To appear in Moallem A. (Ed.) Smart and Intelligent Systems, CRC Press (2021)

Social Factors in Human-Agent Teaming

Celso M. de Melo[1], Benjamin T. Files[1], Kimberly A. Pollard[1], and Peter Khooshabeh[1]

[1] CCDC Army Research Laboratory

Author Note

This research was supported by Mission Funds from the US Army. The content does not necessarily reflect the position or the policy of any Government, and no official endorsement should be inferred.

Correspondence concerning this article should be addressed to Celso M. de Melo, CCDC U.S. Army Research Laboratory, Playa Vista, CA 90094, United States. Email: celso.m.de.melo.civ@mail.mil.

Preprint Manuscript

Abstract

Recent decades have seen impressive progress in the development of autonomous technology, such as robots, drones, self-driving cars, and personal assistants. These intelligent agents are able to engage with their surrounding environment in increasingly sophisticated ways. However, as this technology becomes pervasive in society, its success hinges on effective and efficient collaboration with humans. To accomplish this, agents need not only understand the functional aspects of the task, but also the broader social context. Here, we first review relevant psychological theory explaining why and when humans treat agents in a social manner and are socially influenced by them. Second, we summarize experimental evidence showing the importance of verbal (e.g., natural language conversation) and nonverbal (e.g., emotion expressions) communication for successful collaboration between humans and agents. Third, we review recent work showing how perceptions of social group membership with agents influence cooperation. Fourth, we cover research on key individual differences – e.g., anthropomorphic tendency – shaping social interaction with agents. Finally, we identify open challenges and opportunities in this emerging field.

Social Factors in Human-Agent Teaming

The last two decades have seen an explosion of interest in autonomous agents – such as robots, drones, self-driving cars, and home assistants – and the expectation is that these agents will become even more pervasive in society in the future (Bonnefon et al., 2016; de Melo et al., 2019; Stone, & Lavine, 2014; Waldrop, 2015). As autonomous technology becomes integrated into our personal, social, and professional lives, humans will have to engage with it often and, in many cases, rely on it to accomplish their goals. In fact, human-agent teaming is expected to be critical to accomplishing the mission in increasingly complex and dynamic environments (Kott & Alberts, 2017; Kott & Stump, 2019). However, the success of these hybrid teams relies on a simple premise: humans will successfully collaborate with autonomous agents. But this premise should not be taken for granted. On the one hand, humans are very selective about with whom they cooperate, basing their decision on a multitude of factors including prior interaction, reputation, and shared group identity (Kollock, 1998; Rand & Nowak, 2013). On the other hand, many people may not trust autonomous technology, due for example to lack of experience and understanding of how it works (Gillis, 2017; Hancock et al., 2011; Lee & See, 2004). In this chapter, we review research supporting the argument that building machines that have appropriate social skills will encourage humans to treat them as social partners and, in turn, promote trust and cooperation in human-agent teams.

Humans are inherently social creatures. Our beliefs are influenced by our social context (Manstead & Fischer, 2001), we influence and are influenced by others (Van Kleef et al., 2010), we communicate with others to share information and synchronize our actions (Orbell et al., 1988), and we distinguish those that belong to our social groups from those that do not (Crisp & Hewstone, 2007; Tajfel & Turner, 1986). Moreover, we often anthropomorphize non-human others, and apply social heuristics learnt from interaction with other humans (Epley et al., 2007; Premack & Premack, 1995). Engaging in a social manner with non-human others supports intuitive explanations of others' behavior and brings familiar guidelines to unfamiliar situations (Reeves & Nass, 1996). For these reasons, several researchers argued for the development of social agents that, on the one hand, display cues that promote social engagement from humans and, on the other hand, simulate appropriate social behavior, including verbal and nonverbal communication and adapt behavior to the social context (Bates, 1994; Breazeal, 2003; Cassell, 2000; Gratch et al., 2002; Leite et al., 2013).

In this chapter, we review research on social factors that shape human-agent collaboration. We first look at the importance of natural language communication. Second, we emphasize the importance of nonverbal communication – in particular, emotion expression – to promote cooperation between humans and agents. Third, we review experimental studies indicating that humans readily apply social groups to agents, though tending to perceive agents, by default, as belonging to an out-group. Fourth, we look at how an individual's personality and traits shape their social interaction with agents. Finally, we discuss opportunities and challenges to the development of socially intelligent agents.

Theoretical Foundations

Following a series of experimental studies showing that people treated machines in a social manner (Nass & Moon, 2000; Nass et al., 1996, 1997, 1999, 2000), Nass and colleagues advanced a general theory for human-machine interaction – the media equation theory (Reeves & Nass, 1996). According to this theory, people will intuitively treat machines in social settings as if they were social actors. The idea is that humans carry social heuristics learnt in human-human interaction to human-machine interaction automatically. A strict interpretation of the theory further argues that any social effect we see among humans could carry to human-machine interaction: “Findings and experimental methods from the social sciences can be applied directly to human-media interaction. It is possible to take a psychology research paper about how people respond to other people, replace the word ‘human’ with the word ‘computer’ and get the same results” (Reeves & Nass 1996; pg. 28). Some of the studies supporting this view showed that people were polite to machines (Nass et al., 1999), formed positive impressions of machines perceived to be teammates (Nass et al., 1996), and applied social stereotypes to machines (Nass et al., 1997, 2000).

Blascovich et al. (2002), in contrast, proposed a more refined view – the social influence theory – which argues that machines are more likely to influence people, the higher the agency and realism of the machine. Agency increases with the perception that the machine is being controlled by a human; thus, an autonomous machine being controlled by algorithms would rank lower in this factor. Realism, or fidelity, relates to the photorealism of the machine (i.e., does it look like a human?),behavioral realism of the machine (i.e., does it behave like a human?), and the social realism of the machine (i.e., does it engage socially like a human?; Sinatra et al. 2021). According to this theory, it is possible to compensate for lack of agency by increasing realism. Studies in line with this view indicate that machines mirroring humans’ nonverbal behavior can increase rapport (Gratch et al., 2007), simulating emotion expression in machines can increase cooperation (de Melo et al., 2014), and photorealistic avatars that look like the user can increase compliance with an exercise regime (Fox & Bailenson, 2009).

Evidence from the emerging field of neuroeconomics presents further evidence that, even though humans can treat machines in a social manner, there are still important differences in the way humans behave with machines vs. humans. This research shows that people can reach different decisions and show different patterns of brain activation with machines in decision tasks, when compared to humans. Gallagher et al. (2002) showed that when people played the rock-paper-scissors game with a human there was activation of the medial prefrontal cortex, a region of the brain that had previously been implicated in mentalizing (i.e., inferring of other's beliefs, desires and intentions); however, no such activation occurred when people engaged with a machine that followed a known predefined algorithm. McCabe et al. (2001) found a similar pattern when people played the trust game with humans vs. machines, and others replicated this finding using prisoner's dilemma games (Kircher et al., 2009, Krach et al., 2008, Rilling et al., 2002). Sanfey et al. (2003) further showed that, when receiving unfair offers in the ultimatum game, people showed stronger activation of the bilateral anterior insula – a region associated with the experience of negative emotions – when engaging with humans, when compared to machines. de Melo et al. (2016) also showed that people made more favorable offers to humans than machines in various decision tasks and showed less guilt when exploiting machines. Overall, the evidence suggests that people experienced less emotion and spent less effort inferring mental states with machines than with humans. These findings are compatible with research showing that people perceive, by default, less mind in machines than in humans (Gray et al., 2007; Waytz et al., 2010). Denying mind to others or perceiving inferior mental ability in others, in turn, is known to lead to discrimination (Haslam, 2006) and can form the basis for out-group discrimination (Crisp & Hewstone, 2007; Tajfel & Turner, 1986), which could in some cases negatively impact human-agent team performance. In sum, even though there is increasing evidence that people are able to treat agents in a social manner, there is complementary evidence indicating that important differences remain. In the next sections, we describe mechanisms that support and encourage social interaction between humans and agents and, thus, help bring human-agent collaboration closer to what we see among humans.

Verbal Communication

One of the most fundamental ways in which humans communicate with one another is verbally via natural language. This makes natural language a promising modality for human-machine communication. Recent breakthroughs in cloud computing and machine learning techniques, combined with increased availability of large, well-annotated language corpora, have led to a dramatic uptick in research and development in the past decade, and in the widespread use of these technologies in everyday interactions. Natural Language Processing (NLP) technologies are more advanced – and more ubiquitous – than ever.

Perhaps the most familiar natural language enabled systems today are smart objects in the home (e.g., Alexa) and language-enabled assistants, such as Siri, on mobile devices. In our pockets, millions of us carry an NLP-enabled agent that we team with to do web searches, place orders, get directions, and even tell jokes. Siri, and similar technologies like Alexa, rely extensively on web-based data to answer users' questions and fulfill users' requests. However, such technologies are not well suited to addressing physical tasks or reasoning about items or processes that require sensing the physical world. Siri can look up directions to a sporting goods store where you can buy a basketball, but Siri can't find the basketball you already have stored in your closet.

Physically situated reasoning is a difficult challenge for autonomous and NLP based systems, but progress is being made by uniting techniques (Gratch et al., 2015) used to develop non-physically-situated conversational agents (e.g., Rizzo et al., 2011; DeVault et al., 2014; Traum et al., 2015) with robots and techniques to sense and navigate the real world (such as light-based ranging or image recognition technologies). One example is the Army Research Laboratory's JUDI (Joint Understanding and Dialogue Interface) project which involved developing a natural language-enabled search and navigation robot that can respond to spoken commands from a remotely-located human teammate (Marge et al., 2016, Marge et al., 2017, Bonial et al., 2017; Lukin et al., 2018). The system can take action in the physical world and reply with text-based natural language confirmations and requests for clarification.

Even a remotely-located robot with no visualization of facial features or intended emotional expressions can be perceived as a social partner simply by interacting with human users (Reeves & Nass, 1996). Humans interacting with versions of the JUDI system, for example, expressed interest in naming the robot, gave it encouraging verbal feedback, and inquired about the robot's gender (Henry et al., 2017; Pollard et al., 2018).

Creating a natural language dialogue system requires understanding what humans will want to say to the agent. Language-based systems can be constrained, involving a limited number of pre-determined commands and queries which the machine or agent is built to understand. However, a more flexible and intuitive system can be constructed to allow humans to speak to the agent using whatever phrasing they wish. Minimally-constrained systems are more complex to build and often require extensive data collection to uncover, and account for, the wide diversity of verbal communication that humans may want to use with the system. This must be done so that the system can be built to accommodate this diversity. Humans can provide sample language during interviews or on questionnaires, or such data can be acquired by harvesting existing data in the wild such as from online chat language. A more direct and naturalistic way to collect data for some use cases is the Wizard of Oz method. This method encourages humans to interact with what they believe to be an autonomous agent. The human participant interacts with the agent as they would wish to, while a human researcher behind the scenes listens to (or reads) the language used and generates appropriate responses to act out through the "agent." The human researcher behind the scenes thus performs tasks such as speech recognition (if spoken language is to be used), natural language understanding, and natural language generation in the place of what will eventually become an automated system. This "wizard" may also execute other functions to eventually be performed by the autonomy (such as movement). A body of ecologically valid, application-specific human language samples, along with linked appropriate agent responses, can thus be collected before the envisioned autonomous system exists and can be used to then create the actual system. This method has been used fruitfully in the development of human-agent natural language dialogue systems for a variety of applications, such as counselling (e.g., DeVault et al., 2014), museum and history reenactments (e.g., Traum et al., 2015), and for physically situated navigation for robots (Bonial et al., 2017; Lukin et al., 2018).

Building machine systems to understand natural human utterances is just one piece of the puzzle. On the other side of the interaction, machines or agents built to produce verbal communication must express themselves in a manner that humans can understand and in ways that engender an appropriate level of rapport or trust in the agent. Here, we discuss two key examples of language use patterns that affect user response and performance outcomes in human-agent interactions, focusing specifically on agents used explicitly for social purposes, such as teaching.

Language style can influence a human's perception of, and response to, an autonomous agent. Teaching agents that use conversational language as opposed to formal language are often perceived more positively and lead to better human performance from the interaction (e.g., learning gains). Conversational language uses personalized pronouns (e.g., "you" and "me"), colloquial phrasing, and/or slang terms, which should engender a stronger feeling of social presence, and potentially greater motivation or interest, when interacting with the agent (the *personalization principle,* Moreno & Mayer, 2004). Some studies have found the use of conversational style language by pedagogical virtual agents to result in greater learning or transfer gains across a variety of academic subjects (Moreno & Mayer 2000; Moreno & Mayer 2004; Rey & Steib, 2013; Reichelt et al., 2014; Schneider et al. 2015a; Schrader et al., 2018). In some studies, conversational language features were associated with greater motivation (Kartal 2010; Reichelt et al., 2014) and greater interest (Kartal 2010). However, personalization can lead to worse performance in some cases (e.g., Kühl & Zander, 2017) or higher cognitive load (e.g., Kurt 2011), and users' individual differences can play a role in how users respond to different language styles (Schrader et al., 2018).

While conversational language style can often be beneficial, research shows that it is also often beneficial for language to be polite. As with human-human verbal communication, agents that employ polite and face-saving (Brown & Levinson, 1987) language often yield improved human performance results. Polite, face-saving linguistic acts include the use of indirect wording or suggestions rather than direct wording or commands (e.g., *Could you turn the page?* or *For more information, you can turn the page*. vs. *Turn the page.*) The polite wordings help the receiver feel more as if they are exercising independent agency and personal control in completing the tasks. This can be particularly important when agents are providing feedback regarding human performance (Mikheeva et al., 2019). The use of a polite language style was found to facilitate performance gains in a variety of learning domains (Wang et al., 2008; Schneider et al., 2015b), led to users choosing to spend more time engaging more with the learning material (Mikheeva et al., 2019), and was found to be more natural and less stress-inducing in a conversational interview context (Gebhard et al., 2014). However, users' individual differences can influence the effects of polite agent language on outcomes (e.g., Wang et al., 2008; McLaren et al., 2011), and non-direct language can be problematic in some settings, such as with healthcare robots (Lee et al., 2017). Additional research is needed to understand the various factors influencing human responses to agent politeness and conversational styles. The importance of individual differences and social group signifiers (e.g., accented speech, Khooshabeh et al., 2017) are discussed in separate sections of this chapter.

Nonverbal Communication

Complementing research on verbal communication, there has been growing interest on the role of nonverbal signaling in facilitating collaboration (Boone & Buck, 2003; de Melo et al., 2014; Gratch & de Melo, 2019; Lerner et al., 2015; Tickle-Degnen & Rosenthal, 1990; van Kleef & Côté, 2018; van Kleef et al., 2010). Rapport is emblematic and refers to a social phenomenon that occurs when people are highly engaged with each other, focused, mutually attentive, and enjoying the interaction (Tickle-Degnen & Rosenthal, 1990). Two important components in establishing rapport are nonverbal responsiveness – e.g., listening behaviors – and mimicry – for instance, mimicking the counterpart's posture. Establishing rapport has been shown to facilitate negotiation (Drolet & Morris, 2000), therapy (Tsui & Schultz, 1985), teaching (Fuchs, 1987), and caregiving (Burns, 1984), among others. Accordingly, human-computer interaction researchers have attempted to establish rapport between agents and humans, in particular, through nonverbal behavior (Bailenson & Yee, 2005; Gratch et al., 2006, 2007). Bailenson and Yee (2005) showed that mimicking agents were more persuasive and were rated more positively. Gratch et al. (2006, 2007) also showed that an agent that displayed listening behaviors – e.g., a nod in response to prosodic cues in the counterpart's speech – led to more fluent conversation and received more positive ratings.

One category of nonverbal signals, however, has received considerable attention due to its influence on human decision making and role in promoting cooperation: emotion expressions. In the last twenty years, there has been substantial experimental support for the interpersonal influence of emotion expressions in social decision making (for reviews see: Lerner et al., 2015; van Kleef & Côté, 2018; van Kleef et al., 2010), including effects on concession-making (van Kleef et al., 2004, 2006), emergence of cooperation (de Melo, Carnevale, Read, & Gratch, 2014), fairness (Terada & Takeuchi, 2017; van Dijk et al., 2008), trust building (Krumhuber et al., 2007), and everyday life (Parkinson & Simons, 2009). Progress has also been made in understanding the pathways by which these effects operate. Broadly speaking, emotions can serve to evoke emotions in others via contagion (Lanzetta & Englis, 1989; Niedenthal et al., 2010) or can serve as information, revealing the experiencer's mental state (de Melo et al., 2014; Manstead & Fischer, 2001; van Kleef et al., 2010). This latter path is particularly interesting as it suggests a mechanism whereby people are able to "read other people's minds" by making appropriate inferences from other's emotion displays (de Melo et al., 2014; Gratch & de Melo, 2019).

There is general agreement among emotion theorists that emotions are elicited by ongoing, conscious or nonconscious, appraisal of events with respect to the individual's beliefs and goals (Frijda, 1986; Scherer, 2001; Scherer & Moors, 2019). Different emotions result from different appraisals, as well as their associated patterns of physiological manifestation, action tendencies, and behavioral expressions. Expressions of emotions, therefore, reflect differentiated information about the expresser's appraisals and goals. Accordingly, researchers have noted that emotions serve important social functions, including communicating one's beliefs, desires, and intentions to others (Frijda & Mesquita, 1994; Keltner & Haidt, 1999; Keltner & Lerner, 2010; Morris & Keltner, 2000). In line with this view, Frank (1988, 2004) notes that emotion signals are ideal for identifying cooperators in society, especially since they tend to be harder to fake.

Several studies have now shown that emotion expressions can shape cooperation. de Melo and colleagues (de Melo et al., 2014; de Melo & Terada, 2019, 2020) revealed that emotion expressions compatible with an intention to cooperate (e.g., joy following mutual cooperation and regret following exploitation) led to increased cooperation in the iterated prisoner's dilemma. In contrast, emotion displays compatible with a competitive intention (e.g., joy following exploitation) hindered cooperation. These results emphasize the contextual nature of the effects of emotion, with the same exact expression leading to opposite effects according to the context in which it was shown. de Melo and Terada (2020) further showed that the effect of emotion expressions combine in interesting ways with the effect of actions: when actions were ambiguous or insufficient to convey the individual's intentions, the emotion signal became more relevant in the interaction; in contrast, when the counterpart's actions were clearly indicative of an intention to compete, emotion expressions had no effect. van Kleef and colleagues (van Kleef, 2016; van Kleef & Côté, 2018; van Kleef et al., 2010) further articulated the impact of emotion signals in more complex settings, such as negotiation. For instance, anger in negotiation led to increased concessions, as receivers inferred high aspirations on the sender's side (van Kleef et al., 2004). They further note several moderating factors – such as power and motivation to process information – on the effects of emotion expression on decision making (van Kleef, 2016).

Emotion expressions have also been shown to enhance human-agent interaction (Beale & Creed, 2009). In the context of the prisoner's dilemma, de Melo et al. (2009) showed that simulating facial expressions of emotion in an agent increased cooperation, when compared to an agent that showed no emotion. In follow-up work, de Melo et al. (2012) demonstrated that emotion expressions in agents also had the ability to hinder cooperation with humans, if they reflected competitive intentions (e.g., smile following exploitation). Moreover, several studies used experimental stimuli consisting of virtual faces – i.e., algorithms that simulate prototypical human facial expressions (de Melo, Carnevale, Gratch, 2014) – to research the social effects of emotions and, thus, arguably already provide support to the plausibility of simulating emotion in agents to promote collaboration with humans. Finally, researchers noted that emotion expressions can be used to overcome negative biases people have with agents (more on this in the next section) (de Melo & Terada, 2019; Terada & Takeuchi, 2017).

Social Groups

Humans often categorize others as belonging to distinct social groups, distinguishing "us" versus "them", and this categorization influences collaboration as people are more likely to trust and cooperate with in-group than out-group members (Baillet, Wu, De Dreu, 2014; Brewer, 1979; Crisp & Hewstone, 2007; Tajfel & Turner, 1986). Social identities, however, are complex and multifaceted. In many situations, more than one social category (e.g., gender, age, ethnicity) may be relevant. On the one hand, context can prime one category to become more dominant (or salient) and effectively exclude the influence of others. On the other hand, social categories can be simultaneously salient and have an additive effect on people's behavior (Crisp & Hewstone, 2007). These mechanisms based on multiple categories have, in fact, been proposed as the basis for reducing intergroup bias.

Experimental research indicates that people also engage in social categorization with agents. Nass et al. (1997) showed that participants perceived computers according to gender stereotypes, assigning more competence to computers with a female voice than a male voice on the topic of "love and relationships." Khooshabeh et al. (2017) showed that agents with voices with and without accent of the same culture, impacted perceptions of the appropriateness of the machine's decisions in social dilemmas. Researchers further showed that participants were more likely to cooperate (de Melo & Terada, 2019; de Melo, Carnevale, Gratch, 2014) and trust (Nass et al., 2000) agents that had virtual faces matching the participant's ethnicity.

However, as noted in the Theoretical Foundations section, people tend to make more favorable decisions with humans than agents, thus appearing to treat agents as out-group members by default. Accordingly, researchers looked at the possibility of associating positive social categories to compensate for negative social categories associated with agents. de Melo and Terada (2019) showed that conveying a cue for shared cultural identity in agents –

through the ethnicity of the agent's virtual face – was sufficient to mitigate this bias. de Melo, Carnevale, and Gratch (2014) further showed that creating a sense of belonging to the same team *and* sharing the same ethnicity could lead participants to be even more generous with agents than some humans. Complementary, de Melo and Terada (2019) showed that emotional expressions communicating affiliative intent (e.g., joy following mutual cooperation) could override initial expectations participants formed from social categorization.

Individual Traits

Individual differences influence social interactions. For example, a cross-cultural study of personality measured with the five-factor model found that Agreeableness and Conscientiousness were associated with quality of social interactions (Nezlek et al., 2011). Social Value Orientation is a relatively stable individual difference measure that describes the extent to which a person considers their own interests and the interests of others in their social interactions (Van Lange et al., 1997). Differences in social value orientation are associated with differences in negotiation behaviors (de Dreu & Van Lange, 1995). More rapidly changing individual differences, such as affective state, are associated with differences in social judgment (Forgas & Moylan, 1987). Although these examples come from studies of human social interaction, both the media equation theory (Reeves & Nass, 1996) and social influence theory (Blascovich et al., 2002), discussed earlier, suggest that individual differences that affect social interaction with humans should also affect social interactions with sufficiently realistic non-human social agents.

However, the extent to which findings from social psychology generalize to interactions with social agents might depend on the extent to which people treat agents as they would treat another person. The tendency to treat non-human animals and objects as human is called anthropomorphism (Epley et al., 2007), and individuals differ in the extent they tend to anthropomorphize. Differences in anthropomorphic tendency are associated with differences in the extent to which people interact with non-human agents as they would with a human (Waytz et al., 2010). Anthropomorphic tendency appears to be somewhat stable, but situational factors also affect anthropomorphic tendency. People who are lonelier anthropomorphize more, and being reminded of close, supportive social relationships is associated with less anthropomorphizing behaviors (Bartz et al., 2016). Similarly, Shin & Kim (2020) replicated the relationship between loneliness and increases in anthropomorphizing behavior, and they found that inducing feelings of loneliness with a writing task led to more anthropomorphizing behavior compared to control. More research is needed to understand how individual differences in anthropomorphizing mediate relationships between other characteristics and behavior toward social agents.

Studies of individual differences in social agent interaction can have practical utility, because social agents can potentially adapt to individual users' characteristics to increase the probability of some outcome the agent's designer prefers. Pedagogical agents (Sinatra et al., 2021) have been designed to leverage knowledge about an individual student to optimize that student's learning. As examples, PAL3 (Swartout et al., 2016) and GIFT (Sottilare et al., 2012) are systems that maintain a model of the student's learning based on the student's record of past activities, and they both can customize recommended activities based on a model of learning and forgetting to account for individual differences in background knowledge and learning speed. Both systems support delivering these recommendations via an onscreen agent. Other pedagogical agents have been designed to account for the learner's cultural knowledge in delivering feedback and content in cultural interaction training (Lane & Wray, 2012). Although some work has been done to examine the utility of stable personality traits in learner modeling, inferring the learner's characteristics from behavior might be a more promising approach (Abyaa et al., 2019).

Progress has been made in inferring human characteristics from behavior in the context of human/agent negotiation. For example, Sequeira & Marsella (2018) analyzed the offers humans made in the context of a structured human/agent negotiation task with the goal of discovering for each individual person a negotiating algorithm that reproduced the history of offers that person made. The characteristics of the best algorithms could be interpreted to summarize that person's approach to the negotiation. The human negotiators' social value orientation and Machiavellianism traits were also measured, and people with similar traits also had similar inferred algorithms. This illustrates an approach that could lead to negotiating social agents that learn a human partner's true negotiating style and objectives, potentially enabling the agent to engage in more effective negotiations.

In non-oppositional contexts, a social agent could request information about the human to enable it to more effectively interact. The field of social psychology has developed many well-validated questionnaires and other instruments from which reliable individual differences measures can be calculated. For example, there are stable, population-level effects of message framing on decision outcomes (Tversky & Kahneman, 1981). However, an individual difference measure called regulatory focus (Higgins, 1998) accounts for variability in subjective (Higgins et al., 2003) and objective (Files et al., 2019; Glass et al., 2011) effects of message framing in the context of risk communication and performance feedback, among others. These findings suggest that social agents could more effectively communicate with individuals if they leverage knowledge of the individual's regulatory focus to appropriately frame messages about risk and opportunity.

In summary, individual differences affect several aspects of human social interaction, and the extent to which those same differences affect interactions with social agents might itself depend on individual differences in anthropomorphic tendency. Despite these complications, using individual differences to personalize social agent interactions could have big payoffs by creating a more effective experience. Progress has been made in identifying promising user characteristics on which to base agent customization, as has progress been made in inferring and measuring those characteristics. More research is needed to understand the complex interactions between user characteristics, social agent personalization, and interaction contexts to derive generalizable principals for effective personalized social agent interactions.

General Discussion

We have argued that endowing agents with social skills can promote collaboration with humans. We reviewed literature indicating that humans regularly use these skills when working with others to achieve common goals. Moreover, we reviewed theory and experimental findings suggesting that humans will readily engage in a social manner with agents, especially the higher the social skill displayed by agents. The argument, therefore, suggests that designers cannot afford to ignore the broader social context in human-agent teams. Just as in human-human interaction, building rapport, trust, and cooperation with humans requires agents to skillfully navigate this social context, above and beyond the (non-social) functional aspects of the task.

We, thus, offer the following guidelines for designing socially intelligent agents:

- **Engage verbally and nonverbally with human teammates**: Communication between human and agent teammates can help ground the interaction, repair misunderstandings, and acknowledge mutual understanding through feedback cues. Nonverbal communication, especially through emotion expressions (de Melo et al., 2014), can be particularly important in communicating to humans an intention to cooperate, build trust, and mitigate mistakes (e.g., displays of regret).

- **Enhance perceptions of shared (social) group identity in agents**: Humans are likely to treat social agents as if they were out-group members, but this can be overcome by emphasizing shared social group membership (e.g., same ethnicity or accent as human teammates; Khooshabeh et al., 2017) or overriding negative social groups through clear signals of affiliative intent (e.g., appropriate expressions of emotion; de Melo & Terada, 2019).
- **Adapt the agent's behavior to the individual's personality and traits**: Agents that are able to customize their behavior to the individual's specific traits are more likely to collaborate successfully. Agents, therefore, should seek to learn about their human teammates through pre-interaction subjective personality scales, continuous physiological monitoring, and inferences from their actions. Once a model of the teammate is available, the agent can choose the optimal strategy – e.g., if a teammate is perceived to have a cooperative social value orientation, the agent can engage in cooperation from the start; however, if the teammate is perceived to have a competitive orientation, the agent can engage in tit-for-tat behavior to encourage cooperative behavior.

We, nevertheless, identify a few important open challenges to the successful adoption of socially intelligent agents:

- **Manage human teammates' expectations**: There is a longstanding debate in the robotics community about the so-called "uncanny valley" (Moris, 1970), which reflects a sudden shift in attitude, from empathy to revulsion, when the appearance of the robot begins to appear human-like while simultaneously failing to sufficiently act and look like a human. This uneasiness arises from the mismatch between the robot's appearance and the expectations that it generates in humans engaging with it.

Similarly, with social agents it is critical that designers adjust the way social skill is expressed to the sophistication of the underlying model. For instance, it may be preferable to have a more robotic voice, rather than create the expectation that an agent is able to engage in open-ended conversation. This issue is particularly important as research suggests that people are more reluctant to trust autonomous machines given the lack of experience and understanding about how it works (Gillis, 2017). It is, therefore, essential to continuously manage the expectations in human teammates and help them form the correct mental model about the agents' (social) capabilities.

- **Ethical considerations**: The ability to exert social influence in humans is powerful, but potentially dangerous. Recently, there's been much discussion on the ethics of using technology that is able to perceive people's emotional states (Greene, 2019). There are concerns, on the one hand, about the accuracy of these algorithms and, on the other hand, about the appropriateness of engaging in emotion perception without the user's consent. Similarly, social agents will be able to make inferences about people's mental and affective states (e.g., is the individual trying to cooperate? Is the individual angry?) and engage in behavior to change those states (e.g., express regret to reduce anger and promote cooperation). However, when is it appropriate to engage in this type of social manipulation, especially since often its effect is subconscious? Moreover, unlike humans, it is trivial for agents to fake social signals and, thus, steer humans' behavior towards its own interests. These are not easy issues, and it is beyond the scope of the paper to solve them; however, it is important to acknowledge them, engage in cross-disciplinary debate, and encourage the ethical development of social agent technology from the start.

Socially intelligent agents promise to enhance the efficacy and efficiency of hybrid teams. They add a missing dimension to most current agent technology and bring human-agent collaboration closer to the kind of collaboration we see among humans. Moreover, because these agents can be designed from the ground up to optimally simulate social skill, if used ethically, they introduce a unique opportunity to create a more collaborative society.

References

Abyaa, A., Khalidi Idrissi, M., & Bennani, S. (2019). Learner modelling: Systematic review of the literature from the last 5 years. *Educational Technology Research and Development*, *67*, 1105-1143.

Bailenson, J., & Yee, N. (2005). Digital Chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science*, *16*, 814-819.

Baillet, D., Wu, J., De Dreu, C. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin*, *140*, 1556-1581.

Bartz, J. A., Tchalova, K., & Fenerci, C. (2016). Reminders of social connection can attenuate anthropomorphism: A replication and extension of Epley, Akalis, Waytz, and Cacioppo (2008). *Psychological Science*, *27*, 1644-1650.

Bates, J. (1994). The role of emotion in believable agents. *Communications of the ACM*, *37*.

Beale, R., & Creed, C. (2009). Affective interaction: How emotional agents affect users. *Human-Computer Studies*, 67, 755-776.

Blascovich, J., Loomis, J., Beall, A., Swinth, K., Hoyt, C., & Bailenson, J. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*, 103-124.

Bonial, C., Marge, M., Artstein, R., Foots, A., Gervits, F., Hayes, C. J., Henry, C., Hill, S. G., Leuski, A., Lukin, S. M., Moolchandani, P., Pollard, K. A., Traum, D., & Voss, C. R. (2017). Laying Down the Yellow Brick Road: Development of a Wizard-of-Oz Interface for Collecting Human-Robot Dialogue. *Proceedings of the Association for the Advancement of Artificial Intelligence Fall Symposium (AAAI 2017)*.

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*, 1573-1576.

Boone, R., & Buck, R. (2003). Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior*, *27*, 163-182.

Brewer, M. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, *86*, 307-324.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

Burns, M. (1984). Rapport and relationships: The basis of child care. *Journal of Child Care*, *2*, p. 47-57.

Crisp, R., & Hewstone, M. (2007). Multiple social categorization. *Advances in Experimental Social Psychology*, *39*,163-254.

de Dreu, C., & Van Lange, P. (1995). The impact of social value orientations on negotiator cognition and behavior. *Personality and Social Psychology Bulletin*, *21*, 1178-1188.

de Melo, C., & Terada, K. (2019). Cooperation with autonomous machines through culture and emotion. PLOS ONE, https://doi.org/10.1371/journal.pone.0224758.

de Melo, C., & Terada, K. (2020). The interplay of emotion expressions and strategy in promoting cooperation in the iterated prisoner's dilemma. *Scientific Reports*, *10*, 14959.

de Melo, C., Marsella, S., & Gratch, J. (2016). People do not feel guilty about exploiting machines. *ACM Transactions of Computer-Human Interaction*, *23*.

de Melo, C., Marsella, S., & Gratch, J. (2019). Human cooperation when acting through autonomous machines. *Proceedings of the National Academy of Sciences U.S.A.*, *116*, 3482-3487.

de Melo, C., Carnevale, P., & Gratch, J. (2012). The impact of emotion displays in embodied agents on emergence of cooperation with people. *Presence: Teleoperators and Virtual Environments Journal*, *20*, 449-465.

de Melo, C., Carnevale, P., & Gratch, J. (2014). Using virtual confederates to research intergroup bias and conflict. In *Proceedings of Academy of Management (AoM)*.

de Melo, C., Zheng, L., & Gratch, J. (2009). Expression of moral emotions in cooperating agents. In Proceedings of the International Conference on Intelligent Virtual Agents.

de Melo, C., Carnevale, P., Read, S., & Gratch, J. (2014). Reading people's minds from emotion expressions in interdependent decision making. *Journal of Personality and Social Psychology*, *106*, 73-88.

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, Gale Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Sherer, Giota Stratou, Apar Suri, David Traum, Rachel Wood, … Louis-Philippe Morency. (2014). SimSensei Kiosk: A virtual human interviewer for healthcare decision support. *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*.

Drolet, A., & Morris, M. (2000). Rapport in conflict resolution: accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts. *Experimental Social Psychology*, *36*, 26-50.

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

Files, B., Pollard, K., Oiknine, A., Passaro, A., & Khooshabeh, P. (2019). Prevention focus relates to performance on a loss-framed inhibitory control task. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00726

Forgas, J., & Moylan, S. (1987). After the movies: Transient mood and social judgments. *Personality and Social Psychology Bulletin*, *13*, 467-477.

Fox, J., & Bailenson, J. (2009). Virtual self-modeling: The effects of vicarious reinforcement and identification on exercise behaviors. *Media Psychology*, *12*, 1-25.

Frank, R. (1988). *Passions within reason: The strategic role of the emotions*. New York: Norton.

Frank, R. (2004). Introducing moral emotions into models of rational choice. In A. Manstead, N. Frijda & A. Fischer (Eds.), *Feelings and emotions* (pp. 422-440). New York, NY: Cambridge University Press.

Frijda, N. (1986). *The emotions*. Cambridge, UK: Cambridge University Press.

Frijda, N., & Mesquita, B. (1994). The social roles and functions of emotions. In S. Kitayama & H. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 51-87). Washington, DC: American Psychological Association.

Fuchs, D. (1987). Examiner familiarity effects on test performance: implications for training and practice. *Topics in Early Childhood Special Education*, *7*, 90-104.

Gallagher, H., Anthony, J., Roepstorff, A., & Frith, C. (2002). Imaging the intentional stance in a competitive game. *NeuroImage*, *16*, 814-821.

Gebhard, P., Baur, T., Damian, I., Mehlmann, G., Wagner, J., & André, E. (2014). Exploring Interaction Strategies for Virtual Characters to Induce Stress in Simulated Job Interviews. *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*, 661-668.

Gillis, J. (2017). Warfighter trust in autonomy. *DSIAC*, *4*, 23-29.

Glass, B., Maddox, W., & Markman, A. (2011). Regulatory fit effects on stimulus identification. *Attention, Perception, & Psychophysics*, *73*, 927-937.

Gratch, J., & de Melo, C. (2019). Inferring intentions from emotion expressions in social decision making. In U. Hess, & S. Hareli (Eds.), *The Social Nature of Emotion Expression* (pp. 141-160). Springer.

Gratch, J., Hill, S., Morency, L.-P., Pynadath, D., & Traum, D. (2015). Exploring the Implications of Virtual Human Research for Human-Robot Teams. In R. Shumaker & S. Lackey (Eds.), *Virtual, Augmented and Mixed Reality* (Vol. 9179, pp. 186-196). Springer.

Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007). Creating rapport with virtual agents. In *Proceedings of Intelligent Virtual Agents (IVA'07).* Springer Berlin Heidelberg, 125-138.

Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007). Creating rapport with virtual agents. In Proceedings of 7th International Conference on Intelligent Virtual Agents.

Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., Werf, R., & Morency, L.-P. (2006). Virtual rapport. In Proceedings of 6th International Conference on Intelligent Virtual Agents.

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

Greene, G. (2019). The ethics of AI and emotional intelligence: Data sources, applications, and questions for evaluating ethics risk. *Partnership on AI Report*. Accessed from: https://www.partnershiponai.org/the-ethics-of-ai-and-emotional-intelligence/

Hancock, P., Billings, D., Schaefer, K., Chen, J., de Visser, E., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, *53*, 517-527.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, *10*, 252-264.

Henry, C., Moolchandani, P., Pollard, K. A., Bonial, C., Foots, A., Artstein, R., Hayes, C., Voss, C. R., Traum, D., & Marge, M. (2017). Towards efficient human-robot dialogue collection: Moving Fido into the virtual world. *Proceedings of ACL Workshop Women and Underrepresented Minorities in Natural Language Processing*.

Higgins, E. (1998). Promotion and prevention: Regulatory focus as a motivational principle. *Advances in Experimental Social Psychology*, *30*, 1-46.

Higgins, E., Idson, L., Freitas, A., Spiegel, S., & Molden, D. (2003). Transfer of value from fit. *Journal of Personality and Social Psychology*, *84*, 1140.

Kartal, G. (2010). Does language matter in multimedia learning? Personalization principle revisited. *Journal of Educational Psychology*, *102*, 615-624.

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion*, *13*, 505-521.

Keltner, D. & Lerner, J. (2010). Emotion. In D. Gilbert, S. Fiske & G. Lindzey (Eds.), *The Handbook of Social Psychology* (pp. 312-347). John Wiley & Sons.

Khooshabeh, P., Dehghani, M., Nazarian, A., & Gratch, J. (2017). The cultural influence model: When accented natural language spoken by virtual characters matters. *AI & Society*, *32*, 9-16.

Kircher, T., Blümel, I., Marjoram. D., Lataster, T., Krabbendam, L., Weber, J., van Os, & Krach, S. (2009). Online mentalising investigated with functional MRI. *Neuroscience Letters*, *454*, 176-181.

Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology*, *24*, 183-214.

Kott, A., & Alberts, D. (2017). How do you command an Army of intelligent things? *Computer*, *50*, 96-100.

Kott, A., & Stump, E. (2019). Intelligent autonomous things on the battlefield. In W. Lawlessm, R. Mittu, D. Sofge, I. Moskowitz, & S. Russell (Eds.), *Artificial Intelligence for the Internet of Everything* (pp. 47-65). Academic Press.

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLOS ONE*, *3*, 1-11.

Krumhuber, E., Manstead, A., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, *7*, 730-735.

Kühl, T., & Zander, S. (2017). An inverted personalization effect when learning with multimedia: The case of aversive content. *Computers & Education*, *108*, 71-84.

Kurt, A. (2011). Personalization principle in multimedia learning: Conversational versus formal style in written word. *TOJET: The Turkish Online Journal of Educational Tecnology*, *10*, 8.

Lane, H., & Wray, R. (2012). Individualized cultural and social skills learning with virtual humans. In A. M. Lesgold & P. J. Durlach (Eds.), *Adaptive Technologies for Training and Education* (pp. 204-221). Cambridge University Press.

Lanzetta, J., & Englis, B. (1989). Expectations of cooperation and competition and their effects on observer's vicarious emotional responses. *Journal of Personality and Social Psychology*, *36*, 543-554.

Lee, D., & See, K. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*, 50-80.

Lee, N., Kim, J., Kim, E., & Kwon, O. (2017). The Influence of Politeness Behavior on User Compliance with Social Robots in a Healthcare Service Setting. *International Journal of Social Robotics*, *9*, 727-743.

Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: A survey. *International Journal of Social Robotics*, *5*, 291-308.

Lerner, J., Li, Y., Valdesolo, P., & Kassam, K. (2015). Emotion and decision making. *Annual Review of Psychology*, *66*, 799-823.

Lukin, S. M., Gervits, F., Hayes, C. J., Moolchandani, P., Leuski, A., Rogers III, J. G., Sanchez Amaro, C., Marge, M., Voss, C. R., & Traum, D. (2018). ScoutBot: A Dialogue System for Collaborative Navigation. *Proceedings of ACL 2018, System Demonstrations*, 93-98.

Manstead, A., & Fischer, A. (2001). Social appraisal: The social world as object of and influence on appraisal processes. In K. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 221-232). Oxford, UK: Oxford University Press.

Marge, M., Bonial, C., Foots, A., Hayes, C., Henry, C., Pollard, K. A., Artstein, R., Voss, C. R., & Traum, D. (2017). Exploring Variation of Natural Human Commands to a Robot in a Collaborative Navigation Task. *Proceedings of the First Workshop on Language Grounding for Robotics*, 58-66.

Marge, M., Bonial, C., Pollard, K. A., Artstein, R., Byrne, B., Hill, S. G., Voss, C., & Traum, D. (2016). Assessing Agreement in Human-Robot Dialogue Strategies: A Tale of Two Wizards. In D. Traum, W. Swartout, P. Khooshabeh, S. Kopp, S. Scherer, & A. Leuski (Eds.), *Intelligent Virtual Agents* (pp. 484-488). Springer.

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Science*, *98*, 11832-11835.

McLaren, B. M., DeLeeuw, K. E., & Mayer, R. E. (2011). A politeness effect in learning with web-based intelligent tutors. *International Journal of Human-Computer Studies*, *69*, 70-79.

Mikheeva, M., Schneider, S., Beege, M., & Rey, G. D. (2019). Boundary conditions of the politeness effect in online mathematical learning. *Computers in Human Behavior*, *92*, 419–427.

Moreno, R., & Mayer, R. E. (2000). Engaging Students in Active Learning: The Case for Personalized Multimedia Messages. *Journal of Educational Psychology*, *92*, 724-733.

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology*, *96*, 165-173.

Mori, M. (1970). The uncanny valley. *Energy*, *7*, 33-35.

Morris, M., & Keltner, D. (2000). How emotions work: An analysis of the social functions of emotional expression in negotiations. *Research in Organizational Behavior*, *22*, 1-50.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*, 81-103.

Nass, C., Fogg, B., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, *45*, 669-678.

Nass, C., Isbister, K., & Lee, E.-J. (2000). Truth is beauty: Researching embodied conversational agents. In J. Cassell (Ed.), *Embodied conversational agents* (pp. 374-402). Cambridge, MA: MIT Press.

Nass, C., Moon, Y., & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems. *Journal of Applied Psychology*, *29*, 1093-1110.

Nass, C., Moon, Y., & Green, N. (1997). Are computers gender-neutral? Gender stereotypic responses to computers. *Journal of Applied Social Psychology*, *27*, 864-876.

Nezlek, J. B., Schütz, A., Schröder-Abé, M., & Smith, C. V. (2011). A cross-cultural study of relationships between daily social interaction and the five-factor model of personality. *Journal of Personality*, *79*, 811–840.

Niedenthal, P., Mermillod, M., Maringer, M., & Hess, U. (2010). The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences*, *33*, 417-480.

Orbell, J., van de Kragt, A., & Dawes, R. (1988). Explaining discussion-induced cooperation. *Journal of Personality and Social Psychology*, *54*, 811-819.

Parkinson, B., & Simons, G. (2009). Affecting others: Social appraisal and emotion contagion in everyday decision making. *Personality and Social Psychology Bulletin, 35*, 1071-1084.

Pollard, K. A., Lukin, S. M., Marge, M., Foots, A., & Hill, S. G. (2018). How We Talk with Robots: Eliciting Minimally-Constrained Speech to Build Natural Language Interfaces and Capabilities. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*, 160-164.

Premack, D., & Premack, A. (1995). Origins of human social competence. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 205-218). New York: Bradford.

Rand, D., & Nowak, M. (2013). Human cooperation. *Trends in Cognitive Science*, *17*, 413-425.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Reichelt, M., Kämmerer, F., Niegemann, H. M., & Zander, S. (2014). Talk to me personally: Personalization of language style in computer-based learning. *Computers in Human Behavior*, *35*, 199–210. https://doi.org/10.1016/j.chb.2014.03.005

Rey, G. D., & Steib, N. (2013). The personalization effect in multimedia learning: The influence of dialect. *Computers in Human Behavior*, *29*, 2022-2028.

Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., & Kilts, C. (2002). A neural basis for social cooperation. *Neuron*, *35*, 395-405.

Rizzo, A., Sagae, K., Forbell, E., Kim, J., Lange, B., Buckwalter, J. G., Williams, J., Parsons, T. D., Kenny, P., Traum, D., Difede, J., & Rothbaum, B. O. (2011). SimCoach: An intelligent virtual human system for providing healthcare information and support. *Proc. Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2011*.

Sanfey, A., Rilling, J., Aronson, J., Nystrom, L., & Cohen, J. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*, 1755-1758.

Scherer, K. (2001). Appraisal considered as a process of multi-level sequential checking. In K. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92-120). New York, NY: Oxford University Press.

Scherer, K., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology*, *70*.

Schneider, S., Nebel, S., Pradel, S., & Rey, G. (2015a). Introducing the familiarity mechanism: A unified explanatory approach for the personalization effect and the examination of youth slang in multimedia learning. *Computers in Human Behavior*, *43*, 129–138.

Schneider, S., Nebel, S., Pradel, S., & Rey, G. D. (2015b). Mind your Ps and Qs! How polite instructions affect learning with multimedia. *Computers in Human Behavior*, *51*, 546-555.

Schrader, C., Reichelt, M., & Zander, S. (2018). The effect of the personalization principle on multimedia learning: The role of student individual interests as a predictor. *Educational Technology Research and Development*, *66*, 1387-1397.

Sequeira, P., & Marsella, S. (2018). Analyzing human negotiation using automated cognitive behavior analysis: The effect of personality. *CogSci*.

Shin, H., & Kim, J. (2020). My computer is more thoughtful than you: Loneliness, anthropomorphism and dehumanization. *Current Psychology*, *39*, 445-453.

Sinatra, A., Pollard, K., Files, B., Oiknine, A., Ericson, M., & Khooshabeh, P. (2021). Social fidelity in virtual agents: Impacts on presence and learning. *Computers in Human Behavior*, *114*, 106562. https://doi.org/10.1016/j.chb.2020.106562

Sottilare, R. A., Brawner, K. W., Goldberg, B. S., & Holden, H. K. (2012). The generalized intelligent framework for tutoring (GIFT). *Orlando, FL: US Army Research Laboratory–Human Research & Engineering Directorate (ARL-HRED).*

Stone, R., & Lavine, M. (2014). The social life of robots. *Science*, *346*, 178-179.

Swartout, W., Nye, B., Hartholt, A., Reilly, A., Graesser, A. C., VanLehn, K., Wetzel, J., Liewer, M., Morbini, F., Morgan, B., Wang, L., Benn, G., & Rosenberg, M. (2016). Designing a personal assistant for life-long learning (PAL3). *Proceedings of the 29th International Florida Artificial Intelligence Research Society Conference, FLAIRS 2016*, 491–496. Tajfel, H., & Turner, J. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. Austin (Eds.), *Psychology of intergroup relations* (pp 7-24). Chicago: Nelson-Hall.

Tajfel, H., & Turner, J. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. Austin (Eds.), *Psychology of intergroup relations* (pp 7-24). Chicago: Nelson-Hall.

Terada, K., & Takeuchi, C. (2017). Emotional expression in simple line drawings of a robot's face leads to higher offers in the ultimatum game. *Frontiers of Psychology*, *8*, Article 724.

Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological Inquiry*, *1*, 285-293.

Traum, D., Jones, A., Hays, K., Maio, H., Alexander, O., Artstein, R., Debevec, P., Gainer, A., Georgila, K., Haase, K., Jungblut, K., Leuski, A., Smith, S., & Swartout, W. (2015). New Dimensions in Testimony: Digitally Preserving a Holocaust Survivor's Interactive Storytelling. In H. Schoenau-Fog, L. E. Bruni, S. Louchart, & S. Baceviciute (Eds.), *Interactive Storytelling* (Vol. 9445, pp. 269-281). Springer.

Tsui, P., & Schultz, G. (1985). Failure of rapport: Why psychotheraputic engagement fails in the treatment of Asian clients. *American Journal of Orthopsychiatry*, *55*, 561-569.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453-458.

van Dijk, E., van Kleef, G. A., Steinel, W., & van Beest, I. (2008). A social functional approach to emotions in bargaining: When communicating anger pays and when it backfires. *Journal of Personality and Social Psychology*, *94*, 600–614.

van Kleef, G. (2016). *The interpersonal dynamics of emotion: Toward an integrative theory of Emotions as Social Information*. Cambridge, UK: Cambridge Univ. Press.

van Kleef, G., De Dreu, C., & Manstead, A. (2004). The interpersonal effects of anger and happiness in negotiations. *Journal of Personality and Social Psychology*, *86*, 57-76.

van Kleef, G., De Dreu, C., & Manstead, A. (2006). Supplication and appeasement in negotiation: The interpersonal effects of disappointment, worry, guilt, and regret. *Journal of Personality and Social Psychology*, *91*, 124-142.

van Kleef, G. & Côté, S. (2018). Emotional dynamics in conflict and negotiation: Individual, dyadic, and group processes. *Annual Review of Organizational Psychology and Organizational Behavior*, *5*, 437-464.

van Kleef, G., De Dreu, C., & Manstead, A. (2010). An interpersonal approach to emotion in social decision making: The emotions as social information model. *Advances in Experimental Social Psychology*, *42*, 45-96.

Van Lange, P., De Bruin, E., Otten, W., & Joireman, J. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology*, *73*, 733–746.

Waldrop, M. (2015). No drivers required. *Nature*, *518*, 20-23.

Wang, N., Johnson, W., Mayer, R., Rizzo, P., Shaw, E., & Collins, H. (2008). The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies*, *66*, 98-112.

Waytz, A., Gray, K., Epley, N., & Wegner, D. (2010). Causes and consequences of mind perception. *Trends in Cognitive Science*, *14*, 383-388.