Ann. N.Y. Acad. Sci. ISSN 0077-89

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Special Issue: *The Year in Cognitive Neuroscience*
REVIEW

From automata to animate beings: the scope and limits of attributing socialness to artificial agents

Ruud Hortensius[1,2,a] and Emily S. Cross[1,2,a]

[1]Wales Institute for Cognitive Neuroscience, School of Psychology, Bangor University, Wales, United Kingdom. [2]Institute of Neuroscience and Psychology, School of Psychology, University of Glasgow, Scotland, United Kingdom

Addresses for correspondence: Ruud Hortensius, Institute of Neuroscience and Psychology, School of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, Scotland, UK. ruud.hortensius@glasgow.ac.uk; Emily S. Cross, Institute of Neuroscience and Psychology, School of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, Scotland, UK. emily.cross@glasgow.ac.uk

Understanding the mechanisms and consequences of attributing socialness to artificial agents has important implications for how we can use technology to lead more productive and fulfilling lives. Here, we integrate recent findings on the factors that shape behavioral and brain mechanisms that support social interactions between humans and artificial agents. We review how visual features of an agent, as well as knowledge factors within the human observer, shape attributions across dimensions of socialness. We explore how anthropomorphism and dehumanization further influence how we perceive and interact with artificial agents. Based on these findings, we argue that the cognitive reconstruction within the human observer is likely to be far more crucial in shaping our interactions with artificial agents than previously thought, while the artificial agent's visual features are possibly of lesser importance. We combine these findings to provide an integrative theoretical account based on the "like me" hypothesis, and discuss the key role played by the Theory-of-Mind network, especially the temporal parietal junction, in the shift from mechanistic to social attributions. We conclude by highlighting outstanding questions on the impact of long-term interactions with artificial agents on the behavioral and brain mechanisms of attributing socialness to these agents.

Keywords: socialness attribution; animacy; anthropomorphism; artificial agents; human–robot interaction; social cognition

Introduction

Humans readily attribute socialness to artificial agents[b]. We erupt in anger at the computer that "knowingly" crashes in the middle of an important task, sympathize with a robot character in a film, or attribute a personality to an artificial personal assistant. The ease with which we ascribe agency and socialness to artificial entities has been exploited by writers, artists, and filmmakers for nearly a century, resulting in a rich fiction exploring the relationship between man and sentient machine. How do we make the transition from seeing a robot as a simple automaton to a sentient social being? This and related questions have puzzled scholars for centuries, from the discussion of the uniqueness of human social nature by Aristotle to the study of early automatons by Leonardo da Vinci and most recently the detailed empirical investigations by roboticists, psychologists, and cognitive neuroscientists.[1–3]

These days robots not only work alongside people on factory floors, but can increasingly be found in health care, education, and service industry settings as well. Similarly, with the increasing use of virtual and augmented reality, interactions with virtual humans are also likely to play a key role in the social fabric of society in the near future. Amidst

[a] Both authors contributed equally to this work.

[b] We use the term artificial agents to refer to robots (including those that are machine-like, pet-like, or human-like), virtual agents (including avatars of oneself or other virtual humans or characters), and artificial personal assistants (such as Siri, Cortana, Alexa, etc.).

doi: 10.1111/nyas.13727

this fourth industrial revolution, as social robots and other artificial agents become increasingly sophisticated and resolutely move from fiction to reality, important questions regarding the flexibility and adaptability of human social cognition when interacting with these entities require urgent attention.

Why might we perceive a robot as merely an automaton in some situations, while in other situations we see the same robot as an engaging social partner? Is this process of attributing socialness to artificial agents similar to that for attributing social characteristics to biological agents? To what extent is the same neural machinery we use to navigate our social world, refined over millennia of interacting with other people, co-opted when we reason about and interact with artificial agents? In this review, we aim to formulate answers to these complex and increasingly relevant social cognition questions by using an integrative approach that synthesizes the latest findings in psychology, social robotics, virtual reality, and neuroscience. We first review work on the role of artificial agents' appearance in eliciting responses at the brain and behavioral level. While this work documents how variations in an agents' visual features can shape human social engagement to a certain extent, we discuss noteworthy new findings incorporating insights from psychology and neuroscience on the impact of a person's prior knowledge, beliefs, or expectations about an artificial agent. In addition, we consider the processes of anthropomorphism and dehumanization on social engagement at brain and behavioral levels. We discuss a model that takes into account both perceptual and cognitive factors of human interactions with artificial agents and shows the functional convergence of cognitive factors driving socialness attribution to artificial agents within a specialized neural network. We conclude by discussing current challenges and future directions.

The attribution of socialness

A social interaction between two agents involves a complex cascade of expressions and reactions to social and emotional signals. During these interactions, we are not passive observers, but instead we actively construct the social nature of the other agent. We try to understand and explain the behavior and internal states of the other agent in terms of emotions, intentions, and beliefs. As we attribute emotions or intentions to other agents, artificial or not, we infer and attribute *socialness* to these agents. Socialness can be defined as the presence of intentional goal-directed recursive interactions with other beings. This process of attribution is closely related to the perception of other agents' minds,[4] or the representation of their mental state, also referred to as mentalizing or Theory-of-Mind (ToM).[5] If we attribute socialness to an agent, we adopt what Dennett[6,7] calls the intentional stance. We view and treat the agent as rational, with beliefs, desires, and behavioral consistency. This is in contrast to the design stance, in which we view and treat an agent based on knowledge of its function or design (e.g., predicting the response of a robot based on the knowledge of its software or actuators), or physical stance, in which we view and treat an agent based on knowledge derived from physics and chemistry (e.g., predicting the trajectory of a self-driving car based on its mass and velocity).

The attribution of socialness to artificial agents should not be viewed as a binary decision or an all-or-nothing process. An agent's socialness or the presence of intentional goal-directed recursive interactions with other beings is not dichotomous, but is instead a continuum constructed across multiple dimensions.[4,8] Researchers have described at least two dimensions that summarize different parts of socialness. People use experience, or the ability to sense and feel, and agency, the ability to plan and act, to distinguish the minds of agents.[9] For instance, data suggest that people rate an adult human being as high on both experience and agency, a dog or a chimp as high on experience and low on agency, and a robot as low on experience and medium on agency. Other researchers suggest a different, but related, distinction between warmth (similar to experience) and competence (similar to agency).[10] Yet, other researchers discuss a distinction based on human nature (aspects related to emotion that allow for a human–inanimate distinction) and human uniqueness (aspects related to morality that allow for a human–animal distinction).[11,12] Regardless of which distinction is used, the important point is that socialness can be distinguished on multiple dimensions. In order to be seen as a social being, an agent does not need to qualify as social across all dimensions (e.g., having both agency and experience). While other researchers have discussed humanness or animacy, derived from the Latin word *animat*, meaning "instilled with life," here, we mainly use socialness to highlight the notion of a continuum and multidimensionality, and to capture the full dimensions of these attributions. Socialness comprises capacities like actions, emotions, and intentions, and some scholars have argued that at its essence, a social agent is an agent that is capable of influencing the behavior of another agent.[13]

Crucially, the attribution of socialness is an ongoing, dynamic process between the perceived agent and the observing agent,[4] and is composed of many cues.[14] Some of these cues are derived from features of the artificial agent, such as its form and motion, and are referred to as bottom-up or stimulus cues to socialness. Above and beyond these cues, recent studies, as reviewed here, show that the prior knowledge of the observer, based upon beliefs, expectations, and experience, is key in the attribution of socialness. These observer-oriented factors are collectively referred to as top-down or knowledge cues to socialness. The distinction between these two types of cues provides the necessary framework to distil and disentangle the factors that influence the attribution of socialness to artificial agents.

At a basic level, brain regions associated with the person perception network (PPN), the action observation network (AON), and the ToM network have been shown to selectively respond to animate agents.[15,16] The PPN includes regions in the occipital and temporal cortex, such as the fusiform face area (FFA) and body area, occipital face area, extrastriate body area, and posterior superior temporal sulcus (STS).[17–19] Of course, these regions do not represent discrete animate and inanimate categories, but instead are responsive to a wide variety of stimuli.[20,21] Activity within these regions appears to index an observed agent's features (e.g., facial features, body posture, and motion),[17,18,22] the interactive nature of the situation,[23] context,[24,25] and perceived animacy and socialness.[26,27] The AON comprises frontoparietal regions spanning the posterior inferior frontal gyrus and inferior parietal lobule, which are engaged in a similar manner for executed and observed actions.[28–30] Activity within this network is also modulated by animacy and socialness. The very first work on this system demonstrated a distinction in parietal and premotor neural firing when nonhuman primates observed animate (a grasp being performed by a human hand) compared to inanimate actions (a grasp being performed by pliers).[31,32] Social processing performed by the PPN and AON is further informed by the ToM network, which comprises cortical regions spanning the medial prefrontal cortex (MPFC), temporoparietal junction (TPJ), precuneus, and temporal pole.[33–35] This network is crucial for inferring the mental state of other agents, including inanimate agents.[36]

As we show in the next sections, a mechanistic understanding of the attribution of socialness to artificial agents can be advanced through use of a social and cognitive neuroscientific lens. Some researchers have even suggested that measures of brain activity can serve as a "neural Turing test,"[37,38] a way of assessing the ability of an artificial agent to be indistinguishable from a human being.[39] Whether or not this is (yet) feasible, findings from social and cognitive neuroscience can nonetheless illuminate the factors underpinning the attribution of socialness to artificial agents.

Impact of artificial agent's visual features

Form

Form follows function, not just in the world of architecture, but also in the design of artificial agents. The first cue toward socialness is the form and shape of an observed agent. The perception of the face and body of a human agent provide access to a rich set of cues to socialness that facilitate subsequent behavior. Besides identity, the human face and body communicate emotions and intentions.[17–19] The perception of this information is partly influenced by stimulus cues, such as the shape and gender of an agent. Importantly, the perception of animacy is at the core of face processing.[40] By directly evaluating the perception of artificial compared to human agents, important first insights of the impact of artificial agents' visual features have come to light.

Several studies have looked at the pattern of activation in the PPN when observing emotions expressed by artificial agents. So far, both electroencephalography and functional magnetic resonance imaging (fMRI) studies suggest that activity within this network is not necessarily decreased by the appearance of the artificial agent.[41–47] Two influential studies provided the first insight into the effect of stimulus cues of socialness.[45,46] In both studies, participants observed a wide variety of emotional facial expressions (e.g., happiness, disgust, and anger) made by a humanoid robot or a human. Regardless of the instruction to either passively observe or actively rate the expressions, similar findings emerged in the neuroimaging data. Specifically, both studies reported no attenuation in activity within the PPN when observing robotic facial expressions. In terms of the response profile of individual regions, activity within the superior temporal gyrus did not discriminate between humanoid robot or human facial expressions, while activity in the FFA was increased for the robotic face compared to the human face.[45,46] However, the study by Gobbini and colleagues[46] reported the first evidence of decreased activity in the ToM network, specifically the right MPFC and the right TPJ, when observing artificial agents. This finding has recently been corroborated by Wang and Quadflieg[47] in a study on the perception of human–robot interactions. In this study, participants observed a human interacting with either another human or a humanoid robot, and were instructed to indicate if one agent was helping the other agent. This instruction was given to focus participants' attention on the relational aspect of the interaction between the two agents. Similar to previous findings, no robust differences were observed between the perception of human–robot interactions compared with the perception of human–human interactions within the PPN. Only three out of 10 regions in this network, the right FFA and bilateral posterior STS, showed greater activation for perception of human–human interaction compared with human–robot interaction. However, this study also showed sensitivity to the socialness of the agent within the ToM network, with more activation for human–robot interaction in the vMPFC and precuneus, but less activation in the left TPJ during the observation of these interactions compared to human–robot interaction.

These findings of overlap between artificial and human agents at the level of the PPN are complimented by studies probing this network using schematic faces or bodies.[18,22] With minimal cues present, people readily see faces in face-like objects or even random patterns, with similar activation patterns observed within dedicated brain areas implicated in person perception.[48,49] Differences in how an agent's appearance impacts person perception compared to the cognitive processes of ToM are further borne out by behavioral findings. For instance, emotions expressed by artificial agents, especially in the case of negative emotions, are sometimes difficult to recognize by human observers.[50] Martini *et al.*[51] directly investigated the role of a human-like appearance of an artificial agent on the attribution of different states to the artificial agent, including emotions, goals, and agency. Findings from two experiments suggest that the attribution of these states is a two-step process. While observers seldom attribute socialness to artificial agents below a specific threshold of human-likeness, this attribution linearly increases as the artificial agents gain an increasing number of human-like features. Studies examining these effects on social interactions with artificial agents show decreased human cooperation during direct economic interactions with a small humanoid robot compared to a person,[52] and that people are prone to punish artificial agents more than people.[53,54] In sum, these findings suggest a differential impact of artificial agents' form on processing in the PPN and ToM network. Whereas the PPN might not rely as heavily on them being human-like, the ToM network and related behaviors might. It should be noted that while the study of perceiving artificial agents is an emerging topic, the few studies published have mostly been limited to facial expressions, with the exception of one study on perceiving whole-body interactions between humans and humanoid robots.[47] This reflects a similar face-centric bias observed in studies on the perception of human social signals.[55] As such, it will be valuable for future studies to use a larger variety of social signals to probe the effect of artificial agent's form on engagement of the PPN and ToM network, in order to build a more complete picture of social perception.

Movement

Numerous studies document how the human brain reliably extracts a wealth of socially relevant information from simple motion cues. Since the seminal work by Johansson on point-light displays,[56] it has been shown that videos featuring a handful of points following a biological motion profile, containing no further information on the form of the agent, can be used to distil not only the direction and type of actions performed by an actor, but also the actor's emotions, gender, and identity.[57] Some researchers have argued that biological motion might serve as a "life-detector,"[58–60] which helps us to detect conspecifics and other animals. In an important early study, Pelphrey and colleagues[61] showed that the STS is selective to biological motion cues but not the form of the agent. While biological motion is clearly an important social cue, with some researchers arguing for a biological tuning of the motion node of the PPN,[62] as well as the human motor system and the AON,[63–65] open questions remain concerning whether biological motion is necessary for engagement of each of these networks.

First insights into such questions can be found via studies on the attribution of socialness to simple animated shapes. Since the influential work by Heider and Simmel,[66] multiple studies found that the observation of animations of simple shapes or social animations featuring nonbiological, but self-propelled, motion not only triggers the attribution of goals and intentions to these shapes,[67,68] but also robustly activates the posterior part of the STS (pSTS),[36,69–71] a core node of the PPN. Activity in this region increases when movement parameters suggest an interaction between animated shapes and decreases when movement parameters suggest less interactive and more random motion, implying a "perception of animacy" response gradient.[72] A recent study provides further insights into the role of pSTS that extend beyond motion *per se*.[73] In two experiments, the authors presented participants with short clips of point-light displays with two agents interacting or completing individual actions, as well as animations of simple shapes engaging in helping or hindering social interactions. Results showed that pSTS maximally responds to social interactions between point-light figures as well as simple shapes. Further analysis revealed that activity in this region does not depend on shape, goal, or animacy of the agent, *per se.* Crucially, the pSTS appears to be specifically sensitive to decoding the *nature* of these interactions, whether the interaction was helpful or hindering. Together, these findings suggest a role for the pSTS that is more flexible, moving beyond mere selectivity for biological motion.[62] While these findings complement previously discussed findings of the role of the PPN, it is important to note that perception of social animations also engages brain activity beyond this network. For example, the goal-directed movement of simple shapes triggers activation in the anterior intraparietal sulcus, part of the AON,[74] similar to human goal-directed movements,[75] and social animations can be used to functionally localize the ToM network[76] and robustly activate the TPJ.[77]

These findings are corroborated by studies examining automatic imitation during human–robot interactions. Automatic imitation studies seek to quantify the reflexive imitation of observed behavior, in this case, the interference of an observed robot's movements on the human perceiver's executed movements. Some evidence suggests that a biological motion profile of an observed agent impacts an observer's ongoing or subsequent movement more than a nonbiological motion profile,[64,78,79] and that automatic imitation is greater for robotic movements with quasi-biological motion.[80,81] However, other studies call into question the necessity of biological motion of an observed action in order to interfere with the observer's motor performance. While automatic imitation of robotic actions is smaller in absolute value compared to human actions, it is not completely absent, and several studies document automatic imitation of movements made by real and virtual full-body humanoid robots,[80–83] regardless of the presence of biological motion. A key factor driving automatic imitation appears to be the presence of human-like joint configuration, and not human-like motion *per se.*[84] Interference effects of observed movements on executed movements are reported for both humanoid robot and mechanical robot arms, as long as the latter had human-like joint configurations. No interference effects are observed if the mechanical robot arm had a nonhuman joint configuration, despite having quasi-biological motion. Interference effects are also found for apparent motion movements made by robotic hands.[85–88] Thus, little direct evidence exists for biological tuning at the behavioral level.

In contrast to an early study,[65] the majority of studies report that activity in the AON is not reliably decreased when observing actions performed by artificial agents compared to humans.[37,89–92] Indeed, motion parameters and an agent's appearance appear to not impact the AON in isolation, but rather in combination, and based on context.[89,90] In an innovative study, Saygin and colleagues[89] compared the observation of simple actions performed by a human, an android, or a robot. Importantly, the android and robot shared an identical motion profile and only differed in appearance. This was achieved by removing or replacing all of the external "human-like" features of the android so the appearance looked far more mechanical. Results showed most AON engagement for android compared to human or robot movements. Thus, activity in the AON appears to be mediated by an interaction between form (human-like) and motion (machine-like). The authors interpreted their finding in terms of a predictive coding model.[93] In this model,[94] observing unfamiliar actions can lead to increased activity in the AON due to greater prediction error. Further evidence for the importance of familiarity above and beyond the effect of motion parameters and appearance comes from a study by Cross and colleagues.[90] Across two experiments, they showed that AON activity was reliably greater when participants watched unfamiliar robotic dancing movements compared to natural dancing movements, regardless of whether these movements were performed by a person or a robot. These studies provide critical evidence that it is not simply stimulus cues, like an agent's form and motion, that drive engagement of brain regions involved in social perception, but also an observer's previous experience, familiarity, and expectations about how an artificial or human agent moves. In sum, while biological motion is an important cue to socialness, a number of lines of evidence suggest that it is not necessary. Instead, the core networks implicated in social perception can be flexibly engaged when observing an artificial agent in action, depending on a number of other mostly stimulus-independent factors.

Presence

The first visual encounter with an artificial agent provides a human observer with a first, albeit partial, indication of the agent's socialness. This understanding is only partial at first, since socialness attribution is a dynamic, emergent property of the active social interaction between two or more agents.[95] One important feature that supports social interactions is the physical colocation of both or all agents in the same environment. However, to maximize experimental control and efficiency, most studies so far merely explore how people *perceive* other (artificial) agents, which is a far leap from active social interaction during which each agent's ongoing behavior has the potential to influence and be influenced by the behavior of the other agent. Investigations that use agent observation as the main measure of interest focus on offline social cognition, while investigations employing reciprocal social interaction can delve more deeply into online social cognition, thereby tapping into distinct psychological and neural processes.[96] Physical embodiment and agent presence are crucial features for studying social interaction between human and artificial agents.[97,98] Besides sharing a virtual environment in virtual reality,[99] physical embodiment is another way to ensure that artificial and human agents share the same space.[100] A physically embodied artificial agent is a real, physical agent that is physically present in the same room as the human agent and allows for physical and face-to-face interaction between the two agents. A physically embodied agent can also be physically present in another room but presented on a screen, thereby reducing the presence of the artificial agent but maintaining the potential for face-to-face interaction. Finally, a virtually embodied artificial agent is a virtual construction of an artificial agent presented on a screen, thus having neither presence in the real world nor physical embodiment. Initial evidence documents the impact of embodiment and presence on the attribution of socialness.

Engaging in mutual gaze, compared to averted gaze, with a physically embodied robot increases engagement and can drive perceived human-likeness.[101] Confirming previous observations,[102] humans perceive a physically embodied and collocated robot more positively and persuasively than a visual representation of the same robot.[103] In addition, people also recognize physically present robots' emotions more accurately,[104] and even report higher levels of empathy for robots with whom they share the same space.[105,106] A recent study showed that the presence of an android robot directly influences perceived humanness and spontaneous mimicry by participants.[107] First, participants rated the android higher on human-likeness when it was collocated with them, compared to being presented on a computer screen. Second, while spontaneous mimicry was robustly observed across participants for a collocated android, only participants who rated the visually presented android higher on human-likeness showed spontaneous mimicry for this agent.

While evidence is so far limited to behavioral studies, indirect evidence of the impact of an artificial agent's physical embodiment and presence on social engagement at the brain level can be distilled from studies on gaze interaction.[97,108] Displaying the gaze behavior of a human agent via a virtual avatar can result in increased feelings of presence of the human agent and increased positive evaluations of this agent.[109] Schilbach and colleagues[110] used an interactive gaze task to tease apart the effect of self-initiated and other-initiated joint attention at behavioral and brain levels. They found that other-initiated joint attention increased activation in the MPFC, a core region of the ToM network, and self-initiated joint attention increased activation in the ventral striatum, a region associated with reward processing. Interestingly, activity in these regions is decreased when participants believe that the gaze behavior of the avatar has a computer origin.[111] These findings thus contribute to our understanding of how the physical embodiment and presence of an agent might potentially shape attributions of socialness, in addition to the agent's form and motion characteristics.

In sum, a number of studies have attempted to address the extent to which stimulus cues, including visual features, movement parameters, and presence of an artificial agent, influence behavioral and brain measures of socialness. To date, we argue that there is not enough evidence to suggest a reliable, clear impact of any of these features on socialness attributions. Instead, the work reviewed demonstrates that both the PPN and the AON are flexibly engaged when perceiving a diverse array of artificial agents, from schematic faces to social animations of shapes to mechanical and humanoid robots. Regarding the AON, several studies already suggest that it is not the appearance and motion of an artificial agent, but instead expectations or familiarity that might be more important in driving engagement of this network. In contrast, activity in the ToM network appears to be sensitive to the physical presence of an artificial agent, and whether or not there is a social narrative that can be ascribed to groups of animated shapes, but is not particularly sensitive to the presence or absence of biological motion. In contrast, several researchers have concluded that visual features of the artificial agent influence the attribution of socialness at the brain and behavioral level to some extent.[63,112] Naturally, it seems likely that the prospect of social behaviors being present, and thus the potential for a human interaction partner to attribute socialness, is higher in a robot designed to look like a human than a robot with a more machine-like appearance. However, the attribution of socialness will not necessarily follow from a human-like visual appearance of the artificial agent only. As we describe below, beliefs, expectations, and the prior experience of the individual all contribute to attributions of socialness above and beyond the artificial agent's appearance. For instance, the previous experience of the human agent with a sophisticated mechanical robot or the unrealistic high expectation of the individual for a humanoid robot can counteract any effect of visual appearance.[113]

Impact of knowledge cues the human observer

Belief and expectations

Turning our focus from the artificial agent to the human observer or interaction partner, in this section we explore how the knowledge, thoughts, beliefs, and expectations that a person brings to an interaction with an artificial agent shape the extent to which the agent is perceived in a social or nonsocial manner. In the classic cognitive psychology literature, such factors fall under the category of top-down processes. Top-down cognitive processes are endogenous to the perceiving/acting individual, driven by contexts, knowledge, or goals. Such processes can help facilitate perception, in that our past experience or knowledge can help us to formulate predictions about what is going to happen next.[114,115] Internal knowledge representations guide visual processing,[116] and importantly, the attribution of mental states influences the perception of social cues.[117] Below, we examine how research insights from psychology and neuroscience advance our understanding of the human side of human–robot interaction, and the importance of knowledge cues in attributing socialness to artificial agents and fostering social connections with these agents.

Some of the earliest work on the impact of knowledge cues on social perception comes from Stanley *et al.*[118] In this study, the authors used an elegant paradigm that required participants to follow the trajectory of a bouncing dot with their arm. The dot followed either a biologically plausible or biologically impossible (i.e., mechanical) velocity profile, and participants were instructed that the movement they were watching was either prerecorded human movement or computer-generated movement. The authors found that participants' belief about the human origins of the moving dot stimulus had a stronger impact on their actions than whether or not the velocity profile was biological or mechanical in nature. This led Stanley and colleagues to conclude that even when a cue is a simple bouncing dot, our beliefs about the human origins of a dot's movements can shape behavior more than the whether or not the dot moves in a biologically plausible manner. A related study by Liepelt and Brass[119] also examined the extent to which an observer's actions are influenced by his/her beliefs about an observed action's humanness, but this time, in a further step toward ecological validity, the stimuli featured hands performing finger lifting movements. Critically for our purposes, before this task, half the participants received instructions that the hand they were about to see was a human hand wearing a glove, while the other half were shown that the same gloved hand was in fact a wooden hand artist's model wearing a glove. The authors reported stronger automatic imitation in participants who were told they were watching a human hand wearing a glove, suggesting here again that our beliefs about the human origins of an action strongly shape the extent to which we behaviorally respond to them in a social manner.[119]

Converging evidence comes from a number of other studies employing various belief manipulations and paradigms.[120–127] For example, Wykowska, Wiese, and colleagues showed across a series of gaze cueing experiments that an observer's belief also shapes gaze following.[125,126] When people were told they were observing an intentional agent (a human or a human-controlled robot), gaze following was stronger than when they were told the agent had no intention (a robot or a human-like mannequin). While these studies contrast the belief that an agent is human versus a robot or machine, one study comparing a "human-like" robot versus a "machine-like" robot found similar effects during a joint action task.[121] Crucially, the effect of belief manipulations is related to true imitation,[124] rather than attentional effects in combination with general stimulus-response compatibility effects that can confound automatic imitation studies.[128] It is of note, however, that earlier work contrasting knowledge and stimulus cues found that only stimulus cues appeared to impact automatic imitation of hand movements.[87] On balance, behavioral findings build a largely (though not completely) consistent case for a strong influence of knowledge cues on shaping perceptions of socialness.

In the past several years, a number of brain imaging studies have further advanced our understanding of the influence of knowledge cues to socialness by revealing marked differences in neural processing based on participants' expectations about the human or artificial origins of a perceived agent. The first study along these lines was again performed by Stanley *et al.*[129] This time, the authors asked participants undergoing fMRI scanning to watch a number of point light animations of simple actions (such as walking, kicking a ball, lifting a box, etc.), which could either be presented as originally recorded, or with different levels of noise introduced so that the dots appeared to be moving in an increasingly random manner. Participants were notified before each video whether it featured human or computer-generated movement, and their task was to decide whether this label was accurate or inaccurate. The authors found that participants were more likely to agree that a stimulus looked like a person moving if they were told the video had human origins, while watching the identical video paired with a computer-generated label led to different behavioral responses, as well as different patterns of neural activation. Specifically, the ventral paracingulate cortex was most strongly engaged when watching those videos believed to have human origins, while the dorsal paracingulate cortex was most active when participants viewed ambiguous stimuli (such as scrambled dots with human instructions, or human action-like dots with computer-generated instructions).

A subsequent functional neuroimaging study used a similar automatic imitation paradigm to Liepelt and Brass's[119] paradigm, described above, with a few important differences.[130] First, this study employed a within-subjects design to determine the extent to which differences in stimulus and knowledge cues to human animacy can be observed in the same group of participants. Second, the authors asked participants to watch a 20-min bespoke documentary wherein two kinds of cutting-edge filmmaking techniques were described: human motion capture and computer keyframe animation. They were then shown two different hand stimuli—one that featured an avatar of realistic-looking human hand and the other featured two blocky robot-like fingers. Participants performed the same imitation for both kinds of "hand" stimuli, but half of the trials were preceded with instructions that the following stimuli were made from human motion capture, and the other half with the instructions that the following videos were made with computer keyframe animation. In reality, all stimuli followed the identical motion profile, regardless of instructions. Somewhat in contrast to what Liepelt and Brass[119] reported with their between-subjects behavioral study, Klapper and colleagues found that *any* cue to humanness led to greater motor priming compared to when no cues to humanness were present (in other words, more interference of executed movements was observed when participants observed a human hand avatar and/or either kind of hand paired with the instruction that the video had human motion capture origins). In contrast, the brain imaging data demonstrated that the right TPJ, a brain region that is often associated with cognitive processes that involve self-other distinctions, was most strongly engaged when participants were performing the imitation task when both knowledge and stimulus cues to human animacy were present.[130] The authors suggest that this finding underscores the critical role the right TPJ plays in mediating interactions with other human (but not artificial) agents. These findings are corroborated by a number of other previous and subsequent findings.[131–135] For instance, a recent study using an established gaze cueing paradigm[125,126,136] reported increased activation in bilateral TPJ for gaze following when participants thought the gaze had human origins compared to a preprogrammed, computer-determined origin.[134]

Another recent functional neuroimaging study helped to further illuminate the differential roles played by stimulus and knowledge cues in social perception.[14] In this study, Cross and colleagues made use of the same documentary used by Klapper and colleagues,[130] but this time paired the documentary with an elaborate cover story that the researchers were working for the German Film Commission and were tasked with evaluating participants' perceptions of the smoothness and likeability of different simple action movies, based on their human versus computer origins. Again, the motion profiles were the same for each video and half the videos featured a human avatar performing a number of simple goal-directed actions (such as tidying a table, stacking blocks, or hammering a nail), or an abstract-looking robotic avatar performing the same actions. Participants reported the highest movement smoothness and liking ratings for videos that were paired with human origin instructions (regardless of whether the actions were performed by a human or robot avatar). The brain data demonstrated that visual features of an agent appear to primarily influence ventral temporal brain regions, such as the fusiform gyrus, when observing a robot compared to a human avatar. This increase in activity in the fusiform gyrus for robotic compared to human agents is in accordance with previous studies on the perception of facial expressions of emotions in robotic agents.[45,46] Crucially, brain regions associated with ToM processes, including the precuneus, are more strongly engaged when videos are paired with human compared to computer-generated instructions. Taken together, the studies examined in this section begin to build a compelling case that the knowledge cues that a participant has when perceiving or interacting with another agent, whether human or artificial in appearance, have a strong impact on behavior and underlying brain circuits.

Anthropomorphism and dehumanization

Corroborating evidence of the central role played by knowledge cues to socialness comes from research on anthropomorphism. Anthropomorphism can be defined as our propensity to attribute human-like qualities, characteristics, and behaviors, to nonhuman agents, entities, or objects, by the observer. The tendency to anthropomorphize animals, robots, and computers is a stable characteristic that is robust over time,[137] and already present in young children.[134] Research on anthropomorphism suggests that it is not the visual features of the agent, but instead knowledge factors and the social motivation of the observer that are central in attributing socialness to nonhuman agents.[138]

Several studies have investigated the process of anthropomorphizing and corresponding brain networks. A recent study reported that an individual's disposition to anthropomorphize is related to gray matter volume of the left TPJ.[139] Specifically, they found that a greater tendency to anthropomorphize nonhuman animals is positively correlated with TPJ volume. However, the authors did not observe any correlation between brain structure and a disposition to anthropomorphize nonanimal agents or objects. While this already provides evidence on the role of the ToM network in anthropomorphism, further details come from studies measuring brain activity during the active process of anthropomorphizing.[140,141] An early study by Chaminade *et al.*[140] investigated how the visual appearance of an animated character influenced the perception of motion. The agents ranged from animated point-light displays, to a stick figure, robot, alien, clown, or a human-like jogger. Participants' task was to indicate if the agents' movement was biological or computer-generated. Importantly, all agents moved with the same motion parameters, which came in one condition directly from motion capture data of a human actor's movements (biological motion), while in the other condition they were computer-generated movements (nonbiological motion). Crucially, brain activity did not serve as a function of characters' appearance, but was instead modulated by participants' anthropomorphic bias (in this study, a participant's tendency to report the observed motion as biological). The authors reported a positive correlation between this response bias and activity in the ToM network, specifically the left TPJ and bilateral precuneus. Thus, a tendency to perceive the motion of the agents as biological tracked with increased engagement of the left TPJ. Interestingly, however, this response bias was negatively correlated with activity in regions of the AON.[140]

Besides the features of an artificial agent and an observer's belief, anthropomorphism is also influenced by the motivation to understand and predict the social environment.[142] Waytz and colleagues[141] showed that this so-called effectance motivation is directly related to a tendency to anthropomorphize artificial agents as well as objects. In a series of five experiments, they showed that when uncertainty and unpredictability, two key determinants of effectance motivation, increase, so does anthropomorphism. For example, when participants interacted with an unpredictable robot, they rated this robot higher on anthropomorphic aspects, such as having its own mind, intentions, free will, consciousness, desires, beliefs, and the ability to experience emotions, compared to participants who interacted with a predictable robot. When asked to predict the behavior of an unfamiliar robot, participants' anthropomorphic judgments also increased, compared to when participants were not asked to predict its behavior. Crucially, the authors showed that the tendency to anthropomorphize the actions/behaviors of unpredictable objects is related to increased activity within brain regions associated with the ToM network. Participants first read about gadgets that were either predictable or unpredictable. Next, they answered the question to what extent the gadget had a mind of its own. Behavioral results showed that again participants were more likely to attribute mind-like qualities to unpredictable gadgets compared to predictable gadgets. Importantly, activity in the ventral part of MPFC was increased for unpredictable compared to predictable gadgets. The involvement of the ToM network in these anthropomorphism judgments was further confirmed in a connectivity analysis, showing functional connectivity with the precuneus as well as the anterior cingulate cortex. Furthermore, activity in the ventral part of the MPFC directly covaried with anthropomorphism judgments across participants. In sum, a greater tendency to anthropomorphize is related to structural and functional changes within the ToM network. An area ripe for future exploration concerns how predictability and appearance of the artificial agent interact with the belief and tendency of the human observer to anthropomorphize and the role of the ToM network, especially the MPFC in these processes.[143]

The engagement of brain regions associated with the ToM network is also robustly implicated in studies on dehumanizing behavior, a similar but opponent process to the attribution of socialness to artificial agents. Dehumanization is the process by which a human agent or group of individuals is seen to possess fewer human-like qualities compared to another person or group of individuals. People not only view themselves as more human than others,[144] but can view other humans or groups as closer to animals or automata.[145] An early study by Harris and Fiske[146] found that activity in the MPFC was decreased when participants viewed pictures of members of an extreme outgroup that was rated low on both warmth (similar to experience or human nature) and competence (similar to agency or human uniqueness), such as homeless individuals or drug addicts. Activity in the MPFC, as well as other regions of the ToM network, has been documented to show similar effects for other aspects of dehumanization, such as sexual objectification of women,[147] or viewing people as products.[148] A recent study showed that activity in the ToM network, namely, the MPFC and right TPJ, is decreased when participants are told that the observed person is similar to a machine.[149] Findings from research on dehumanization will inform us further on the factors that might (negatively) influence the attribution of socialness to artificial agents.

Eyssel and colleagues showed across a series of experiments[150–152] that similar biases are present in human–robot interaction as in human–human interaction. That is, people favor a robot that belongs to their in-group and ascribe more human-like characteristics to these robots. Similar biases were observed for people interacting with a computer[153–155] or a virtual human.[156,157] Thus, an increase in the human-likeness of the artificial agent might increase the potential for biases or stereotypes to enter the socialness equation. Another influence on socialness attribution comes from the perspective put forward by Ferrari *et al.*[158] These authors suggest that artificial agents' threat to distinctiveness also has the potential to influence socialness attribution and social behavior toward artificial agents. The authors argue that robots with human-like appearances threaten to blur the boundaries between humans, machines, and other artificial agents. This perception of threat to human uniqueness can therefore reduce the attribution of socialness to artificial agents. This process could be similar to dehumanization of human agents. Thus, the attribution of socialness to artificial agents is a highly contextual process that depends on the observer, the artificial agent, and the environment.[159]

In sum, these studies confirm the vital role played by the ToM network, in particular the TPJ and vMPFC (but see Ref. 160) in the attribution of socialness and extend our understanding of factors that influence these attributions. It is not so much the features of the artificial agent that drive the attribution of socialness to artificial agents, but the belief and expectations of the human observer as well as her/his tendency to attribute human-like qualities to human and artificial agents as seen in the process of anthropomorphism and dehumanization.

Integrative perspective

When attempting to link together the broad range of findings covered in this review, it is instructive to ask whether a common theoretical thread links much of this work examining whether, how, and when people engage with artificial agents in a social manner. In light of this, we argue that a particular theoretical position borrowed from developmental psychology is useful for framing past findings and future questions about humans' social future with artificial agents. This theoretical position, termed the "like me" hypothesis, states that understanding the basic similarity between self and other forms the foundation of social cognition, and that humans have evolved to seek out self-other equivalence in others.[161,162] This account further proposes that actions performed by oneself and another are represented in common cognitive codes,[161] and early neurophysiological work establishing the existence of mirror neurons in the nonhuman primate brain[31,32] has helped inspire much of the foundational work in this domain.

However, attributing socialness to another agent involves far more than linking perceived and executed actions between oneself and an observed other, as we have seen throughout this review. The representation of other agents' minds is a three-stage developmental process,[159] which involves not just linking observed and executed acts (via imitation), but also the first-person everyday experience of action-intention coupling, and finally the ultimate step of understanding other agents by projection of one's own state onto another agent. This correspondence between self and other at multiple levels is vital to the attribution of socialness to artificial agents. As we have seen, this process is flexible in nature, and the "like me" hypothesis provides a useful point of departure for contextualizing brain and behavioral findings about the impact of form, motion, knowledge, experience, anthropomorphizing, and dehumanization on the attribution of socialness to artificial agents. For instance, some of the studies reviewed above adhere neatly to the "like me" hypothesis by demonstrating evidence of behavioral overlap[63,64,78,79] and increased engagement of the AON when participants observe familiar actions or interact with agents similar to themselves,[65,129,130] while others do not. For example, other studies have demonstrated similar or greater AON engagement when participants observe very much *unlike*-me robotic actions compared to more familiar human actions,[37,89–92] with similar findings for behavior engagement.[80–83,85–88] Likewise, almost no correspondence with the "like me" hypothesis is found at the level of the PPN.

However, neuroimaging findings consistently support the "like-me" nature of ToM network engagement, especially within the MPFC

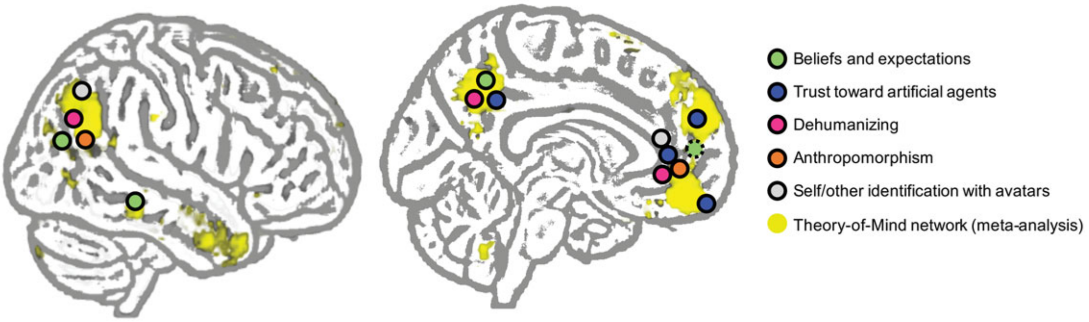
Beliefs and expectations
Trust toward artificial agents
Dehumanizing
Anthropomorphism
Self/other identification with avatars
Theory-of-Mind network (meta-analysis)

Figure 1. **Functional convergence of cognitive factors driving socialness attribution to artificial agents within the Theory-of-Mind network.** Studies on knowledge factors, anthropomorphism, and dehumanization robustly report engagement of the temporoparietal junction, precuneus, and dorsal and ventral medial prefrontal cortex, regions within the Theory-of-Mind network. The dots illustrate the clusters of activation found in the studies and do not reflect exact coordinates. The automated term-based meta-analytic brain activation map for the Theory-of-Mind network was created and downloaded from the Neurosynth database (http://neurosynth.org, December 1, 2017).[176] The maps are based on 124 studies using the term "mentalizing" and are corrected at FDR < 0.01.

and the TPJ, when interacting with socially similar others. Stimulus cues,[46,47,111] knowledge cues,[130–135] anthropomorphism,[140,141] and dehumanization[146–149] processes all impact activity in these regions (Fig. 1). Further evidence for MPFC-mediated self-other equivalence comes from studies on self/other identification with a virtual avatar[163] or trust during interactions with an artificial agent.[164] Similarly, a wealth of evidence from human–human interaction supports the notion of a crucial role of the TPJ in inferring the mental states of others,[33–35] in differentiating the self from others during joint attention and perspective taking tasks,[77,165] as well when making judgments of in-group versus out-group members.[165]

A recent perspective on the TPJ suggests that it codes information about social context.[165,166] This perspective dovetails with a mechanistic model on downstream effects of mental state attribution,[117] whereby the attribution of a mind in an observed agent influences perception of social cues associated with this agent. A study by Carter and colleagues[177] provides evidence for this social bias model during interactions with artificial agents. In this study, participants played a game of poker against human or computer opponents. The authors used activity from 55 bilateral brain regions to predict the decisions of the participant throughout the game. While activity in regions associated with the ToM network was predictive for these decisions, this was independent of social context. Only activity within the TPJ predicted future decisions while taking into account the social context. That is, only when the participant deemed the human opponent superior to the computer opponent, and therefore more socially relevant, was activity in the TPJ informative of subsequent decisions. Combined, these findings suggest that the highly context-dependent implicit or explicit decision of the observer that the agent is "like me," partly coded in the TPJ, is key in fostering the attribution of socialness to artificial agents.

Conclusions and future directions

The studies reviewed here support the notion that knowledge cues consistently impact engagement of behavioral and brain mechanisms supporting the attribution of socialness to artificial agents. The data also suggest that knowledge cues play an extremely (if not more) important role in socialness attributions than stimulus cues. Knowledge cues, as well as the process of anthropomorphism and dehumanization, influence ToM network engagement and determine the attribution of socialness. While the research reviewed here begins to provide answers as to why we can sometimes perceive a robot as an automaton, and at other times as a social agent, several questions remain.

One important question concerns how socialness attributions unfold over time. Most studies so far have examined passive observation of artificial agents or one-off interactions. A challenge for future work is to investigate real, repeated, and ongoing interactions with artificial agents in order to map functional changes in socialness attribution at behavioral and brain levels across time.[97,98] Several examples already exist that show how incorporating the temporal dimension of human–robot *relationships* can enrich our understanding of the mechanisms of socialness attribution.[167] For instance, repeated exposure to robotic actions was shown to increase automatic imitation of these actions to levels comparable to human actions.[88] Moreover, several recent studies on trust during interactions with artificial agents or machines provide compelling first evidence on how expectations and changes in the involvement of the ToM network shape these interactions over time.[164,168,169] An exciting opportunity for future research is to study in more detail the dynamic, temporal dimensions of these processes.

Work in this domain already builds upon a solid foundation of findings from psychology and neuroscience on social cognition during human–human interaction. Integrating findings and approaches from additional related fields stands to advance understanding of not only the factors that drive the attribution of socialness to artificial agents, but also the temporal dynamics. To this end, future work could benefit from considering work on human attachment and relationship formation,[170] as well as the emerging field of human–animal interactions.[171–173] Studies on these latter interactions not only provide converging evidence on behavioral and brain mechanisms of socialness attribution,[172] but also can help us to understand how long-term interaction with nonhuman agents shapes these attributions over time by studying pet owners versus nonpet owners.[171,173] Thus, thinking openly and creatively about how work from distinct but complementary disciplines might inform our understanding about humans' evolving relationship with socially savvy technology.

Given the importance of knowledge factors during interactions with artificial agents, as we highlight here, it will also be enormously important for future research to acknowledge and investigate interindividual and group differences, as well as developmental changes in the attributions of socialness. Dispositional levels of anthropomorphism and dehumanization work in concert with knowledge and stimulus cues. Similarly, not only are developmental considerations important to further our understanding of the behavioral and brain mechanisms supporting the attribution of socialness, they are crucial given that both children and the elderly are target groups for the deployment of social robots, and age-dependent effects have been reported with regard to socialness attribution.[174,175] Finally, group and cultural differences that determine the scope of attribution of socialness are likely at play. With the potential for far-reaching consequences, for instance on perceived human uniqueness, a nuanced approach that takes into account dispositional, situational, and cultural factors is warranted to truly capture the mechanisms and dynamics of socialness attribution across space and time.

Acknowledgments

The authors thank Kami Koldewyn, Anna Henschel, and Katie A. Riddoch, as well as the anonymous reviewers for valuable comments on an earlier version of the manuscript and gratefully acknowledge funding from the European Research Council to E.S.C. (H2020 ERC-2015-StG-67720-SOCIAL ROBOTS).

Competing interests

The authors declare no competing interests.

References

1. Broadbent, E. 2017. Interactions with robots: the truths we reveal about ourselves. *Annu. Rev. Psychol.* **68:** 627–652.
2. Chaminade, T. & G. Cheng. 2009. Social cognitive neuroscience and humanoid robotics. *J. Physiol. Paris* **103:** 286–295.
3. Dautenhahn, K. 2007. Socially intelligent robots: dimensions of human–robot interaction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362:** 679–704.
4. Waytz, A., K. Gray, N. Epley, *et al.* 2010. Causes and consequences of mind perception. *Trends Cogn. Sci.* **14:** 383–388.
5. Frith, U. & C.D. Frith. 2009. The social brain: allowing humans to boldly go where no other species has been. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365:** 165–176.
6. Dennett, D.C. 1989. *The Intentional Stance.* Cambridge, MA: MIT Press.
7. Dennett, D.C. 1971. Intentional systems. *J. Philos.* **68:** 87–106.
8. Haslam, N. & S. Loughnan. 2014. Dehumanization and infrahumanization. *Annu. Rev. Psychol.* **65:** 399–423.
9. Gray, H.M., K. Gray & D.M. Wegner. 2007. Dimensions of mind perception. *Science* **315:** 619.
10. Fiske, S.T., A.J.C. Cuddy, P. Glick, *et al.* 2002. A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *J. Pers. Soc. Psychol.* **82:** 878–902.

11. Haslam, N. 2006. Dehumanization: an integrative review. *Pers. Soc. Psychol. Rev.* **10:** 252–264.
12. Kagan, J. 2004. The uniquely human in human nature. *Daedalus* **133:** 77–88.
13. Mason, P. & H. Shan. 2017. A valence-free definition of sociality as any violation of inter-individual independence. *Proc. R. Soc. Lond. B Biol. Sci.* **284:** pii: 20170948.
14. Cross, E.S., R. Ramsey, R. Liepelt, *et al.* 2016. The shaping of social perception by stimulus and knowledge cues to human animacy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371:** 20150075.
15. Adolphs, R. 2009. The social brain: neural basis of social knowledge. *Annu. Rev. Psychol.* **60:** 693–716.
16. Frith, C.D. & U. Frith. 2012. Mechanisms of social cognition. *Annu. Rev. Psychol.* **63:** 287–313.
17. de Gelder, B. 2006. Towards the neurobiology of emotional body language. *Nat. Rev. Neurosci.* **7:** 242–249.
18. Peelen, M.V. & P.E. Downing. 2007. The neural basis of visual body perception. *Nat. Rev. Neurosci.* **8:** 636–648.
19. Adolphs, R. 2017. Emotion perception from face, voice, and touch: comparisons and convergence. *Trends Cogn. Sci.* **21:** 216–228.
20. Huth, A.G., S. Nishimoto, A.T. Vu, *et al.* 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* **76:** 1210–1224.
21. Kriegeskorte, N., M. Mur, D.A. Ruff, *et al.* 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* **60:** 1126–1141.
22. Grill-Spector, K., K.S. Weiner, K. Kay, *et al.* 2017. The functional neuroanatomy of human face perception. *Annu. Rev. Vis. Sci.* **3:** 167–196.
23. Quadflieg, S. & K. Koldewyn. 2017. The neuroscience of people watching: how the human brain makes sense of other people's encounters. *Ann. N.Y. Acad. Sci.* **1396:** 166–182.
24. Van den Stock, J., M. Vandenbulcke, C.B.A. Sinke, *et al.* 2014. Affective scenes influence fear perception of individual body expressions. *Hum. Brain Mapp.* **35:** 492–502.
25. Van den Stock, J., M. Vandenbulcke, C.B.A. Sinke, *et al.* 2014. How affective information from faces and scenes interacts in the brain. *Soc. Cogn. Affect. Neurosci.* **9:** 1481–1488.
26. Shultz, S. & G. McCarthy. 2014. Perceived animacy influences the processing of human-like surface features in the fusiform gyrus. *Neuropsychologia* **60:** 115–120.
27. Looser, C.E., J.S. Guntupalli & T. Wheatley. 2013. Multivoxel patterns in face-sensitive temporal regions reveal an encoding schema based on detecting life in a face. *Soc. Cogn. Affect. Neurosci.* **8:** 799–805.
28. Caspers, S., K. Zilles, A.R. Laird, *et al.* 2010. ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage* **50:** 1148–1167.
29. Keysers, C. & V. Gazzola. 2009. Expanding the mirror: vicarious activity for actions, emotions, and sensations. *Curr. Opin. Neurobiol.* **19:** 666–671.
30. Molenberghs, P., R. Cunnington & J.B. Mattingley. 2012. Brain regions with mirror properties: a meta-analysis of 125 human fMRI studies. *Neurosci. Biobehav. Rev.* **36:** 341–349.
31. Gallese, V., L. Fadiga, L. Fogassi, *et al.* 1996. Action recognition in the premotor cortex. *Brain* **119**(Pt 2): 593–609.
32. Rizzolatti, G., L. Fadiga, V. Gallese, *et al.* 1996. Premotor cortex and the recognition of motor actions. *Brain Res. Cogn. Brain Res.* **3:** 131–141.
33. Frith, U. & C.D. Frith. 2003. Development and neurophysiology of mentalizing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358:** 459–473.
34. Molenberghs, P., H. Johnson, J.D. Henry, *et al.* 2016. Understanding the minds of others: a neuroimaging meta-analysis. *Neurosci. Biobehav. Rev.* **65:** 276–291.
35. Schurz, M., J. Radua, M. Aichhorn, *et al.* 2014. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci. Biobehav. Rev.* **42:** 9–34.
36. Castelli, F., F. Happé, U. Frith, *et al.* 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage* **12:** 314–325.
37. Oberman, L.M., J.P. McCleery, V.S. Ramachandran, *et al.* 2007. EEG evidence for mirror neuron activity during the observation of human and robot actions: toward an analysis of the human qualities of interactive robots. *Neurocomputing* **70:** 2194–2203.
38. Saygin, A.P., I. Cicekli & V. Akman. 2000. Turing test: 50 years later. *Minds Mach.* **10:** 463–518.
39. Turing, A.M. 1950. Computing machinery and intelligence. *Mind* **59:** 433–460.
40. Koldewyn, K., P. Hanus & B. Balas. 2014. Visual adaptation of the perception of "life": animacy is a basic perceptual dimension of faces. *Psychon. Bull. Rev.* **21:** 969–975.
41. Dubal, S., A. Foucher, R. Jouvent, *et al.* 2011. Human brain spots emotion in non humanoid robots. *Soc. Cogn. Affect. Neurosci.* **6:** 90–97.
42. Chammat, M., A. Foucher, J. Nadel, *et al.* 2010. Reading sadness beyond human faces. *Brain Res.* **1348:** 95–104.
43. Craig, R., R. Vaidyanathan, C. James, *et al.* 2010. Assessment of human response to robot facial expressions through visual evoked potentials. In *2010 10th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, Nashville, TN, pp. 647–652.
44. Moser, E., B. Derntl, S. Robinson, *et al.* 2007. Amygdala activation at 3T in response to human and avatar facial expressions of emotions. *J. Neurosci. Methods* **161:** 126–133.
45. Chaminade, T., M. Zecca, S.-J. Blakemore, *et al.* 2010. Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PLoS One* **5:** e11577.
46. Gobbini, M.I., C. Gentili, E. Ricciardi, *et al.* 2011. Distinct neural systems involved in agency and animacy detection. *J. Cogn. Neurosci.* **23:** 1911–1920.
47. Wang, Y. & S. Quadflieg. 2015. In our own image? Emotional and neural processing differences when observing human–human vs human–robot interactions. *Soc. Cogn. Affect. Neurosci.* **10:** 1515–1524.

48. Hadjikhani, N., K. Kveraga, P. Naik, *et al.* 2009. Early (M170) activation of face-specific cortex by face-like objects. *Neuroreport* **20:** 403–407.
49. Liu, J., J. Li, L. Feng, *et al.* 2014. Seeing Jesus in toast: neural and behavioral correlates of face pareidolia. *Cortex* **53:** 60–77.
50. Hortensius, R., F. Hekele & E.S. Cross. 2017. The perception of emotion in artificial agents. https://doi.org/10.17605/OSF.IO/UFZ5W.
51. Martini, M.C., C.A. Gonzalez & E. Wiese. 2016. Seeing minds in others—can agents with robotic appearance have human-like preferences? *PLoS One* **11:** e0146310.
52. Sandoval, E.B. 2016. Reciprocity in human–robot interaction: a quantitative approach through the prisoner's dilemma and the ultimatum game. *Int. J. Soc. Robot.* **8:** 303–317.
53. Bartneck, C. & J. Hu. 2008. Exploring the abuse of robots. *Interact. Stud.* **9:** 415–433.
54. Slater, M., A. Antley, A. Davison, *et al.* 2006. A virtual reprise of the Stanley Milgram obedience experiments. *PLoS One* **1:** e39.
55. de Gelder, B. & R. Hortensius. 2014. The many faces of the emotional body. In *New Frontiers in Social Neuroscience.* J. Decety & Y. Christen, Eds.: 153–164. Berlin, Cham: Springer.
56. Johansson, G. 1973. Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* **14:** 201–211.
57. Puce, A., A. Rossi & F.J. Parada. 2015. Biological motion. *Brain Mapp.* **3:** 125–130.
58. Simion, F., E. Di Giorgio, I. Leo, *et al.* 2011. The processing of social stimuli in early infancy: from faces to biological motion perception. *Prog. Brain Res.* **189:** 173–193.
59. Troje, N.F. & C. Westhoff. 2006. The inversion effect in biological motion perception: evidence for a "life detector"? *Curr. Biol.* **16:** 821–824.
60. Johnson, M.H. 2006. Biological motion: a perceptual life detector? *Curr. Biol.* **16:** R376–R377.
61. Pelphrey, K.A., T.V. Mitchell, M.J. McKeown, *et al.* 2003. Brain activity evoked by the perception of human walking: controlling for meaningful coherent motion. *J. Neurosci.* **23:** 6819–6825.
62. Blake, R. & M. Shiffrar. 2007. Perception of human motion. *Annu. Rev. Psychol.* **58:** 47–73.
63. Press, C. 2011. Action observation and robotic agents: learning and anthropomorphism. *Neurosci. Biobehav. Rev.* **35:** 1410–1418.
64. Kilner, J.M., Y. Paulignan & S.-J. Blakemore. 2003. An interference effect of observed biological movement on action. *Curr. Biol.* **13:** 522–525.
65. Tai, Y.F., C. Scherfler, D.J. Brooks, *et al.* 2004. The human premotor cortex is "mirror" only for biological actions. *Curr. Biol.* **14:** 117–120.
66. Heider, F. & M. Simmel. 1944. An experimental study of apparent behavior. *Am. J. Psychol.* **57:** 243–259.
67. Johnson, S.C. 2003. Detecting agents. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358:** 549–559.
68. Opfer, J.E. 2002. Identifying living and sentient kinds from dynamic information: the case of goal-directed versus aimless autonomous movement in conceptual change. *Cognition* **86:** 97–122.
69. Wheatley, T., S.C. Milleville & A. Martin. 2007. Understanding animate agents: distinct roles for the social network and mirror system. *Psychol. Sci.* **18:** 469–474.
70. Schultz, J., H. Imamizu, M. Kawato, *et al.* 2004. Activation of the human superior temporal gyrus during observation of goal attribution by intentional objects. *J. Cogn. Neurosci.* **16:** 1695–1705.
71. Blakemore, S.-J., P. Boyer, M. Pachot-Clouard, *et al.* 2003. The detection of contingency and animacy from simple animations in the human brain. *Cereb. Cortex* **13:** 837–844.
72. Schultz, J., K.J. Friston, J. O'Doherty, *et al.* 2005. Activation in posterior superior temporal sulcus parallels parameter inducing the percept of animacy. *Neuron* **45:** 625–635.
73. Isik, L., K. Koldewyn, D. Beeler, *et al.* 2017. Perceiving social interactions in the posterior superior temporal sulcus. *Proc. Natl. Acad. Sci. USA* **114:** E9145–E9152.
74. Ramsey, R. & F.C. Hamilton. 2010. Triangles have goals too: understanding action representation in left aIPS. *Neuropsychologia* **48:** 2773–2776.
75. Hamilton, A.F. & S.T. Grafton. 2006. Goal representation in human anterior intraparietal sulcus. *J. Neurosci.* **26:** 1133–1137.
76. Jacoby, N., E. Bruneau, J. Koster-Hale, *et al.* 2016. Localizing pain matrix and theory of mind networks with both verbal and non-verbal stimuli. *Neuroimage* **126:** 39–48.
77. Schurz, M., M.G. Tholen, J. Perner, *et al.* 2017. Specifying the brain anatomy underlying temporo–parietal junction activations for theory of mind: a review using probabilistic atlases from different imaging modalities. *Hum. Brain Mapp.* **38:** 4788–4805.
78. Kilner, J., A.F. Hamilton & S.-J. Blakemore. 2007. Interference effect of observed human movement on action is due to velocity profile of biological motion. *Soc. Neurosci.* **2:** 158–166.
79. Bisio, A., A. Sciutti, F. Nori, *et al.* 2014. Motor contagion during human–human and human–robot interaction. *PLoS One* **9:** e106172–10.
80. Cook, J.L., D. Swapp, X. Pan, *et al.* 2014. Atypical interference effect of action observation in autism spectrum conditions. *Psychol. Med.* **44:** 731–740.
81. Chaminade, T., D.W. Franklin, E. Oztop, *et al.* 2005. Motor interference between humans and humanoid robots: effect of biological and artificial motion. In *Proceedings of the 4th International Conference on Development and Learning.* 96–101. Osaka: IEEE.
82. Oztop, E., D.W. Franklin & T. Chaminade. 2005. Human–humanoid interaction: is a humanoid robot perceived as a human? *Int. J. Humanoid Robotics* **2:** 537–559.
83. Hofree, G., B.A. Urgen, P. Winkielman, *et al.* 2015. Observation and imitation of actions performed by humans, androids, and robots: an EMG study. *Front. Hum. Neurosci.* **9:** 364.

84. Kupferberg, A., M. Huber, B. Helfer, *et al.* 2012. Moving just like you: motor interference depends on similar motility of agent and observer. *PLoS One* **7:** e39637–e39638.
85. Press, C., G. Bird, R. Flach, *et al.* 2005. Robotic movement elicits automatic imitation. *Brain Res. Cogn. Brain Res.* **25:** 632–640.
86. Bird, G., J. Leighton, C. Press, *et al.* 2007. Intact automatic imitation of human and robot actions in autism spectrum disorders. *Proc. R. Soc. Lond. B Biol. Sci.* **274:** 3027–3031.
87. Press, C., H. Gillmeister & C. Heyes. 2006. Bottom-up, not top-down, modulation of imitation by human and robotic models. *Eur. J. Neurosci.* **24:** 2415–2419.
88. Press, C., H. Gillmeister & C. Heyes. 2007. Sensorimotor experience enhances automatic imitation of robotic action. *Proc. R. Soc. Lond. B Biol. Sci.* **274:** 2509–2514.
89. Saygin, A.P., T. Chaminade, H. Ishiguro, *et al.* 2012. The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* **7:** 413–422.
90. Cross, E.S., R. Liepelt, A.F. Hamilton, *et al.* 2012. Robotic movement preferentially engages the action observation network. *Hum. Brain Mapp.* **33:** 2238–2254.
91. Gazzola, V., G. Rizzolatti, B. Wicker, *et al.* 2007. The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage* **35:** 1674–1684.
92. Urgen, B.A., M. Plank, H. Ishiguro, *et al.* 2013. EEG theta and Mu oscillations during perception of human and robot actions. *Front. Neurorobot.* **7:** 19.
93. Saygin, A.P. & W. Stadler. 2012. The role of appearance and motion in action prediction. *Psychol. Res.* **76:** 388–394.
94. Kilner, J.M., K.J. Friston & C.D. Frith. 2007. Predictive coding: an account of the mirror neuron system. *Cogn. Process.* **8:** 159–166.
95. Gangopadhyay, N. & L. Schilbach. 2012. Seeing minds: a neurophilosophical investigation of the role of perception–action coupling in social perception. *Soc. Neurosci.* **7:** 410–423.
96. Schilbach, L. 2014. On the relationship of online and offline social cognition. *Front. Hum. Neurosci.* **8:** 278.
97. Schilbach, L., B. Timmermans, V. Reddy, *et al.* 2013. Toward a second-person neuroscience. *Behav. Brain Sci.* **36:** 393–414.
98. Wykowska, A., T. Chaminade & G. Cheng. 2016. Embodied artificial agents for understanding human social cognition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371:** pii: 20150375.
99. Sanchez-Vives, M.V. & M. Slater. 2005. From presence to consciousness through virtual reality. *Nat. Rev. Neurosci.* **6:** 332–339.
100. Jung, Y. & K.M. Lee. 2004. Effects of physical embodiment on social presence of social robots. In *PRESENCE 2004*, Valencia, pp. 80–87.
101. Kompatsiari, K., V. Tikhanoff, F. Ciardo, *et al.* 2017. The importance of mutual gaze in human–robot interaction. In *Intelligent Virtual Agents.* W.-P. Brinkman, J. Broekens & D. Heylen, Eds.: 443–452. Cham: Springer International Publishing.
102. Kiesler, S. 2008. Anthropomorphic interactions with a robot and robot-like agent. *Soc. Cogn.* **26:** 169–181.
103. Li, J. 2015. The benefit of being physically present: a survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *Int. J. Hum. Comput. Stud.* **77:** 23–37.
104. Lazzeri, N., D. Mazzei, A. Greco, *et al.* 2015. Can a humanoid face be expressive? A psychophysiological investigation. *Front. Bioeng. Biotechnol.* **3:** 21.
105. Seo, S.H. 2015. Poor thing! Would you feel sorry for a simulated robot? A comparison of empathy toward a physical and a simulated robot. In *HRI' 15 Proceedings of the 10th Annual ACM/IEEE International Conference on Human–Robot Interaction.* ACM, Portland, OR, pp. 125–132.
106. Kwak, S.S., Y. Kim, E. Kim, *et al.* 2013. What makes people empathize with an emotional robot?: the impact of agency and physical embodiment on human empathy for a robot. In *2013 IEEE RO-MAN*, pp. 180–185.
107. Hofree, G., P. Ruvolo, M.S. Bartlett, *et al.* 2014. Bridging the mechanical and the human mind: spontaneous mimicry of a physically present android. *PLoS One* **9:** e99934.
108. Schilbach, L. 2015. Eye to eye, face to face and brain to brain: novel approaches to study the behavioral dynamics and neural mechanisms of social interactions. *Curr. Opin. Behav. Sci.* **3:** 130–135.
109. Cross, E.S., D.J.M. Kraemer, A.F. de C. Hamilton, *et al.* 2009. Sensitivity of the action observation network to physical and observational learning. *Cereb. Cortex* **19:** 315–326.
110. Schilbach, L., M. Wilms, S.B. Eickhoff, *et al.* 2010. Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. *J. Cogn. Neurosci.* **22:** 2702–2715.
111. Pfeiffer, U.J., L. Schilbach, B. Timmermans, *et al.* 2014. Why we interact: on the functional role of the striatum in the subjective experience of social interaction. *Neuroimage* **101:** 124–137.
112. Cracco, E., L. Bardi, C. Desmet, *et al.* 2018. Automatic imitation: a meta-analysis. *Psychol. Bull.* https://doi.org/10.1037/bul0000143.
113. Dautenhahn, K. 1998. The art of designing socially intelligent agents: science, fiction, and the human in the loop. *Appl. Artif. Intell.* **12:** 573–617.
114. Gregory, R.L. 1968. Perceptual illusions and brain models. *Proc. R. Soc. Lond. B Biol. Sci.* **171:** 279–296.
115. Summerfield, C., T. Egner, M. Greene, *et al.* 2006. Predictive codes for forthcoming perception in the frontal cortex. *Science* **314:** 1311–1314.
116. Smith, M.L., F. Gosselin & P.G. Schyns. 2012. Measuring internal representations from behavioral and brain data. *Curr. Biol.* **22:** 191–196.
117. Teufel, C., P.C. Fletcher & G. Davis. 2010. Seeing other minds: attributed mental states influence perception. *Trends Cogn. Sci.* **14:** 376–382.
118. Stanley, J., E. Gowen & R.C. Miall. 2007. Effects of agency on movement interference during observation of a moving dot stimulus. *J. Exp. Psychol. Hum. Percept. Perform.* **33:** 915–926.
119. Liepelt, R. & M. Brass. 2010. Top-down modulation of motor priming by belief about animacy. *Exp. Psychol.* **57:** 221–227.

120. Tsai, C.-C., W.-J. Kuo, D.L. Hung, *et al.* 2008. Action co-representation is tuned to other humans. *J. Cogn. Neurosci.* **20:** 2015–2024.

121. Stenzel, A., E. Chinellato, M.A.T. Bou, *et al.* 2012. When humanoid robots become human-like interaction partners: corepresentation of robotic actions. *J. Exp. Psychol. Hum. Percept. Perform.* **38:** 1073–1077.

122. Stenzel, A., T. Dolk, L.S. Colzato, *et al.* 2014. The joint Simon effect depends on perceived agency, but not intentionality, of the alternative action. *Front. Hum. Neurosci.* **8:** 595.

123. Longo, M.R. & B.I. Bertenthal. 2009. Attention modulates the specificity of automatic imitation to human actors. *Exp. Brain Res.* **192:** 739–744.

124. Gowen, E., E. Bolton & E. Poliakoff. 2016. Believe it or not: moving non-biological stimuli believed to have human origin can be represented as human movement. *Cognition* **146:** 431–438.

125. Wiese, E., A. Wykowska, J. Zwickel, *et al.* 2012. I see what you mean: how attentional selection is shaped by ascribing intentions to others. *PLoS One* **7:** e45391.

126. Wykowska, A., E. Wiese, A. Prosser, *et al.* 2014. Beliefs about the minds of others influence how we process sensory information. *PLoS One* **9:** e94339.

127. Shen, Q., H. Kose-Bagci, J. Sanders & K. Dautenhahn. 2011. The impact of participants' beliefs on motor interference and motor coordination in human–humanoid interactions. *IEEE Trans. Auton. Ment. Dev.* **3:** 6–16.

128. Cho, Y.S. & R.W. Proctor. 2003. Stimulus and response representations underlying orthogonal stimulus-response compatibility effects. *Psychon. Bull. Rev.* **10:** 45–73.

129. Stanley, J., E. Gowen & R.C. Miall. 2010. How instructions modify perception: an fMRI study investigating brain areas involved in attributing human agency. *Neuroimage* **52:** 389–400.

130. Klapper, A., R. Ramsey, D.H.J. Wigboldus, *et al.* 2014. The control of automatic imitation based on bottom-up and top-down cues to animacy: insights from brain and behavior. *J. Cogn. Neurosci.* **26:** 2503–2513.

131. Gallagher, H.L., A.I. Jack, A. Roepstorff, *et al.* 2002. Imaging the intentional stance in a competitive game. *Neuroimage* **16:** 814–821.

132. Krach, S., F. Hegel, B. Wrede, *et al.* 2008. Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS One* **3:** e2597.

133. Chaminade, T., D. Rosset, D. Da Fonseca, *et al.* 2012. How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. *Front. Hum. Neurosci.* **6:** 103.

134. Özdem, C., E. Wiese, A. Wykowska, *et al.* 2016. Believing androids—fMRI activation in the right temporo-parietal junction is modulated by ascribing intentions to non-human agents. *Soc. Neurosci.* **7:** 1–12.

135. McCabe, K., D. Houser, L. Ryan, *et al.* 2001. A functional imaging study of cooperation in two-person reciprocal exchange. *Proc. Natl. Acad. Sci. USA* **98:** 11832–11835.

136. Caruana, N., P. de Lissa & G. McArthur. 2017. Beliefs about human agency influence the neural processing of gaze during joint attention. *Soc. Neurosci.* **12:** 194–206.

137. Waytz, A., J.T. Cacioppo & N. Epley. 2010. Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspect. Psychol. Sci.* **5:** 219–232.

138. Waytz, A., N. Epley & J.T. Cacioppo. 2010. Social cognition unbound: insights into anthropomorphism and dehumanization. *Curr. Dir. Psychol. Sci.* **19:** 58–62.

139. Cullen, H., R. Kanai, B. Bahrami, *et al.* 2014. Individual differences in anthropomorphic attributions and human brain structure. *Soc. Cogn. Affect. Neurosci.* **9:** 1276–1280.

140. Chaminade, T., J. Hodgins & M. Kawato. 2007. Anthropomorphism influences perception of computer-animated characters' actions. *Soc. Cogn. Affect. Neurosci.* **2:** 206–216.

141. Waytz, A., C.K. Morewedge, N. Epley, *et al.* 2010. Making sense by making sentient: effectance motivation increases anthropomorphism. *J. Pers. Soc. Psychol.* **99:** 410–435.

142. Epley, N., A. Waytz & J.T. Cacioppo. 2007. On seeing human: a three-factor theory of anthropomorphism. *Psychol. Rev.* **114:** 864–886.

143. Gowen, E. & E. Poliakoff. 2012. How does visuomotor priming differ for biological and non-biological stimuli? A review of the evidence. *Psychol. Res.* **76:** 407–420.

144. Haslam, N., P. Bain, L. Douge, *et al.* 2005. More human than you: attributing humanness to self and others. *J. Pers. Soc. Psychol.* **89:** 937–950.

145. Loughnan, S. & N. Haslam. 2007. Animals and androids: implicit associations between social categories and nonhumans. *Psychol. Sci.* **18:** 116–121.

146. Harris, L.T. & S.T. Fiske. 2006. Dehumanizing the lowest of the low: neuroimaging responses to extreme out-groups. *Psychol. Sci.* **17:** 847–853.

147. Cikara, M., J.L. Eberhardt & S.T. Fiske. 2011. From agents to objects: sexist attitudes and neural responses to sexualized targets. *J. Cogn. Neurosci.* **23:** 540–551.

148. Harris, L.T., V.K. Lee, B.H. Capestany, *et al.* 2014. Assigning economic value to people results in dehumanization brain response. *J. Neurosci. Psychol. Econ.* **7:** 151–163.

149. Jack, A.I., A.J. Dawson & M.E. Norr. 2013. Seeing human: distinct and overlapping neural signatures associated with two forms of dehumanization. *Neuroimage* **79:** 313–328.

150. Eyssel, F.A. & D. Kuchenbrandt. 2011. My robot is more human than yours: effects of group membership on anthropomorphic judgments of social robots. In *Proceedings of the 24th IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011).*

151. Eyssel, F. & D. Kuchenbrandt. 2012. Social categorization of social robots: anthropomorphism as a function of robot group membership. *Br. J. Soc. Psychol.* **51:** 724–731.

152. Kuchenbrandt, D., F. Eyssel, S. Bobinger, *et al.* 2011. Minimal group–maximal effect? Evaluation and anthropomorphization of the humanoid robot NAO. In *ICSR 2011: Social Robotics.* pp. 104–113.

153. Nass, C., Y. Moon & N. Green. 1997. Are machines gender neutral? Gender-stereotypic responses to computers with voices. *J. Appl. Psychol.* **27:** 864–876.

154. Nass, C. 1996. Can computers be teammates? *Int. J. Hum. Comput. Stud.* **45:** 669–678.

155. Nass, C. & Y. Moon. 2000. Machines and mindlessness: social responses to computers. *J. Soc. Issues* **56:** 81–103.

156. Sacheli, L.M., A. Christensen, M.A. Giese, *et al.* 2015. Prejudiced interactions: implicit racial bias reduces predictive simulation during joint action with an out-group avatar. *Sci. Rep.* **5:** 8507.

157. Sacheli, L.M., M. Candidi, E.F. Pavone, *et al.* 2012. And yet they act together: interpersonal perception modulates visuo-motor interference and mutual adjustments during a joint-grasping task. *PLoS One* **7:** e50223.

158. Ferrari, F., M.P. Paladino & J. Jetten. 2016. Blurring human–machine distinctions: anthropomorphic appearance in social robots as a threat to human distinctiveness. *Int. J. Soc. Robot.* **8:** 287–302.

159. Waytz, A. & M.I. Norton. 2014. Botsourcing and outsourcing: robot, British, Chinese, and German workers are for thinking—not feeling—jobs. *Emotion* **14:** 434–444.

160. Kühn, S., T.R. Brick, B.C.N. Müller, *et al.* 2014. Is this car looking at you? How anthropomorphism predicts fusiform face area activation when seeing cars. *PLoS One* **9:** e113885–14.

161. Meltzoff, A.N. 2007. "Like me": a foundation for social cognition. *Dev. Sci.* **10:** 126–134.

162. Meltzoff, A. & W. Prinz. 2003. *The Imitative Mind: Development, Evolution, and Brain Bases.* Cambridge: Cambridge University Press.

163. Ganesh, S., H.T. van Schie, F.P. de Lange, *et al.* 2012. How the human brain goes virtual: distinct cortical regions of the person–processing network are involved in self-identification with virtual agents. *Cereb. Cortex* **22:** 1577–1585.

164. Riedl, R., P.N.C. Mohr, P.H. Kenning, *et al.* 2014. Trusting humans and avatars: a brain imaging study based on evolution theory. *J. Manag. Inf. Syst.* **30:** 83–114.

165. Schuwerk, T., M. Schurz, F. Müller, *et al.* 2017. The rTPJ's overarching cognitive function in networks for attention and theory of mind. *Soc. Cogn. Affect. Neurosci.* **12:** 157–168.

166. Carter, R.M. & S.A. Huettel. 2013. A nexus model of the temporal–parietal junction. *Trends Cogn. Sci.* **17:** 328–336.

167. Tanaka, F., A. Cicourel & J.R. Movellan. 2007. Socialization between toddlers and robots at an early childhood education center. *Proc. Natl. Acad. Sci. USA* **104:** 17954–17958.

168. Goodyear, K., R. Parasuraman, S. Chernyak, *et al.* 2016. Advice taking from humans and machines: an fMRI and effective connectivity study. *Front. Hum. Neurosci.* **10:** 181–185.

169. Goodyear, K., R. Parasuraman, S. Chernyak, *et al.* 2017. An fMRI and effective connectivity study investigating miss errors during advice utilization from human and machine agents. *Soc. Neurosci.* **12:** 570–581.

170. Vrtička, P. & P. Vuilleumier. 2012. Neuroscience of human social interactions and adult attachment style. *Front. Hum. Neurosci.* **6:** 212.

171. Hayama, S., L. Chang, K. Gumus, *et al.* 2016. Neural correlates for perception of companion animal photographs. *Neuropsychologia* **85:** 278–286.

172. Spunt, R.P., E. Ellsworth & R. Adolphs. 2017. The neural basis of understanding the expression of the emotions in man and animals. *Soc. Cogn. Affect. Neurosci.* **12:** 95–105.

173. Stoeckel, L.E., L.S. Palley, R.L. Gollub, *et al.* 2014. Patterns of brain activation when mothers view their own child and dog: an fMRI study. *PLoS One* **9:** e107205–e107212.

174. McLoughlin, N. & H. Over. 2017. Young children are more likely to spontaneously attribute mental states to members of their own group. *Psychol. Sci.* **28:** 1503–1509.

175. McLoughlin, N., S.P. Tipper & H. Over. 2017. Young children perceive less humanness in outgroup faces. *Dev. Sci.* **12:** e12539.

176. Yarkoni, T., R.A. Poldrack, T.E. Nichols, *et al.* 2011. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* **8:** 665–670.

177. Carter, R.M., D.L. Bowling, C. Reeck, *et al.* 2012. A distinct role of the temporal-parietal junction in predicting socially guided decisions. *Science* **337:** 109–111.