer, 2017, 2018).
Computer literacy. Computer novices proved to be more nervous when interacting with an IVA than other users (Krämer, Hoffmann, and Kopp, 2010). This is in line with previous findings that computer laypeople do not benefit from IVAs in the way in which it is typically hoped (Krämer, Rüggenberg, Meyer zu Kniendorf, & Bente, 2002). Additional research will need to demonstrate under which conditions, non-computer-literate users can be supported in their interactions with IVAs.
Personality. Personality traits (as for example the so called Big Five personality traits, extraversion, neuroticism, conscientiousness, openness and agreeableness) have long been discussed as potential influencing factors in human-agent-interactions. However, the Big Five themselves seem to have only limited exploratory value: Results of a study with the Rapport Agent show that participants´ personality traits affected their subjective feelings after the interaction, as well as their evaluation of the agent and their actual behavior (von der Pütten, Krämer, & Gratch, 2010). From the various personality traits, those traits which relate to withstanding behavioral patterns in social contact (agreeableness, extraversion, approach avoidance, self-monitoring sensitivity, shyness, public self-consciousness) were found to be predictive for the positive and negative feelings participants reported after the interaction, the evaluation of the agent, and the amount of words they used during the conversation. However, other personality traits (e.g. openness, neuroticism) as well as gender and age did not affect the evaluation. For instance, the higher one's rating on extraversion and public self-consciousness, the more words were used. Furthermore, the more shy people are, the more negatively they evaluate the agent, whereas agreeableness increases positive feelings after the interaction.

In conclusion, several personality characteristics were shown to influence how people experience interactions with artificial entities. Whether social reactions are also directly influenced by user attributes has not yet been analysed in detail. The fact that users´ experiences are different depending on their attributes does not necessarily mean that social reactions are altered – especially when these are assumed to be unconscious reactions that are deeply-rooted in humanity´s social nature.

In parallel to these empirical studies, ethical guidelines have been developed. While researchers in psychology are first and foremost interested in understanding the mechanisms, they usually refrain from deriving normative conclusions in the sense of recommendations which effects should be avoided and which should be strived for. Psychological research is of course prone to – based on empirical results - describe potential risks and chances but would usually not use this to prescribe normative guidelines. On the contrary, ethics emphasizes the fact that „[t]he things we call ‚technologies' are ways of building order in our world. Many technical devices and systems important in everyday life contain possibilities for many different ways of ordering human actvity. Consciously or unconsciously, deliberately or inadvertently societies choose structures for technologies that influence how people are going to work, communicate, travel, consume and so forth over a very long time. […] In that sense technological innovations are similar to legislative acts or political foundings that establish a framework for public order that will endure over many generations«. (Winner, 1980, S. 128f). Every invention, thus, unfolds its specific normativity. This is why research on ethics is needed to accompany the development of such guidelines and to help implement structures which can assist when developers want to receive ethical advice. Recently, positive experiences have been made in the so-called FAT (fairness, accountability and transparency) community. Here, AI researchers have used ethical expertise to develop guidelines on how to ensure fairness, accountability and transparency in machine learning (https://www.fatml.org/). Similarly, the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems have suggested general principles of *ethically aligned design*. In addition, edited books give an overview of practically usable ethical aspects in the sense of value sensitive design (van den Hoven, Vermaas & van de Poel, 2015).

3.4 Similarities and Differences in IVAs and SRs

With increasing progress in the development of social robots, the question of whether IVAs and robots as physically embodied characters differ in their effects has become prevalent. The emergence of social robots is accompanied by a larger public debate which might lead to the conclusion that humans are more fascinated but at the same time more appalled and – potentially – frightened by social robots. Therefore, it is important to ask what the differences and similarities regarding the effects of virtual agents and robots are.

Surprisingly, however, there is not much controlled research on this question which directly compares robots and agents. Most studies have focused on obvious characteristics of robots versus agents: Robots are generally taller than characters on a screen, which might lead to more favorable perceptions of the robot (Powers, Kiesler, Fussel, & Torrey, 2007). Also, the "physical proximity" (Powers et al., 2007) might play a role. Whereas virtual characters are only graphical 2D or 3D representations on a monitor, robots have a material embodiment. Accordingly, they are able to physically touch humans, carry things and move on the ground. Therefore, it has been assumed that robots are perceived more as autonomous 'living' systems than virtual characters, meaning that more social presence might be sensed when facing them (Jung & Lee, 2004). Although some studies indeed demonstrate that robots elicit stronger experiences and social behaviors than agents (Bartneck, 2002; Kiesler, Powers, Fussel, & Torrey, 2008), this has not been shown consistently (Wainer, Feil-Seifer, Shell, & Matarić, 2007; Yamato, Shinozawa, Naya, & Kogure, 2001).
Hoffmann and Krämer (2013) identified that one reason for these inconsistencies might be the lack of comparability due to different operationalizations of the embodiments (e.g. different robots, animation, video recordings, etc.), dependent variables and interaction scenarios. In particular, findings by Shinozawa, Naya, Yamato and Kogure (2005) suggest that the interaction scenario and task might be decisive. Therefore, Hoffmann and Krämer (2013) conducted a 2x2 between-subjects experiment in which embodiment (rabbit-shaped Nabaztag robot vs. virtual version) and the interaction scenario (cognitive task vs. persuasive conversation) were systematically varied. For the purpose of the study, a three-dimensional virtual character was designed that resembles the rabbit-shaped robot that was used and displays the same voice and behavior. Two different interaction scenarios were created: a persuasive conversation about health habits (cf. Kiesler et al., 2008), and a task-oriented scenario in which participants had to solve a Towers of Hanoi puzzle, which was set on the table, under the guidance of the artificial counterpart. As dependent variables, subjective evaluation criteria (affective state, perceived social presence, attractiveness, and general evaluation of the interaction) and objective measures were assessed. In the conversational scenario, persuasion was analyzed as objective measurement by means of the amount of healthy food participants ate after the interaction. In the task-oriented setting, the amount of moves to solve the Towers of Hanoi task was counted as performance measure.
However, contrary to the hypotheses assuming that more social presence should be experienced in the presence of the physical robot, no differences emerged. Moreover, no differences caused by the embodiment were observable with respect to the participants' affective state, acceptance of the artificial entity, performance or persuasion. Still, two main effects of embodiment emerged: First, participants perceived the robot as more competent with regard to the fulfillment of tasks. The second main effect of embodiment occurred for the factor of control. Participants stated that they perceived more control during the interaction with the screen animation than with the robot. Additionally, a significant interaction between embodiment and interaction scenario with regard to the fulfillment of tasks emerged. In line with the findings of Shinozawa et al. (2005), task-oriented attractiveness was higher for the robot in the Towers of Hanoi condition, whereas it was higher for the screen animation in the conversational scenario.
In summary, the results of this study suggest that virtual characters can be used instead of more expensive robots when the aim of the application is of a persuasive nature. For scenarios in which physical manipulation is necessary at least on the side of the user, robots seem to be beneficial because they share the space of reference. Most importantly, the study underlines the importance of the consideration of different contexts (i.e. task or interaction scenario) while analyzing the impact of different embodiments. Whether social effects (like persuasion by an artificial companion) can be observed will therefore depend not only on the form of embodiment alone but also on the appropriateness of the specific embodiment for the specific task or scenario (see Krämer, Rosenthal-von der Pütten, & Hoffmann, 2015).

There are other studies - focusing only on robots – which give evidence that robots imply specific social reactions. For example, Horstmann et al. (2018) investigate people´s reactions when they are confronted with a robot that objects against being switched off. When the robot voiced an objection, a number of people let the robot switched on. When asked why they decided to do so, the majority of these participants said they felt sorry for the robot since it had told them about its fear of the darkness and that they did not want the robot to be scared ("He asked me to leave him on, because otherwise he would be scared. Fear is a strong emotion and as a good human you do not want to bring anything or anyone in the position to experience fear", m, 21). Almost as many participants explained that they did not want to act against the robot's will, which was expressed through the robot's objection to being switched off ("it was fun to interact with him, therefore I would have felt guilty, when I would have done something, what affects him, against his will.", m, 21). In addition, participants mentioned being surprised or being afraid of doing something wrong. In a second condition it was manipulated whether the style of the interaction was either social (mimicking human behavior) or functional (displaying machinelike behavior). After the functional interaction, people evaluated the robot as less likeable, which in turn led to a reduced stress experience after the switching off situation. However, individuals hesitated longest when they had experienced a functional interaction in combination with an objecting robot. This unexpected result might be due to the fact that the impression people had formed based on the task-focused behavior of the robot conflicted with the emotional nature of the objection. In sum, the results show that the robot´s human-mimicking behavior had a surprisingly strong impact on the participants. Instead of dismissing the objection to be switched off as weird for a machine, they were largely affected emotionally.

Similarly, Rosenthal-von der Pütten et al. (2014) show that humans are emotionally affected when they see that a robot is being "tortured". Here, the reactions are also documented on a neural level: in an fMRI study, participants were presented videos showing a human, the toy dinosaur robot "Pleo" and an inanimate object (a green box), being treated in either an affectionate (e.g., caressing the skin) or in a violent way (e.g., hitting, being choked). Self-reported emotional states and functional imaging data revealed that participants indeed reacted emotionally when seeing the affectionate and violent videos. Overall, the patterns were similar for robot and human and differed from people´s reactions towards watching the box being caressed or tortured. While no different neural activation patterns emerged for the affectionate interaction towards both, the robot and the human, we still found differences in neural activity when comparing only the videos showing abusive behavior indicating that participants experience more emotional distress and show negative empathetic concern for the human in the abuse condition. This indicates that although robots evoke emotional reactions, there is still a difference to the – more intense – reactions towards humans.

In additional studies, it was investigated whether robots trigger different kinds of reactions depending on their human likeness. Here, according to the uncanny valley hypothesis, humans are expected to prefer anthropomorphic appearance but to reject the artificial entity when it becomes too human-like (but is not yet perfectly human). A study with 40 standardized pictures of robots with varying human-likeness, however, indicated that the relation of robot characteristics and evaluation is not best explained by a cubic function (which would be closest to the uncanny valley thesis) but rather by a linear function – indicating that the evaluation gets more positive with increasing human-likeness (Rosenthal-von der Pütten, 2014). An additional study employing fmri demonstrates that, on a neural level, areas which encode a linear humanlikeness continuum (temporoparietal junction) can be differentiated from areas which encode nonlinear representations and a human-nonhuman distinction (dorsomedial prefrontal cortex and fusiform gyrus) (Rosenthal-von der Pütten et al., 2020). If uncanny valley reactions happen, in the sense of a selective dislike of highly humanlike agents, it is based on nonlinear value-coding in the ventromedial prefrontal cortex, a key component of the brain´s reward system. In consequence, it can be concluded that a basic principle known from sensory coding—neural feature selectivity from linear–nonlinear transformation—may also underlie human responses to artificial social partners.

Given these results regarding the social effects of robots, it needs to be concluded that from the perspective of ethics social robots might be more problematic than IVAs in the sense that robots offer physical encounter and interaction. Physical interactions inscribe specific patterns of understanding oneself and the world. This is on the one hand due to their physical embodiment and the corresponding ability to actually *act* in a real environment. On the other hand, the results on social reactions suggest that they might be more powerful in eliciting emotional reactions – which might make humans more prone for manipulations. Henceforth human beings have to deal with the question what it means to co-operate with a robot in comparison with the cooperation with other human beings or animals. The answer to the question depends very much on how human beings will design the robots – which is, not least, a normative decision.

3.5 Current Challenges

Current challenges include the necessity to observe how social reactions change when human-machine interaction becomes increasingly ‚natural'. Although current systems – even when built on machine learning with millions of user utterances (see Siri or Alexa) – are not yet comparable with human-human-communication, further improvements in the near future are likely. Given that people already react to basic social cues, it is not clear yet whether social reactions will increase in line with better dialogue capabilities. Especially from an ethical perspective, however, it has to be discussed whether the observable social reactions actually entail “full” sociality. Even if empirical reactions are the same as those observable in human-human-interaction, it is questionable whether this indeed indicates that interactions with technology can be viewed as comparably social as interactions with fellow human beings or even animals.

Another challenge pertains to the question of how to further improve dialogue with machines. Here, psychology should contribute knowledge as well as methods. How cooperation in the area might yield synergies is described in Kopp and Krämer (submitted).

When systems become increasingly intelligent it is also important to ask what people know about and expect from human-like machines. At the moment, it is clearly visible that people derive their knowledge mainly from media. Based on the uncertainty reduction theory it was demonstrated that people’s experiences with robots in the media lead to high expectations regarding the skills of robots, which in turn increase people’s general expectations that social robots will be part of the society as well as of their personal lives. Furthermore, knowledge of negatively perceived fictional robots increases negative expectancies of robots becoming a threat to humans, while technical affinity reduces general robot anxiety. In sum, it can be concluded that especially fictional media material which depicts mighty robots which compete against or threaten humans lead to skepticism against artificial intelligence, machine learning and digitalization.

Based on these results and further evidence which indicates that the average human user is not able to grasp what intelligent algorithms and social robots can and cannot do, let alone how they basically function (DeVito et al., 2019; Krämer et al. 2019), it seems necessary to increase the computational literacy of the general public. In order to enable a truly informed consent to use socially interactive agents including their collecting and processing of personal data, people need to be taught the basic steps of intelligent algorithms. First studies which try to describe the mental models and concepts people developed suggest, however, that it is a long way to go to achieve at least a basic understanding which enables informed decisions of whether to take or avoid the risks of using an intelligent interlocutor (Ngo, Kunkel & Ziegler, 2020). Still, from an ethical and also from an empirical point of view, it needs to be asked whether this knowledge will indeed prevent us from misleading perceptions and presuppositions.

This area also poses ethical questions: From an ethical perspective, one can question whether people should be better informed about the artificial nature and the algorithms. It might be doubted that, for example, people can protect themselves better against unconscious social reactions if they understand more about the functioning. Also, for other technology, it is well known that people use technology without knowing in detail how it actually functions. There is a way of familiarizing oneself with a technology without knowing its details – and without needing to take the effort to learn something new or to inform oneself. Becoming familiar with a technology means to establish habits and routines which allows to interact with the technology. Hans Blumenberg, a German philosopher claimed that this acquaintance is one side of the coin, the other one is a loss of reflection and even »Sinnverzicht« [renunciation of meaning].

3.6 Future Directions

From a psychological point of view the most important research question for future research is to observe whether social reactions become stronger when machines get increasingly intelligent and show improved dialogue abilities (see above). One methodological strategy that will be vital in these investigations are longterm measurements. Also, it needs to be asked whether people will react with less social reactions when they understand better what the nature of intelligent algorithms and machines is. Both aspects need to be looked at in (experimental) field studies. From an anthropological and ethical point of view it will be indispensable to investigate the deeper meaning of ‚intelligence‘ when attributing it to human beings as a precondition of agency, decision, responsibility and so on. On the one hand, we have take into account that not all human beings have these capacities and still count as vis-à-vis in human interaction. On the other hand, we should acknowledge that intelligence with regard to machines differs from intelligence of human beings. The understanding of this difference must not necessarily be based on detailed knowledge of algorithms, machine learning or bayesian networks but the ontological difference might be illustrated by other, ‚simpler‘ means.

Most of the studies conducted so far have been experimental laboratory studies which tried to increase internal validity. However, in order to learn whether social effects of artificial entities are actually relevant in humans' everyday encounters with agents or robots, field studies and long-term studies also have to be conducted, which provide higher external validity. Field studies with an open, exploratory approach are also important in order to be able to identify new aspects of human-agent or human-robot interaction which can then be checked in controlled laboratory studies. Contrary to mainly quantitative laboratory studies, field studies can also be conceptualized in a way that makes inter-individual differences more visible: Here, one study described the emergence of huge inter-individual differences as one of the most important results of a long-term study in which six elderly participants interacted with a robot serving as a health advisor (von der Pütten, Krämer, & Eimler, 2011). Following a multi-methodological approach, the continuous quantitative and qualitative description of user behavior on a very fine-grained level gave insights into when and how people interacted with the robot companion. Post-trial semi-structured interviews explored how the users perceived the companion and revealed their attitudes. Based on this large data set, it was found that users are willing to start interactions and even conversation with a robot even though its perceptive and expressive capabilities are limited. Although aware of the fact that they are interacting with an artificial being, some of the users built relationships with the robot, while others treated the robot as a tool. Some people tended to like the robot as long as it was helping and they had a feeling of being in control. Others seemed to integrate the companion into their life, even though it did not always work properly and was perceived as being of limited use. Thus, the fine-grained analysis showed a particular prevalence of idiosyncratic reactions. Therefore, future research should extend field and long-term studies in order to identify patterns and factors that determine the conditions under which people show social reactions and whether they ultimately even develop relationships with artificial entities. As an experimental factor, information on the advance knowledge on the functioning of the machine can be provided to only half of the participants.

Also, future research needs to foster more interdisciplinary research. In the research realm described here, especially a combination of psychology, computer science, ethics and law is relevant (Krämer et al., 2019).

3.7 Summary

As concluding message we would like to repeat that numerous studies demonstrate that people react socially towards artificial entities – as soon as these display social cues. It still has to be analyzed in greater depths whether the availability of the number of social cues affects the degree of social reactions and whether social robots indeed lead to larger effects compared to IVAs. Independent of these questions, the majority of reactions seem to happen unconsciously and this suggests that reactions are based on human nature and human´s unique sociality. This is in line with assumptions that humans are indeed "free monadic radicals" who are persistently searching for a social partner to interact and bond with (Kappas, 2005). Especially due to the unconsciousness and probable inevitability of the reactions, it has to be asked from an ethical point of view whether people have to be protected against manipulations based on social influence – potentially by keeping in mind that is human design which affects people and may manipulate them consciously or inadvertently. Also, it needs to be asked whether computational literacy in the sense of more knowledge about intelligent algorithms and the basic functioning of intelligent conversing machines will help to reduce social effects. This, however, can just as well be doubted as it can be doubted that trying to impose more knowledge is desirable from an ethical point of view.

References

Bartneck, C. (2002). *eMuu - an embodied emotional character for the ambient intelligent home* [Doctoral dissertation, Eindhoven University of Technology]. UC Research Repository. http://www.bartneck.de/publications/2002/eMuu/bartneckPHDThesis2002.pdf

Baylor, A. L., & Ryu, J. (2003). The effects of image and animation in enhancing pedagogical agent persona. *Journal of Educational Computing Research, 28*(4), 373–394. https://doi.org/10.2190/V0WQ-NWGN-JB54-FAT4

Blascovich, J. (2002, September 30–October 2). *A theoretical model of social influence for increasing the utility of collaborative virtual environments.* Proceedings of the 4th International Conference on Collaborative Virtual Environments (pp. 25–30), Bonn, Germany. https://doi.org/10.1145/571878.571883

Blumenberg, H. (1959). Kontingenz. In K. Galling (Ed.), *Religion in Geschichte und Gegenwart* [Religion and history in the present age] (3rd ed., Vol. 3, Sp. 1793–1794). Mohr.

Carlotto, T., & Jaques, P. A. (2016). The effects of animated pedagogical agents in an English-as-a-foreign-language learning environment. *International Journal of Human-Computer Studies, 95*, 15–26. https://doi.org/10.1016/j.ijhcs.2016.06.001

Christen, M. (2004). Schuldige Maschinen? Autonome Systeme als Herausforderung für das Konzept der Verantwortung [Guilty machines? Autonomous systems as a challenge for the concept of responsibility]. In L. Honnefelder & C. Streffer (Eds.), *Jahrbuch für Wissenschaft und Ethik* (Vol. 9, pp. 163–191). Walter de Gruyter.

de Waal, F. (2019). *Mama's Last Hug: Animal Emotions and What They Tell Us about Ourselves.* Norton & Company.

DeVito, M. A., Birnholtz, J., Hancock, J. T., French, M., & Liu, S. (2018). How people form folk theories of social media feeds and what it means for how we study self-presentation. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 311-314). New York, USA: ACM. doi:10.1145/3173574.3173694

Domagk, S. (2010). Do pedagogical agents facilitate learner motivation and learning outcomes? The role of the appeal of agent's appearance and voice. *Journal of Media Psychology, 22*(2), 84–97. https://doi.org/10.1027/1864-1105/a000011

Fogg, B. J., & Nass, C. (1997). Silicon sycophants: the effects of computers that flatter. *International Journal of Human-Computer Studies, 46*(5), 551–561. https://doi.org/10.1006/ijhc.1996.0104

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 effects of agency on social influence. *Human-Computer Interaction, 30*, 401-432. doi: 10.1080/07370024.2014.921494

Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R. J., & Morency, L.-P. (2006). Virtual rapport. In J. Gratch, M. Young, R. Aylett, D. Ballin, & P. Olivier (Eds.), *Lecture notes in computer science: Vol. 4133*. *Intelligent virtual agents* (6th ed., pp. 14–27). Springer. https://doi.org/10.1007/11821830_2

Hall, J. A. (1984). *Nonverbal sex differences. Communication accuracy and expressive style.* Johns Hopkins University Press.

Hoffmann, L. (2017). *That robot touch that means so much. On the psychological effects of human-robot touch.* PhD-thesis, University Duisburg-Essen. Available: https://duepublico2.uni-due.de/receive/duepublico_mods_00043331

Hoffmann, L., & Krämer, N. C. (2013). Investigating the effects of physical and virtual embodiment in task-oriented and conversational contexts. *International Journal of Human-Computer Studies, 71*(7–8), 763–774. https://doi.org/10.1016/j.ijhcs.2013.04.007

Hoffmann, L., Krämer, N. C., Lam-Chi, A., & Kopp, S. (2009). Media equation revisited: Do users show polite reactions towards an embodied agent? In Z. Ruttkay, M. Kipp, A. Nijholt, & H. H. Vilhjálmsson (Eds.), *Lecture Notes in Computer Science: Vol. 5773*. *Intelligent virtual agents* (9th ed., pp. 159–165). Springer. 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

Hosseinpanah, A., Krämer, N. C., & Straßmann, C. (2018, December). *Empathy for everyone? The effect of age when evaluating a virtual agent.* Proceedings of the 6th International Conference on Human-Agent Interaction (pp. 184–190), Southampton, United Kingdom. https://doi.org/10.1145/3284432.3284442

IEEE | Institute of Electrical and Electronics Engineers. (2018). Ethically aligned design. A vision for prioritizing human well-being with autonomous and intelligent systems. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf

Jung, Y., & Lee, K. M. (2004, October 13–15). *Effects of physical embodiment on social presence of social robots.* Proceedings of PRESENCE (pp. 80–87), Valencia, Spain.

Kappas, A. (2005, July). *My happy vacuum cleaner* [Paper presentation]. ISRE General Meeting, Symposium on Artificial Emotions, Bari, Italy.

Kiesler, S. B., & Sproull, L. (1997). Social responses to "social" computers. In B. Friedman (Ed.), *Human values and the design of technology* (pp. 191–200). Cambridge University Press/CSLI.

Kiesler, S., Powers, A., Fussell, S. R., & Torrey, C. (2008). Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition, 26*(2), 169–181. https://doi.org/10.1521/soco.2008.26.2.169

Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior, 28*(1), 241–250. https://doi.org/10.1016/j.chb.2011.09.006

Klatt, J., Haferkamp, N., Tetzlaff, L., &. Krämer, N. C. (2012, June). *How to be… a leader - Examining the impact of gender and nonverbal behavior on the perception of leaders* [Paper presentation]. International Communication Association 62nd Annual Meeting, Phoenix, Arizona, United States.

Kopp, S. & Krämer, N. C. (submitted). Revisiting human-agent communication: The importance of incremental co-construction and understanding mental states. *Frontiers in Cognitive Science.*

Kopp, S., Brandt, M., Buschmeier, H., Cyra, K., Freigang, F., Krämer, N., Kummert, F., Opfermann, C., Pitsch, K., Schillingmann, L., Straßmann, C., Wall, E., Yaghoubzadeh, R. (2018). Conversational assistants for elderly users – The importance of socially cooperative dialogue. In E. André, T. Bickmore, S. Vrochidis, & L. Wanner (Eds.), *CEUR workshop proceedings: Vol. 2338. Proceedings of the AAMAS workshop on intelligent conversation agents in home and geriatric care applications co-located with the federated AI meeting* (pp. 10–17). RWTH.

Krämer, N. C. et al. (2019). *KI-basierte Sprachassistenten im Alltag: Forschungsbedarf aus informatischer, psychologischer, ethischer und rechtlicher Sicht.* [AI-based speech assistants in everyday-life: Research gaps from a computer-science, psychology, ethical and law perspective]. Policy Paper. Universität Duisburg-Essen. https://doi.org/10.17185/duepublico/70571

Krämer, N. C. (2005). Social communicative effects of a virtual program guide. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, & T. Rist (Eds.), *Lecture notes in computer science: Vol. 3661. Intelligent virtual agents* (5th ed., pp. 442–543). Springer. https://doi.org/10.1007/11550617_37

Krämer, N. C. (2017). The immersive power of social interaction. In D. Liu, C. Dede, R. Huang, & J. Richards (Eds.), *Smart computing and intelligence: Vol. 1. Virtual, augmented, and mixed realities in education* (pp. 55–70). Springer. https://doi.org/10.1007/978-981-10-5490-7_4

Krämer, N. C., Bente, G., & Piesk, J. (2003). The ghost in the machine. The influence of embodied conversational agents on user expectations and user behaviour in a TV/VCR application. In G. Bieber & T. Kirste (Eds.), *IMC workshop, assistance, mobility, applications* (pp. 121–128*).* IRB Verlag.

Krämer, N. C., Hoffmann, L., & Kopp, S. (2010). Know your users! Empirical results for tailoring an agent´s nonverbal behavior to different user groups. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, & A. Safonova (Eds.), *Lecture notes in computer science: Vol. 6356. Intelligent virtual agents* (10th ed., pp. 468–474). Springer. http://doi.org/10.1007/978-3-642-15892-6_50

Krämer, N. C., Karacora, B., Lucas, G., Dehghani, M., Rüther, G. & Gratch, J. (2016). Closing the gender gap in STEM with friendly male instructors? On the effects of rapport behavior and gender of a virtual agent in an instructional interaction. *Computers & Education, 99*, 1–13. https://doi.org/10.1016/j.compedu.2016.04.002

Krämer, N. C., Kopp, S., Becker-Asano, C., & Sommer, N. (2013). Smile and the world will smile with you – The effects of a virtual agent's smile on users' evaluation and behavior. *International Journal of Human-Computer Studies, 71*(3), 335–349. http://doi.org/10.1016/j.ijhcs.2012.09.006

Krämer, N. C., Rosenthal-von der Pütten, A. M., & Hoffmann, L. (2015). Social effects of virtual and robot companions. In S. S. Sundar (Ed.), *The handbook of the psychology of communication technology* (1st ed., pp. 137–159). John Wiley & Sons. https://doi.org/10.1002/9781118426456.ch6

Krämer, N. C., Rüggenberg, S., Meyer zu Kniendorf, C., & Bente, G. (2002). Schnittstelle für alle? Möglichkeiten zur Anpassung anthropomorpher Interface Agenten an verschiedene Nutzergruppen. In M. Herczeg, W. Prinz, & H. Oberquelle (Eds.*), Berichte des German Chapter of the ACM: Vol. 56. Mensch und Computer* (pp. 125–134). Teubner. http://doi.org/10.1007/978-3-322-89884-5_13

Krämer, N. C., Simons, N., & Kopp, S. (2007). The effects of an embodied conversational agent's nonverbal behavior on user's evaluation and behavioral mimicry. In C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis, & D. Pelé (Eds.), *Lecture notes in computer science: Vol 4722. Intelligent virtual agents* (7th ed., pp. 238–251). Springer. http://doi.org/10.1007/978-3-540-74997-4_22

Langer, E. J. (1989). *Mindfulness*. Addison-Wesley.

Leary, M. R. (1995). *Self-presentation: Impression management and interpersonal behavior.* Brown & Benchmark Publishers.

Lee, E.-J. (2010). What triggers social responses to flattering computers? Experimental tests of anthropomorphism and mindlessness explanations. *Communication Research, 37*(2), 191–214. https://doi.org/10.1177/0093650209356389

Lin, P., Jenkins, R., & Abney, K. (2017*). Robot ethics 2.0. From autonomous cars to artificial intelligence.* Oxford University Press.

Manzeschke, A., Assadi, G., Karsch, F., & Viehöver, W. (2016). Funktionale Emotionen und emotionale Funktionalität. Über die neue Rolle von Emotionen und Emotionalität in der Mensch-Technik-Interaktion. In A. Manzeschke & F. Karsch (Eds.), *Roboter, Computer und Hybride. Was ereignet sich zwischen Menschen und Maschinen?* (pp. 107–130). Nomos. https://doi.org/10.5771/9783845272931-107

Maricchiolo, F., Gnisci, A., Bonaiuto, M., & Ficca, G. (2009). Effects of different types of hand gestures in persuasive speech on receivers' evaluations. *Language and Cognitive Processes, 24*(2), 239–266. https://doi.org/10.1080/01690960802159929

Misselhorn, C. (2018). Artificial Morality. Concepts, Issues and Challenges. *Society, 55*, 161–169. Doi: 10.1007/s12115-018-0229-y

Moreno, R. (2003). *The role of software agents in multimedia learning environments: When do they help students reduce cognitive load?* [Paper presentation]. European Association for Research on Learning and Instruction Annual Conference, Padova, Italy.

Morkes, J., Kernal, H. K., & Nass, C. (1999). Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of SRCT theory. *Human-Computer Interaction, 14*(4), 395–435. https://doi.org/10.1207/S15327051HCI1404_2

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

Nass, C., & Yen, C. (2012). *The man who lied to his laptop. What we can learn about ourselves from our machines.* Penguin.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies, 45*(6), 669–678. https://doi.org/10.1006/ijhc.1996.0073

Nass, C., Moon, Y., & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems. *Journal of Applied Social Psychology, 29*(5), 1093–1110. https://doi.org/10.1111/j.1559-1816.1999.tb00142.x

Nass, C., Moon, Y., Morkes, J., Kim, E-Y., & Fogg, B. J. (1997). Computers are social actors: A review of current research. In B. Friedman (Ed.), *Human values and the design of computer technology* (pp. 137–162). CSLI Publications.

Nass, C., Steuer, J., & Tauber, E. R. (1994). *Computers are social actors.* Proceedings of the SIGCHI Conference of Human Factors in Computing Systems (pp. 72–78), Boston, Massachusetts, United States. https://doi.org/10.1145/191666.191703

Ngo, T., Kunkel, J. & Ziegler, J. (in press). The Netflix Experience: Mental Models of Recommender Systems Exemplified by Netflix. In *Proceedings of the ACM UMAP, 28th Conference on User Modeling, Adaptation and Personalization*. New York: ACM.

Nowak, K. L., & Rauh, C. (2005). The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer-Mediated Communication, 11*(1), 153–178. https://doi.org/10.1111/j.1083-6101.2006.tb00308.x

Parise, S., Kiesler, S., Sproull, L., & Waters, K. (1999). Cooperating with life-like interface agents. *Computers in Human Behavior, 15*(2), 123–142. https://doi.org/10.1016/S0747-5632(98)00035-1

Plessner, H. (1970). *Philosophische Anthropologie. Lachen und Weinen. Das Lächeln. Anthropologie der Sinne* [Philosophical anthropology. Laughing and crying. Anthropology of the senses]. Frankfurt: Fischer.

Powers, A., Kiesler, S., Fussel, S., & Torrey, C. (2007). *Comparing a computer agent with a humanoid robot.* Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction (pp. 145–152), Arlington, Virginia, United States. https://doi.org/10.1145/1228716.1228736

Rajan, S., Craig, S. D., Gholson, B., Person, N. K., Graesser, A. C., & The Tutoring Research Group. (2001). AutoTutor: Incorporating back-channel feedback and other human-like conversational behaviors into an intelligent tutoring system. *International Journal of Speech Technology, 4*, 117–126. https://doi.org/10.1023/A:1017319110294

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.* Proceedings of the SIGCHI conference on Human Factors in

Computing Systems (pp. 49–56), The Hague, Netherlands. https://doi.org/10.1145/332040.332406

Rosenthal-von der Pütten, A., Krämer, N. C., Maderwald, S., Brand, M. & Grabenhorst, F. (2019). Neural Mechanisms for Accepting and Rejecting Artificial Social Partners in the Uncanny Valley. *Journal of Neuroscience, 39* (33), 6555-6570. DOI: 10.1523/JNEUROSCI.2956-18.2019

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. *Computers in Human Behavior, 33*, 201–212. https://doi.org/10.1016/j.chb.2014.01.004

Rosenthal-von der Pütten, A. M., Straßmann, C., Yaghoubzadeh, R., Kopp, S., & Krämer, N. C. (2019). Dominant and submissive nonverbal behavior of virtual agents and its effects on evaluation and negotiation outcome in different age groups*. Computers in Human Behavior, 90*, 397–409. https://doi.org/10.1016/j.chb.2018.08.047

Schroeder, N. L., & Adesope, O. O. (2014). A systematic review of pedagogical agents‘ persona, motivation, and cognitive load implications for learners. *Journal of Research on Technology in Education, 46*(3), 229–251. https://doi.org/10.1080/15391523.2014.888265

Shechtman, N., & Horowitz, L. M. (2003). *Media inequality in conversation: How people behave differently when interacting with computers and people.* Proceedings of the SIGCHI conference on Human Factors in Computing Systems (pp. 281–288), Fort Lauderdale, Florida, United States. https://doi.org/10.1145/642611.642661

Shinozawa, K., Naya, F., Yamato, J., & Kogure, K. (2005). Differences in effect of robot and screen agent recommendations on human decision-making. *International Journal of Human-Computer Studies, 62*(2), 267–279. https://doi.org/10.1016/j.ijhcs.2004.11.003

Sproull, L., Subramani, M., Kiesler, S. Walker, J. H., & Waters, K. (1996). When the interface is a face. *Human-Computer Interaction, 11*(2), 97–124.

Sträfling, N., Fleischer, I., Polzer, C., Leutner, D., & Krämer, N. C. (2010). Teaching learning strategies with a pedagogiocal agent. The effects of a virtual tutor and its appearance on learning and motivation. *Journal of Media Psychology, 22*(2), 73–83. https://doi.org/10.1027/1864-1105/a000010

Straßmann, C. (2017). *All eyes on the agent´s appearance?! Investigation of target-group-related social effects of a virtual agent's appearance in longitudinal human-agent interactions*. PhD-Thesis, University of Duisburg-Essen. Available: https://duepublico2.uni-due.de/receive/duepublico_mods_00070205

Straßmann, C., & Krämer, N. C. (2017). A categorization of virtual agent appearances and a qualitative study on age-related user preferences. In J. Beskow, C. Peters, G. Castellano, C. O'Sullivan, I. Leite, & S. Kopp (Eds.), *Lecture Notes in Computer Science: Vol. 10498. Intelligent virtual agents* (17th ed., pp. 413–422). Springer. https://doi.org/10.1007/978-3-319-67401-8_51

Straßmann, C., & Krämer, N. C. (2018). A two-study approach to explore the effect of user characteristics on users' perception and evaluation of a virtual assistant's appearance. *Multimodal Technologies and Interaction, 2*(4), 1–25. https://doi.org/10.3390/mti2040066

Sturma, D. (2004). Ersetzbarkeit des Menschen? Robotik und menschliche Lebensform. In L. Honnefelder & C. Streffer (Eds.), *Jahrbuch für Wissenschaft und Ethik* (Vol. 9, pp. 141–162). Walter de Gruyter.

Sundar, S. S. (1994, August 10–13). *Is human-computer interaction social or parasocial?* [Paper presentation]. Association for Education in Journalism and Mass Communication 77th Annual Meeting, Atlanta, Georgia, United States.

Sundar, S. S., & Nass, C. (2000). Source orientation in human-computer interaction: Programmer, networker, or independent social actor. *Communication Research, 27*(6), 683–703. https://doi.org/10.1177%2F009365000027006001

van den Hoven, J., Vermaas, P. E., & van de Poel, I. (Eds.). (2015). *Handbook of ethics, values, and technological design. Sources, theory, values and application domains.* Springer.

van Vugt, H. C., Konijn, E. A., Hoorn, J. F., Keur, I., & Eliéns, A. (2007). Realism is not all! User engagement with task-related interface characters. *Interacting with Computers, 19*(2), 267–280. https://doi.org/10.1016/j.intcom.2006.08.005

von der Pütten, A. M., Hoffmann, L., Klatt, J., & Krämer, N. C. (2011). Quid pro quo? Reciprocal self-disclosure and communicative accomodation towards a virtual interviewer. In H. H. Vilhjálmsson, S. Kopp, S. Marsella, & K. R. Thórisson (Eds.), *Lecture notes in computer science: Vol: 6895. Intelligent virtual agents* (11th ed., pp. 183–194). Springer. https://doi.org/10.1007/978-3-642-23974-8_20

von der Pütten, A. M., Krämer, N. C., & Gratch, J. (2010). How our personality shapes our interactions with virtual characters – Implications for research and development. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, & A. Safonova (Eds.), *Lecture notes in computer science: Vol: 6356. Intelligent virtual agents* (10th ed., pp. 208–221). Springer. https://doi.org/10.1007/978-3-642-15892-6_23

von der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S.-H. (2010). "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Computers in Human Behavior, 26*(6), 1641–1650. https://doi.org/10.1016/j.chb.2010.06.012

von der Pütten, A.M., Krämer, N.C., & Eimler, S.C. (2011). *Living with a robot companion – Empirical study on the interaction with an artificial health advisor.* Proceedings of the 13th International Conference on Multimodal Interaction (pp. 327–334), Alicante, Spain. https://doi.org/10.1145/2070481.2070544

Wainer, J., Feil-Seifer, D. J., Shell, D. A., & Matarić, M. J. (2007, August 26–29). *Embodiment and human-robot interaction: A task-based perspective.* Proceedings of the 16th IEEE International Conference on Robot and Human Interactive Communication (pp. 872–877), Jeju, Korea.

Winner, L. (1980). Do artifacts have politics? *Daedalus 1*, 109, 121–136.

Yamato, J., Shinozawa, K., Naya, F. & Kogure, K. (2001). *Evaluation of communication with robot and agent: Are robots better social actors than agents?* Proceedings of IFIP INTERACT01: Human-Computer Interaction (pp. 690–691), Tokyo, Japan.