Cognitive, Affective, & Behavioral Neuroscience (2018) 18:837–856
https://doi.org/10.3758/s13415-018-0608-2

Seeing minds in others: Mind perception modulates low-level social-cognitive performance and relates to ventromedial prefrontal structures

Eva Wiese[1] · George A. Buzzell[1] · Abdulaziz Abubshait[1] · Paul J. Beatty[1]

Published online: 10 July 2018

Abstract

In social interactions, we rely on nonverbal cues like gaze direction to understand the behavior of others. How we react to these cues is affected by whether they are believed to originate from an entity with a mind, capable of having internal states (i.e., mind perception). While prior work has established a set of neural regions linked to social-cognitive processes like mind perception, the degree to which activation within this network relates to performance in subsequent social-cognitive tasks remains unclear. In the current study, participants performed a mind perception task (i.e., judging the likelihood that faces, varying in physical human-likeness, have internal states) while event-related fMRI was collected. Afterwards, participants performed a social attention task outside the scanner, during which they were cued by the gaze of the same faces that they previously judged within the mind perception task. Parametric analyses of the fMRI data revealed that activity within ventromedial prefrontal cortex (vmPFC) was related to both mind ratings inside the scanner and gaze-cueing performance outside the scanner. In addition, other social brain regions were related to gaze-cueing performance, including frontal areas like the left insula, dorsolateral prefrontal cortex, and inferior frontal gyrus, as well as temporal areas like the left temporo-parietal junction and bilateral temporal gyri. The findings suggest that functions subserved by the vmPFC are relevant to both mind perception and social attention, implicating a role of vmPFC in the top-down modulation of low-level social-cognitive processes.

Keywords Mind perception · gaze following · social interaction · fMRI · medial-frontal cortex · TPJ

Engaging in social interactions requires the ability to infer internal states of others, such as beliefs, intentions, and emotions (*mentalizing*; Baron-Cohen, 1997), and to use this information to predict their behavior (C. D. Frith & Frith, 2006). The primate brain is equipped with neural networks specialized in processing social information (*social brain*; Adolphs, 2009), responsible for making inferences about internal states and understanding the goals that underlie observed actions (Brothers, 2002; Bzdok et al., 2013; C. D. Frith & Frith, 2006; Van Overwalle, 2009; Van Overwalle & Baetens, 2009). Activation within social brain areas is modulated by the degree to which others are perceived as "having a mind"(Spunt, Meyer, & Lieberman, 2015) and the ability to experience internal states and execute goal-directed actions (*mind perception*; H. M. Gray, Gray, & Wegner, 2007). Mind perception is not exclusive to interactions with human agents; it can also be triggered in social interactions with nonhuman entities like animals or robots, as long as their behavior and/or appearance evoke associations with humanness (*anthropomorphism*; Abell, Happé, & Frith, 2000; Castelli, Happé, Frith, & Frith, 2000; DiSalvo, Gemperle, Forlizzi, & Kiesler, 2002; Kiesler, Powers, Fussell, & Torrey, 2008; Looser & Wheatley, 2010; Pfeiffer, Timmermans, Bente, Vogeley, & Schilbach, 2011; Waytz, Gray, Epley, & Wegner, 2010). Agents that do not trigger mind perception recruit social brain areas less than agents believed to have a mind (Gallagher, Jack, Roepstorff, & Frith, 2002; Harris & Fiske, 2006; Krach et al., 2008; Özdem et al., 2016; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003; Waytz, Gray, et al., 2010), and have a negative impact on performance during social interactions (Caruana, McArthur, Woolgar, & Brock, 2016; Wiese, Wykowska, Zwickel, & Müller, 2012; Wykowska, Wiese,

Eva Wiese and George A. Buzzell contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.3758/s13415-018-0608-2) contains supplementary material, which is available to authorized users.

✉ George A. Buzzell
Gbuzzell@gmu.edu

[1] Psychology Department, George Mason University, 4400 University Drive, Fairfax, VA 22030, USA

Prosser, & Müller, 2014). What has not been investigated so far is whether the degree to which mind perception activates social brain areas is directly related to human performance during social-cognitive tasks. To address this question, the current experiment employed parametric analyses of fMRI data to relate brain activation during a mind perception task (i.e., judging the likelihood that agents, varying in physical human-likeness, have internal states) with performance on a separate social attention task (i.e., attentional orienting to agents' gaze cues).

We expect networks that are activated during mind perception and social attention to be located in the social brain network, consisting of the action perception system (APS) involved in understanding the goals underlying observed actions, and the mentalizing system (MS) involved in inferring others' internal states (Adolphs, 2009). The APS consists of a distributed network of temporal areas like the extrastriate body area (EBA) and posterior superior temporal sulcus (pSTS), as well as parietal and frontal areas like the inferior parietal cortex (IPC) and ventral premotor cortex (vPMC); the temporal areas are thought to detect the presence of intentional agents and label their actions as goal-directed based on observed motion patterns, while the parietal and frontal areas are believed to identify particular goals underlying these actions (e.g., "What is the outcome of an action?"; Becchio, Adenzato, & Bara, 2006; Grafton & Hamilton, 2007; Pobric & Hamilton, 2006; Saxe, 2006; Saygin, 2007; Saygin, Wilson, Hagler, Bates, & Sereno, 2004). Action understanding in the primate brain is based on the principles of resonance, where shared representations are activated both when an action is executed and when a similar action is observed in others (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). In nonhuman primates, resonance is associated with mirror neurons located in the IPC and vPMC, which fire during both action observation and execution, and may support inferences about the goals underlying observed actions of others (Gallese et al., 1996; Gallese, Keysers, & Rizzolatti, 2004; Iacoboni, 2005; Keysers & Perrett, 2004; Rizzolatti & Craighero, 2004). Although there is agreement that action understanding in humans is also based on the principles of resonance (Kilner, Paulignan, & Blakemore, 2003; Oztop, Franklin, Chaminade, & Cheng, 2005; Press, Bird, Flach, & Heyes, 2005; Rizzolatti & Craighero, 2004; Umiltà et al., 2001), the particular role of mirror neurons in this process is still a matter of debate (Chong, Cunnington, Williams, Kanwisher, & Mattingley, 2008; Dinstein, Hasson, Rubin, & Heeger, 2007; Kilner, Neal, Weiskopf, Friston, & Frith, 2009; Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010; Saygin, Chaminade, Ishiguro, Driver, & Frith, 2012).

The mentalizing system is a distributed network involving posterior areas like the temporo-parietal junction (TPJ), superior temporal sulcus (STS), and fusiform gyrus (FG), as well as anterior areas like the medial and ventromedial prefrontal cortex (mPFC, vmPFC), and anterior cingulate cortex (ACC; Saygin et al., 2012; Van Overwalle, 2009). Within the posterior part of the network, the STS is involved in processing biological motion and inferring intentions underlying biological cues, like changes in gaze or head direction, while the FG is responsible for encoding invariable facial information, such as identity (Nummenmaa & Calder, 2009). The TPJ is involved in inferring particular intentions, beliefs and higher-order action goals in a situation-specific manner ("Why is an observed action executed?"; Chaminade & Decety, 2002; Farrer et al., 2003; Gallagher et al., 2000; Grèzes, Berthoz, & Passingham, 2006; Grèzes, Frith, & Passingham, 2004; Ohnishi et al., 2004; Perner, Aichhorn, Kronbichler, Staffen, & Ladurner, 2006; Ruby & Decety, 2001; Saxe & Kanwisher, 2003; Saxe & Powell, 2006), and allows differentiating self from other intentions via perspective taking (Chaminade & Decety, 2002; Farrer et al., 2003; Ruby & Decety, 2001). Although still a matter of debate, social functions seem to be lateralized within TPJ, with lTPJ being more involved in perspective taking (Samson, Apperly, Chiavarino, & Humphreys, 2004) and anthropomorphism (Chaminade, Hodgins, & Kawato, 2007; Cullen, Kanai, Bahrami, & Rees, 2013; Perner et al., 2006; Zink et al., 2011), and rTPJ being more responsible for discriminating intentional from nonintentional actions (Cavanna & Trimble, 2006; Chaminade et al., 2012; Gallagher et al., 2002; Krach et al., 2008) and reasoning about others' particular internal states (Costa, Torriero, Oliveri, & Caltagirone, 2008; Gallagher et al., 2000; Saxe, 2006; Saxe & Kanwisher, 2003). The rTPJ also serves as convergence point for social and nonsocial processes (Chang et al., 2013; Krall et al., 2015; Krall et al., 2016; Mitchell, 2008; Scholz, Triantafyllou, Whitfield-Gabrieli, Brown, & Saxe, 2009), and is involved in processing language and semantics (Binder, Desai, Graves, & Conant, 2009). The anterior part of the MS includes areas like the mPFC, vmPFC, and ACC, and is involved in making inferences about others based on enduring dispositions, such as traits or preferences rather than inferring particular internal states on a trial-by-trial basis (Amodio & Frith, 2006; Brothers, 2002; Saxe, 2006; Saxe & Kanwisher, 2003; Saygin et al., 2012; Van Overwalle, 2009). This requires neurons with the ability to represent behavior over a longer period of time, across different circumstances, and with different social partners, a feature that applies to neurons in the mPFC (Amodio & Frith, 2006; Decety & Chaminade, 2003; U. Frith & Frith, 2001; Gallagher & Frith, 2003; Huey, Krueger, & Grafman, 2006; Leslie, Friedman, & German, 2004; Wood & Grafman, 2003). Activation within mPFC is positively correlated to the degree of background knowledge one possesses about another person (Saxe & Wexler, 2005), as well as to the social relevance ascribed to information about others (Grèzes et al., 2004). In contrast, activity within vmPFC has been associated with reasoning about the emotional states of others (Hynes, Baird, & Grafton, 2006; Völlm et al., 2006). Medial prefrontal areas are also involved in impression formation by providing access to

general social knowledge (Mitchell, Macrae, & Banaji, 2006; Szczepanski & Knight, 2014), with activity in mPFC being linked to retrieving stereotypical knowledge about people (Contreras, Banaji, & Mitchell, 2012; Fairhall, Anzellotti, Ubaldi, & Caramazza, 2014), and activity in vmPFC being related to retrieving script-based social knowledge (Ghosh, Moscovitch, Melo Colella, & Gilboa, 2014; van Kesteren, Ruiter, Fernández, & Henson, 2012). Medial prefrontal areas are also related to egocentric mentalizing about similar others (Jenkins, Macrae, & Mitchell, 2008; Mitchell, Macrae, & Banaji, 2004, 2006), and are activated when viewing social scenes containing human versus nonhuman agents (Wagner, Kelley, & Heatherton, 2011). The ACC is specifically activated during interactions that require mentalizing in real time (Gallagher et al., 2000; McCabe, Houser, Ryan, Smith, & Trouard, 2001) and has been suggested as a neural correlate of mind perception, as human-like agents capable of executing intentional actions activate this brain area more strongly than nonhuman agents (Gallagher et al., 2002).

Previous research has shown that during interactions with others, activity within social brain areas is modulated by the degree to which others are perceived as "having a mind," with stronger activation for agents believed to have a mind than for those who do not (Gallagher et al., 2002; Krach et al., 2008; Sanfey et al., 2003). For instance, observing the actions of nonhuman agents recruits the APS to a smaller degree than observation of human actions (Kilner et al., 2003; Oberman, McCleery, Ramachandran, & Pineda, 2007; Oztop et al., 2005; Press et al., 2005); the actual degree of activation has been shown to depend on features like physical appearance (Chaminade et al., 2007; Kupferberg et al., 2012), motion kinematics (Bisio et al., 2014), and familiarity (Press, Gillmeister, & Heyes, 2007). Agents failing to trigger mind perception also underactivate the MS, with reduced activation for nonhuman versus human agents, as well as agents who are deprived of their ability of "having a mind" due to dehumanization (Gallagher et al., 2002; Harris & Fiske, 2006; Krach et al., 2008; Özdem et al., 2016; Sanfey et al., 2003; Spunt et al., 2015; Waytz, Morewedge, et al., 2010a; Wykowska et al., 2014).

In addition to activation in social brain areas, mind perception also modulates performance and attitudes during social interactions. For example, mind perception has been shown to influence prosocial behaviors (Bering & Johnson, 2005; Epley, Waytz, Akalis, & Cacioppo, 2008; Graham & Haidt, 2010; Gray, Young, & Waytz, 2012; Shariff & Norenzayan, 2007), reactions to observing negative consequences for others (Cushman, 2008; Gray & Wegner, 2008; Ohtsubo, 2007), and the motivation to perpetuate moral standards (Haley & Fessler, 2005). Similarly, attitudes and performance in interactions with nonhuman agents can be improved when the agents trigger mind perception by displaying human features or behaviors (Bennewitz, Faber, Joho, Schreiber, & Behnke, 2005; Fussell, Kiesler, Setlock, & Yew, 2008; Huang & Thomaz, 2011; Mutlu, Forlizzi, & Hodgins, 2006; Mutlu, Kanda, Forlizzi, Hodgins, & Ishiguro, 2012; Pfeiffer-Leßmann, Pfeiffer, & Wachsmuth, 2018; Sidner, Kidd, Lee, & Lesh, 2004; Staudte & Crocker, 2011; Wiese, Metta, & Wykowska, 2017; Yamazaki, Yamazaki, Burdelski, Kuno, & Fukushima, 2010). In contrast, agents not triggering mind perception negatively impact performance in social interactions (Caruana et al., 2016; Wiese et al., 2012; Wykowska et al., 2014) and fail to induce social facilitation (Bartneck, 2003; Park & Catrambone, 2007; Riether, Hegel, Wrede, & Horstmann, 2012; Woods, Dautenhahn, & Kaouri, 2005). Specifically, it has been shown that social signals, like changes in gaze direction, are followed to a larger extent when they are believed to reflect the actions of a mind compared to a preprogrammed algorithm (Caruana et al., 2016; Wiese et al., 2012; Wykowska et al., 2014), with faster responses to targets presented at gazed-at locations (*gaze-cueing effect*; Friesen & Kingstone, 1998).

Aim of study

Prior research indicates that mind perception has the capacity to modulate activation in social brain areas, as well as performance during social-cognitive tasks. However, relations between activation in brain areas related to mind perception and performance during social-cognitive tasks have yet to be established. That is, prior studies have not tested whether within-subject variation in brain activation during mind perception is related to subsequent variation in performance on social-cognitive tasks and, if so, which brain areas are most closely related to social-cognitive performance. We address this question by relating brain activation during a mind perception task (i.e., judging the likelihood that agents have internal states; Martini, Gonzalez, & Wiese, 2016) to performance on a low-level social-cognitive task (i.e., attentional orienting to gaze cues; Friesen & Kingstone, 1998). These tasks were chosen based on previous studies showing that (a) judgments regarding others' capacity of having internal states require mind perception (Cheetham, Suter, & Jancke, 2014; Hackel, Looser, & Van Bavel, 2014; Looser & Wheatley, 2010; Martini et al., 2016; Waytz, Gray, et al., 2010), and (b) the degree to which others' gaze is followed is linked to mind perception and other more complex social-cognitive processes like mentalizing (Baron-Cohen, Leslie, & Frith, 1985). In both tasks, we used a set of images that varied in their degree of physical humanness and were created by morphing separate images of a human and a robot face into each other in steps of 20%. Manipulating physical humanness via morphing has been used in previous studies as a reliable tool to manipulate the degree to which mind is perceived in others (Cheetham et al., 2014; Hackel et al., 2014; Looser &

Wheatley, 2010; Martini et al., 2016; Waytz, Gray, et al., 2010). The mind perception task was performed inside an fMRI scanner to determine the degree to which reasoning about the agents' capability of having internal states elicited activation within the social brain network; the social attention task was performed outside the scanner, and reaction times were collected in order to assess the degree to which the agents' gaze triggered shifts of spatial attention to gazed-at locations.

We first confirmed that the mind perception task activated the social brain network by employing a parametric analysis of the fMRI data utilizing the mind perception ratings as weights. As a second step, to test whether activation in the social brain network was also related to subsequent performance on a social attention task, a parametric analysis of the fMRI data was performed utilizing each participant's variation in gaze cueing across the different levels of physical humanness as weights. Together, these two parametric analyses of the fMRI data provide insight about the neural regions involved in mind perception and relations with subsequent low-level social-cognitive performance, respectively. Of particular interest was whether any neural regions were activated not only during the mind perception task but also were related to subsequent low-level social-cognitive performance during gaze cueing. An overlap in activity between these two analyses would provide evidence that initial neural activity related to mind perception, for a particular agent, is related to subsequent low-level social-cognitive performance involving that same agent. In line with the notion that mind perception is a prerequisite for low-level social-cognitive processes like social attention, we predicted that overlapping fMRI activation would be identified within the social brain network.

Method and materials

Participants

Twenty-two undergraduate students (seven female, mean age = 24.36, *SD* = 4.73) were recruited from George Mason University and paid $15 per hour for their participation. All were right-handed, had normal or corrected-to-normal vision, had no known neurological deficits, and were not currently taking any medications known to affect the central nervous system. The office of integrity and assurance approved all procedures, and participants provided informed consent prior to the experiment.

Stimuli

Six agent images were created that varied in their degree of physical humanness (in %) from machine-like (100% robot) to human-like (100% human) and were used both for the mind perception task and the social attention task.[1] Changing the physical appearance of an agent in a parametric fashion has been shown to modulate the degree to which mind is attributed to an agent in previous studies (Hackel et al., 2014; Martini et al., 2016) and to alter activation within social brain areas (e.g., Gao, McCarthy, & Scholl, 2010; Looser & Wheatley, 2010; Waytz, Morewedge, et al., 2010; Wheatley, Weinberg, Looser, Moran, & Hajcak, 2011).

The stimuli were created using FantaMorph, which allows two images to be blended together at specified increments (in %). The images used to create the stimuli were the Meka S2 humanoid robot head and a male human face (Lundqvist, Flykt, & Öhman, 1998). Morphing occurred at 20% increments, yielding a total of six images (0%, 20%, 40%, 60%, 80%, 100% physical humanness; see Fig. 1). Each image was presented on white background in full frontal orientation and subtended 7.8° wide and 8.6° high. For both the mind perception and the social attention tasks, the eyes were centered on the horizontal axis of the screen. In the mind perception task, the pupils always remained centered relative to the vertical axis of the screen, looking straight ahead; in the social attention task, irises and pupils were additionally shifted with Photoshop to deviate 0.4° from direct gaze in order to create the impression of an eye movement.

Tasks

Mind perception task

The mind perception task was performed inside of an fMRI scanner and involved making judgments about the capability of different agents (varying in their degree of physical humanness) of having internal states (see Martini et al., 2016). The sequence of events on a given trial is shown in Fig. 2. Each trial began with the presentation of a question (see Supplementary Table S1) regarding an internal state (e.g., "How likely is it that this agent has a mind?"), followed by a series of images depicting the different morphed images in a randomized order. As each agent image was presented, participants were instructed to rate the agent on the particular question that had just been presented using a Likert scale from 1 (*very unlikely*) to 8 (*very likely*). Responses were entered using a pair of fMRI-safe button boxes. Each internal state question was presented for 5 seconds, followed by a screen that contained only a fixation cross for a jittered time period of 12 to 16 seconds. During the sequence of agent

[1] Physical humanness refers to the percentage amount of the human image that is contained in the morphed image. For instance, a 60% human morph contains 40% features of the robot image and 60% features of the human image. Please note that although physical humanness is manipulated parametrically, perceived humanness as measured in mind ratings, follows a qualitative pattern (in line with Hackel et al., 2014; Looser & Wheatley, 2010; Martini et al., 2016; see Results section).

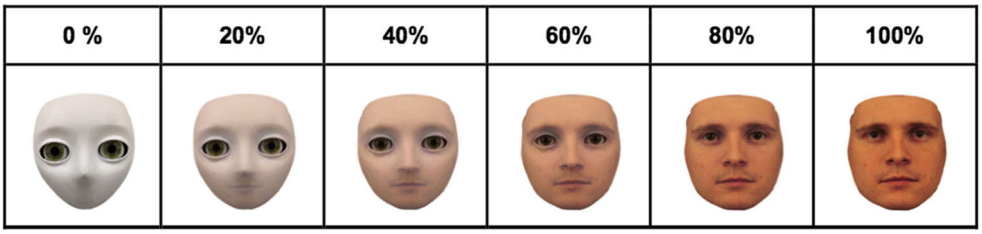

Fig. 1 Stimuli used for the mind perception and social attention tasks. The images were created by morphing a robot face (Meka robot; image on the very left) into a human face (adult male; image on the very right) in steps of 20%

images, each agent was presented for 2 seconds, and participants were given an additional 4 seconds on average (jittered between 2 and 6 seconds) to give a response (i.e., the minimum amount of time for a response was 4 seconds). The mind perception task was divided into four blocks, each consisting of 12 questions and an average of 72 agent presentations (six agents times 12 questions) and lasting approximately 12 minutes each (total task time = 48 minutes). During the task, each of the six distinct agent images (i.e., 0%, to 100% physical humanness in steps of 20%) was presented 72 times, while each of the 24 questions was presented twice (see Supplementary Table S1).

Social attention task

A gaze-cueing paradigm (Friesen & Kingstone, 1998) was used to measure low-level social-cognitive performance in the current experiment. This task was chosen for two reasons: (1) being able to attend to where others are looking is a prerequisite for mentalizing and other more complex social-cognitive functions and is thus a good proxy for social-cognitive performance (Frischen, Bayliss, & Tipper, 2007; for a review), and (2) the degree to which a mind is perceived in others has been shown to modulate mechanisms of social attention like gaze cueing in previous studies (Teufel et al., 2009; Wiese et al., 2012; Wykowska et al., 2014). In contrast to the mind perception task, the gaze-cueing task was performed outside the fMRI scanner and required participants to respond to the identity of a target letter (*F* or *T*) while reaction times were measured. The target either appeared at the location that was looked at by the agent (i.e., valid trial) or opposite of where the face was looking (i.e., invalid trial). Gaze-cueing effects were calculated by subtracting reaction times for valid trials from reaction times for invalid trials (i.e., difference score). In the current experiment, we used a reversed gaze-cueing task, where targets appeared with a higher likelihood opposite of where the agent was looking (80% of the cases) compared with locations that were looked

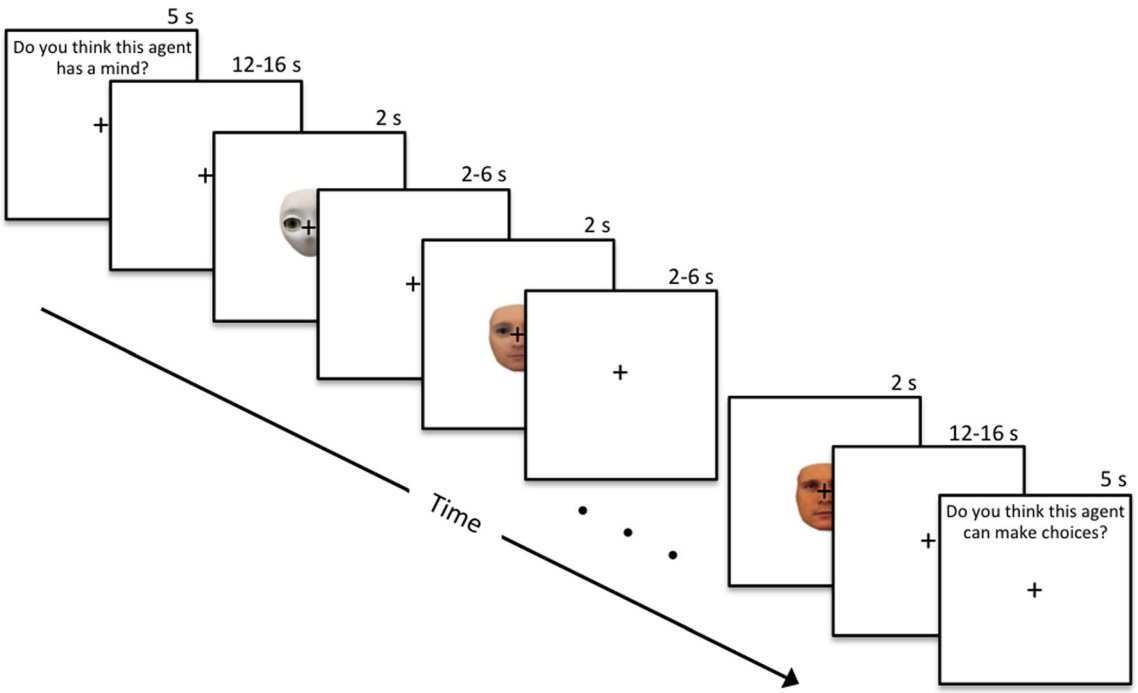

Fig. 2 Mind perception task: While inside of an fMRI scanner, participants were presented with theory-of-mind questions and a series of morphed images to judge. Participants rated each image on a 1–8 scale

at by the agent (20% of the cases; see Friesen, Ristic, & Kingstone, 2004). This was done in order to distinguish between bottom-up components of gaze cueing, which are apparent if participants attend to the gazed-at location despite the target being more likely to appear at the uncued location (i.e., shorter reaction times at the cued location), and top-down influences on gaze cueing, which would be apparent if participants orient away from the gaze cue and shift their attention to the uncued location, which is more likely to contain the target (i.e., shorter reaction times at the uncued location).

The sequence of events on a given trial is shown in Fig. 3. At the beginning of each trial, a black fixation cross appeared on white screen for a jittered time interval of 700 to 1,000 ms, followed by the image of one of the agents displaying a straight gaze. After a jittered time interval of 700 to 1,000 ms, the agent changed gaze direction and looked either to the left or right side of the screen for 400 to 600 ms, followed by the presentation of the target letter (*F* or *T*, measuring .5° in width and .9° in height) that either occurred where the face was looking or opposite of where the face was looking. Targets appeared on the horizontal axis of the screen and were located 14.7° from the center of the screen. The image of the agent and the target remained on the screen until the participant gave a response or a time-out criterion was reached (1,200 ms after target presentation), whichever came first. The intertrial interval (ITI) was 680 ms.

Participants used the index finger of each hand to respond to the identity of the target letter by pressing either the key that was marked with "F" or "T." For half of the participants, "F" was assigned to the "D" key, and "T" was assigned to the "K" key of a regular keyboard, with reversed key assignment for the other half of the participants; key labels were counterbalanced across participants throughout the study. Participants were instructed to maintain fixation on the center of the screen throughout all trials and to respond as quickly and accurately as possible to the target letters. Before the actual experiment started, participants first completed a practice block that mirrored the experimental task but used a different agent stimulus (EDDIE; developed at Technische Universitaet Muenchen; see Wiese et al., 2012) to avoid priming effects or other response biases. Participants then performed six experimental blocks, with each block employing one of the six agent images; the order in which agents were presented was counterbalanced across participants. Total time for the social attention task was approximately 20 minutes.

Procedure

The experiment started with the mind perception task in the scanner, followed by the social attention task and a series of questionnaires outside the scanner. Participants were screened for fMRI safety and completed a demographic questionnaire approximately 1 week prior to participating in the experiment. When participants arrived on the day of the experiment, they were first provided with the instructions of the mind perception task and then positioned in the fMRI scanner in order to perform the task. Following the mind perception task, participants exited the scanner and were then provided with the instructions for a social attention task, which took place in a separate room. Critically, the same agents were employed for both the mind perception and the social attention task. After completion of the social attention task, participants filled out questionnaires and were debriefed.

Analyses

Behavioral data

The behavioral data of the mind perception and social attention tasks were analyzed using the LME4 and the Mediation packages in R (version 3.2.4). We first tested if the relationship of the three variables (physical humanness, gaze-cueing

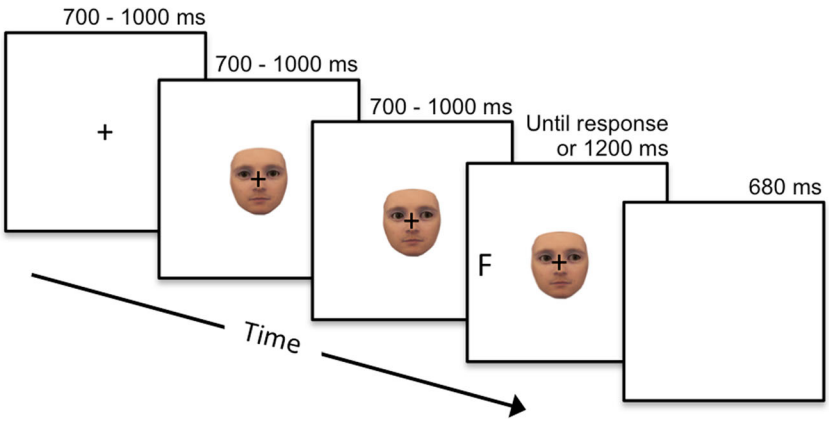

Fig. 3 Social attention task: Outside of the fMRI scanner, participants performed the gaze cueing task using the same morphed images from the mind perception task. Participants were required to identify the identity of a target letter, presented in the periphery that was preceded by either a congruent or incongruent looking morph image

behavior, and mind ratings) were linear or nonlinear. To do so, we constructed four mixed-effects regression models (i.e., one linear and three polynomials: quadratic, cubic, and fourth level) to model the data. This step was done for each of our three predictive relationships. In other words, we tested whether the mind ratings could be predicted by physical humanness in a linear or nonlinear method, if gaze-cueing behavior could be predicted by physical humanness, in a linear or nonlinear fashion, and if gaze-cueing behavior could be predicted by mind ratings in a linear or nonlinear way (see Fig. 4). After we examined the linear and nonlinear relationships for all pairs of our three variables (mind ratings and physical humanness, gaze-cueing behavior and physical humanness, gaze-cueing behavior and mind ratings), we compared the linear model to the nonlinear models in a nested model comparison to determine which of the models represented the data best. Choosing the model of best fit was decided based on a chi-square test that compares more complex, polynomial models (i.e., quadratic, cubic, and fourth level) to a linear reference model (i.e., the simplest model fit; all models with a chi-square test result of $p < .05$ differ significantly from the linear model in terms of model fit). Moreover, the model with the smallest Bayesian information criterion (BIC) constitutes the best, and at the same time most parsimonious, model fit for a given data set (Konishi & Kitagawa, 2008). This step was repeated for each pair of relationships (mind ratings and physical humanness, gaze-cueing behavior and physical humanness, and gaze-cueing behavior and mind ratings).

After testing for which of the relationships were linear and which were nonlinear, we investigated a mediation model that predicted gaze-cueing behavior from physical humanness through mind ratings as a mediator. To avoid overfitting the model and to aid interpretation, we specified a more simplistic mediation model by allowing only linear relations for the mediation analysis, regardless of how well the polynomials predicted the outcome of the nested model comparisons described above.

$$(A)\ y = b_1 x$$

$$(B)\ y = b_1 x + b_2 x^2$$

$$(C)\ y = b_1 x + b_2 x^2 + b_3 x^3$$

$$(D)\ y = b_1 x + b_2 x^2 + b_3 x^3 + b_4 x^4$$

Fig. 4 Equations tested in the nested model comparison. Data were modeled using a linear model (**a**), as well as a quadratic (**b**), cubic (**c**), and fourth-level (**d**) polynomial. The error terms and intercept have been omitted in all of the equations

fMRI data

Image acquisition and preprocessing We acquired fMRI data using a Siemens Allegra 3T scanner, equipped with a standard one-channel quadrature birdcage head coil. During each run, T2* gradient-echo, echo-planar imaging was acquired, with a TR/TE of 2300/30 ms, flip angle = 90 degrees, 40 interleaved axial slices 3 mm thick/1 mm gap, FOV = 192 mm, and matrix size = 64 × 64 (in-plane resolution of 3 mm^2). Following fMRI acquisition, a whole-head, T1 structural scan was acquired using a three-dimensional, magnetization-prepared, rapid-acquisition gradient echo (MPRAGE) pulse sequence. During the MPRAGE sequence, 160 1-mm-thick slices (256 × 256 matrix, field of view = 260, .94 mm voxels) were acquired with a TR/TE of 2300/3 ms.

All analyses of fMRI data were performed using FSL (www.fmrib.ox.ac.uk/fsl). In order to allow the scanner to reach equilibrium magnetization, the first five volumes were removed prior to analysis. The fMRI data were high-pass filtered (128-s cutoff), slice timing corrected (Hanning-windowed sinc interpolation to shift each time series relative to the middle of the TR period), and motion corrected using FMRIB's Linear Registration Tool (MCFLIRT). Prewhitening using fMRIB's Improved Linear Model (FILM) was performed to remove temporal autocorrelation in the fMRI time-series data. Data were smoothed using a 6-mm full-width at half-maximum (FWHM) Gaussian kernel. Coregistration was completed in a two-step process. Functional data were first registered to a high-resolution structural image (MPRAGE) using FMRIB's Linear Registration Tool (FLIRT) following brain extraction using the Brain Extraction Tool (BET) with the fractional intensity threshold set to .35. Registration to standard space (T1 2-mm MNI template) was then performed using FLIRT.

Neural activation associated with mind perception The first analysis of the fMRI data sought to identify whether the mind perception task reliably activated regions within the social brain network. To this end, a parametric analysis of the fMRI data was carried out, with a parametric regressor being used to identify neural regions that tracked trial-by-trial variation in mind perception. The initial, a priori analysis of the data employed all four blocks (separate runs of fMRI acquisition). However, while results from this initial analysis yielded a cluster of activation within vmPFC (see Fig. 9), no activations survived a whole-brain correction for multiple comparisons. Due to concerns that the lack of statistical robustness for this initial analysis was the result of habituation and repeated exposure to the same agent images over extended periods of time, we performed a second, post hoc, analysis of these data in which only the first two blocks (separate runs of fMRI acquisition) were employed. This second analysis was performed in an effort

to optimize the likelihood of identifying statistically robust neural regions that tracked trial-by-trial variation in mind perception; indeed, as described within the Results section, this post hoc analysis revealed a qualitatively similar cluster of activation within the vmPFC that survived correction for multiple comparisons.

For both the a priori and post hoc analyses of the fMRI data, a parametric regressor modeled the onset of each agent at a magnitude determined by the mean-centered Likert-scale rating provided on each trial, whereas a second task-related regressor modeled the onset of each agent image at a fixed magnitude. A nuisance regressor was also included to model the onset of each question, using a fixed magnitude. All task-related regressors were convolved with a canonical double-gamma hemodynamic response function (HRF) with no phase delay. Six motion parameters (three translation, three rotation) were also added to the GLM model as confound regressors in order to account for residual motion effects after correction by MCFLIRT (nine regressors total). A second-level analysis was used to average across the first two runs for each participant using a fixed-effects model. Data were then averaged across participants in a third-level analysis, using FMRIB's Local Analysis of Mixed Effects (FLAME1). We then conducted a whole-brain analysis investigating the parametric effect of gaze cueing. The family-wise error rate (FWER) were controlled for at an alpha level of .05, using cluster-based correction following Gaussian random field (GRF) theory and a cluster-defining threshold of $Z = 1.96$.

Association of neural activation during mind perception and social attention Following the identification of neural regions associated with mind perception, we sought to identify how brain activity elicited by the mind perception task was related to the degree of gaze cueing during the social attention task. To this end, we again performed a parametric analysis; however, in this second analysis, the parametric regressor was modulated based on average gaze-cueing effect values that an individual exhibited for each agent during the social attention task. Therefore, the second parametric analysis allowed us to identify which brain regions that were activated during the mind perception task were directly related to performance in a separate low-level social-cognitive task. Moreover, we identified neural regions that were significantly activated by both the parametric analysis based on gaze cuing and the parametric analysis based on mind ratings. In line with the a priori analysis of neural activity associated with the mind perception task, all four blocks of fMRI data were also employed for the analysis of relations between neural activity during mind perception and behavior during the social attention task. All other aspects of this second fMRI analysis were identical to those described for the analysis focusing on mind perception (see above).

Results

Behavioral data

Mind perception task Results of the nested model comparison predicting mind ratings from physical humanness revealed that both the cubic model, $\chi^2(2) = 11.04$, $p = .003$, BIC = 430.05, and the fourth-level polynomial model, $\chi^2(1) = 12.54$, $p < .001$, BIC = 422.39, fit the data significantly better than the linear model; the fourth-level polynomial model constitutes the overall model of best fit based on the BIC estimate (i.e., smallest BIC; see Table 1 and Fig. 5). These results suggest that linear changes in physical human-likeness do not lead to linear changes in ratings of mind perception; in contrast, linear increases in human-likeliness where associated, on average, with a nonlinear (fourth-level polynomial) increase in ratings of mind perception.

Social attention task The nested model comparison of models predicting gaze-cueing behavior from physical humanness showed that only the cubic model fit significantly different better than the linear model, $\chi^2(2) = 4.51$, $p = .03$; however, the cubic model was not the most parsimonious model based on the BIC (i.e., BIC for the cubic model was larger than the BIC for the linear model; see Table 2). Thus, the linear model constitutes the overall best model fit for the gaze-cueing data (see Fig. 6). This result suggests that linear increases in physical human likeness lead to linear increases in gaze cueing. That is, although gaze cues invalidly cued the target location on 80% of trials, increases in physical humanness led to increased reflexive attentional orienting in direction of the gaze cue (and slower response times at the uncued location).

Link between physical humanness, mind perception, and social attention Before examining the nested model comparison of models predicting gaze-cueing behavior from mind ratings, we controlled for the agents' physical humanness by adding it as a covariate in the model. After controlling for physical humanness, the nested model comparison of models predicting gaze-cueing behavior showed that none the polynomial models fit significantly better than the linear model, as indicated by the chi-square test (see Table 3). This indicates that the linear model is the best fit for the relationship between

Table 1 Nested model comparison predicting mind rating data from physical humanness

	BIC	χ^2	*p* value
Linear model	431.33		
Quadratic model	433.86	2.35	.12
Cubic model	430.05	8.68	<.01
Fourth-level polynomial	422.39	12.54	<.001

gaze-cueing behavior and mind ratings (after controlling for physical humanness). This finding illustrates that reflexive orienting to gaze cues decreases as mind ratings increase, but that voluntary attentional orienting to predicted target locations increases (see Fig. 7). In other words, after controlling for physical humanness, increases in mind ratings were associated with reductions of the "bottom up" component and enhancements of the "top down" component of gaze cueing, which led to a greater reliance on the predictivity of the gaze cue and faster response times to targets appearing at the uncued location. These data are consistent with the notion that increased mind ratings lead to a greater reliance on higher-level behavioral attributes of the agent (i.e., the predictivity of its gaze direction).

After investigating the models of best fit for all three relationships (mind ratings and physical humanness, gaze-cueing behavior and physical humanness, and gaze-cueing behavior and mind ratings), we tested whether mind ratings partially mediated the relationship between physical humanness and gaze-cueing behavior, despite their having effects of opposite directions on gaze-cueing performance. As indicated before, we only used linear models to avoid overfitting the model with too many parameters, as well as to simplify the interpretation. The mediation analysis revealed a nonsignificant total effect of ($\beta = .15$, 95% CI [−.01, .33], $p = .08$), a significant positive direct effect ($\beta = .44$, 95% CI [.14, .73], $p < .01$) of physical humanness on gaze-cueing behavior, as well as a significant negative indirect effect of mind perception on gaze cueing ($\beta = -.28$, 95% CI [−.52, −.05], $p < .01$; see Fig. 8). Since physical humanness and mind ratings were highly correlated ($r = .83$), the observed negative indirect relationship between mind perception and gaze-cueing behavior needs to be interpreted with caution, as it could be an artifact due to issues with multicollinearity (Cohen, Cohen, West, & Aiken, 2003); this would mean that we could not be certain of the direction of this effect as the sign (positive or negative) of the weight, could flip. However, since multicollinearity, if anything, decreases the power of detecting an effect and thus decreases the

Table 2 Nested model comparison predicting gaze cueing data from physical humanness

	BIC	χ^2	*p* value
Linear model	−258.63		
Quadratic model	−253.75	.01	.93
Cubic model	−253.39	2.51	.03
Fourth-level polynomial	−248.53	.02	.87

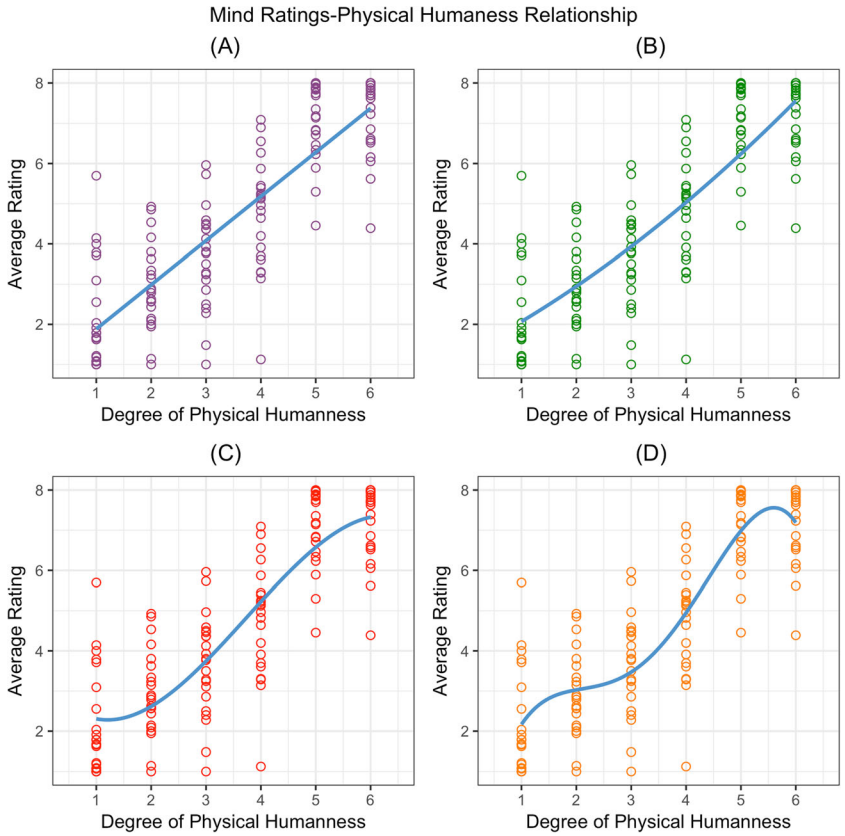

Fig. 5 Average mind ratings as a function of physical humanness (1 = 0% physical humanness; 6 = 100% physical humanness, as modeled by a linear (**a**), quadratic (**b**), cubic (**c**), and fourth-level polynomial (**d**) model. The fourth-level polynomial model constituted the overall best model fit (see Table 1)

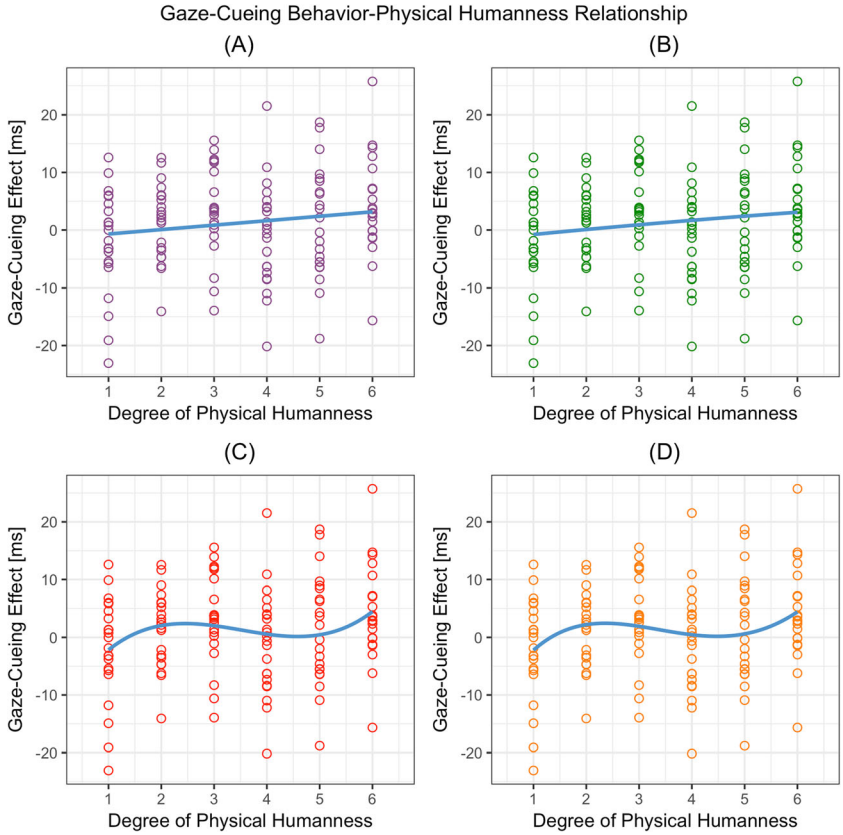

Fig. 6 Average gaze-cueing effects as a function of the degree of physical of humanness. For the *x*-axis, 1 = 0% human, 6 = 100% human, as modeled by a linear (**a**), quadratic (**b**), cubic (**c**), and fourth-level polynomial (**d**) model. The linear model constituted the overall best model fit

probability of rejecting the null hypothesis (Cohen et al., 2003), it is unlikely that the negative direction of the indirect effect is a mere artifact of multicollinearity, as the indirect effect of mind perception on gaze cueing is statistically significant despite such multicollinearity. What is more likely is that the mediation model is showing a suppression phenomenon; unlike in consistent mediation models (i.e., models that have the same direction for all of their paths), suppression occurs when two variables that are related to each other (i.e., an independent variable and a mediator) cause the dependent variable to move in opposite directions (Mackinnon, Krull, & Lockwood, 2000). Consistent with suppression, we find that adding the mediator (i.e., mind ratings) increases the strength of the relationship between physical humanness and gaze-cueing behavior (β increased from .15 to .44 after including the mediator). Taken together, the data suggest that physical humanness affects social attention performance in two potentially opposing ways: On the one hand, increases in physical humanness seem to enhance reflexive attentional orienting to gazed-at locations (i.e., increases in gaze-cueing effects) despite the fact that the predictivity of the gaze cue is low (i.e., 20%), suggesting that changes in gaze direction are more automatically followed as the stimulus looks more human-like. On the other hand, physical humanness also exerts an indirect effect on gaze-cueing behavior by increasing mind perception, such that more human-like agents are ascribed a greater degree of mind, which in turn seems to facilitate voluntary shifts of attention away from the gazed-at location toward the likely target location (i.e., 80%).

Table 3 Nested model comparison predicting mind rating data from gaze-cueing data

	BIC	χ^2	*p* value
Linear model	−273.16		
Quadratic model	−255.46	1.58	.2
Cubic model	−251.34	.76	.38
Fourth-level polynomial	−246.50	.04	.83

FMRI data

Neural activation associated with mind perception

The neural basis of mind perception was investigated using a parametric regressor of agent image onset during the mind perception task, using trial-by-trial mind ratings to weight the regressor. This analysis allowed for testing whether the

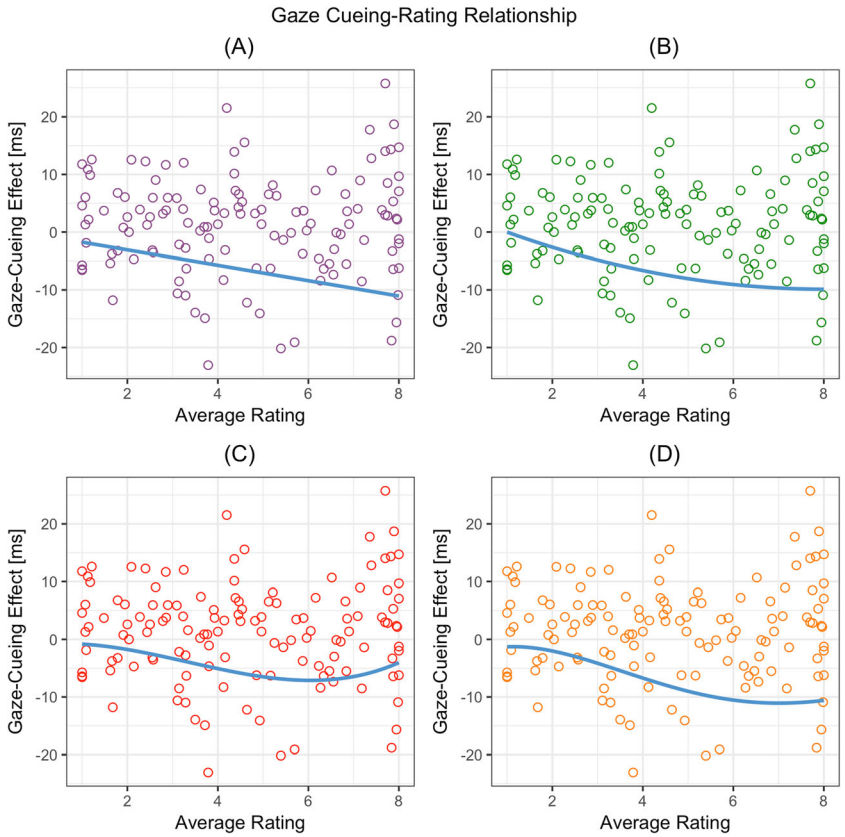

Fig. 7 Average gaze-cueing effects as a function of mind ratings. Data were modeled by a linear (**a**), quadratic (**b**), cubic (**c**), and a fourth-level polynomial (**d**) model. None of the nonlinear models fit significantly better than the linear model, which is evidence that the linear model was the best predictor of gaze-cueing behavior

mind perception task indeed activated the social brain network, and if so, which subdivisions of this network were related to mind perception. The initial, a priori analysis employing all four blocks of the task revealed a cluster of activation within vmPFC, although this cluster of activation did not survive correction for multiple comparisons (see Fig. 9). However, a post hoc analysis that employed only the first two blocks of data, due to concerns overhabituation, revealed statistically robust activation within a similar region of vmPFC that indeed survived correction for multiple comparisons. Specifically, this post hoc analysis revealed a significant cluster located primarily within vmPFC, but also extending into more anterior and less ventral subdivisions of the mPFC such as the frontal poles (peak z = 3.14; 2, 68, 22; 1,050 voxels). No other effects within the whole-brain analysis survived correction for multiple comparisons (see Table 4 and Fig. 10 for the results of the post hoc parametric analysis).

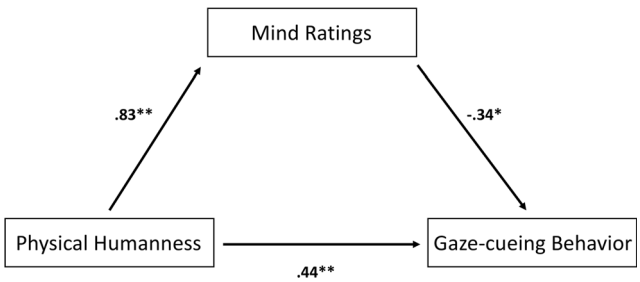

Fig. 8 Path diagram illustrating the mediation model. The mediation analysis revealed both a significant and positive direct effect of physical humanness on gaze cueing, as well as a negative indirect effect, as mediated by mind ratings. Values over the directional arrows reflect standardized coefficients produced from each regression model in the mediation. *p < .05. **p < .01

Association of neural activation during mind perception and social attention

The relationship between brain activation related to mind perception and gaze-cueing performance was investigated using a parametric regressor of agent image onset during mind perception, using gaze-cueing effects to weight the regressor. This analysis allows for testing whether neural activity during the mind perception task significantly matches the patterns of gaze-cueing behavior with the respective agent. Similar to the post hoc analysis of mind ratings described above, the parametric analysis based on gaze-cueing effects revealed significant activation within the vmPFC (peak z = 3.59; −46, 38, −2; 982 voxels). Several other neural regions, such as the left TPJ

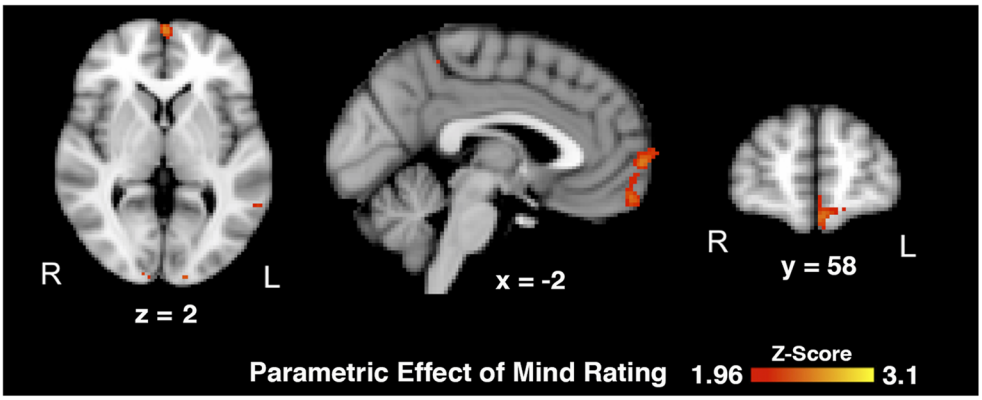

Fig. 9 A priori parametric analysis of fMRI activations based on mind ratings. *Z* maps reflecting onset of the morph images, using mind ratings to weight the parametric regressor. From left to right: coronal (y = 58), sagittal ($x = -2$), and axial ($z = 2$) slices; no activations survived correction for multiple comparisons. (Color figure online)

and insula, right fusiform cortex and middle temporal gyrus, and bilateral occipital cortex were also significant for the parametric analysis based on gaze cueing (see Table 4 and Fig. 11). Most importantly, an overlapping region of the vmPFC was identified for both parametric analyses (see Fig. 11), suggesting that vmPFC may relate not only to mind perception, but also low-level social-cognitive performance during a gaze-cueing task.

Discussion

The goal of the present experiment was to investigate whether within-subject variation in brain activation during mind perception is directly related to variations in social attention performance, and, if so, which social brain areas are most strongly related to this performance measure. For that purpose, we manipulated the physical appearance of social agents (i.e., on a spectrum from robot to human) and measured the effect of this manipulation on two orthogonal tasks: social judgments regarding the agents' capability of having a mind (i.e., ratings and brain activation), and low-level social-cognitive performance during a social attention task (i.e., gaze-cueing effects). Patterns within the behavioral data (i.e., ratings and gaze-cueing effects) were analyzed using a nested model comparison. We used a mediation model to test the complex relations between physical humanness, mind ratings, and gaze-cueing effects. Moreover, a set of parametric analyses of fMRI data was used to investigate the relations between brain activation, mind perception, and low-level social-cognitive performance. In particular, vmPFC was found to be activated not only during mind perception, but the level of vmPFC activity during mind perception was also directly related to subsequent low-level social-cognitive performance on a separate gaze-cueing task. This pattern of results suggests that initial activity within vmPFC actually influences subsequent social-cognitive behavior. However, future research that measures vmPFC activation not only during mind perception but also during subsequent social interactions, is critical in order to identify whether the vmPFC indeed serves as a direct link between mind perception and subsequent low-level social-cognitive behavior. Moreover, additional work using larger sample sizes and more ecologically valid measures of social interaction will be needed to confirm the exact role of the vmPFC in low-level social-cognitive performance.

The linear mixed models revealed that increasing levels of physical humanness were associated with a general increase in mind ratings (i.e., positive social judgments) and low-level social-cognitive performance (i.e., stronger gaze cueing). This is consistent with prior research, demonstrating that increasing levels of physical humanness are associated with increased mind perception (Cheetham et al., 2014; Hackel et al., 2014; Looser & Wheatley, 2010; Martini et al., 2016) and improved low-level social-cognitive performance (Teufel et al., 2009; Wiese et al., 2012; Wykowska et al., 2014). However, we also found that for the counterpredictive social attention task employed here, mind perception (after controlling for physical humanness) seemed to affect gaze cueing in a different manner than physical humanness; that is, increasing levels of physical humanness directly enhanced reflexive attentional orienting to gazed-at locations (i.e., faster reaction times to targets presented at valid compared to invalid locations), suggesting that changes in gaze direction were more automatically followed the more the stimulus looked human-like despite the fact that the gazed-at location was unlikely to contain the target (i.e., counterpredictive cue: 20%). Increasing levels of physical humanness, however, also lead to an increase in mind perception, which seemed to facilitate voluntarily shifts of attention away from the gazed-at location and toward the location that most likely contained the target (i.e., predicted location: 80%). This pattern of results is interesting in the light of previous reports that attentional orienting to gaze cues is hard to suppress given the high social relevance of eye gaze for social learning and the development of close relationships (see Friesen & Kingstone,

Table 4 FMRI activations for the parametric analyses

Anatomical area	Clust.	*x*	*y*	*z*	Z max	Vox.
Mind-rating parametric analysis						
Frontal pole (MFC)	1	2	68	22	3.14	1,050
Frontal pole (vmPFC)	1	0	62	−8	3.04	
Frontal pole (vmPFC)	1	−24	66	−8	2.92	
Frontal medial cortex (vmPFC)	1	−8	52	−16	2.9	
Frontal pole (vmPFC)	1	−8	52	−20	2.88	
Frontal pole (vmPFC)	1	−14	56	−16	2.68	
Gaze-cuing parametric analysis						
Anterior middle temporal gyrus	5	52	−4	−32	3.24	2,175
Anterior middle temporal gyrus	5	50	0	−26	3.16	
White matter	5	14	34	−12	3.12	
Frontal pole (vmPFC)	5	20	44	16	3.09	
Frontal medial cortex (vmPFC)	5	4	48	−22	3.03	
Frontal pole	5	48	36	−4	2.99	
Inferior lateral occipital cortex	4	36	−72	4	3.48	2,130
Inferior lateral occipital cortex	4	46	−74	−2	3.36	
Occipital pole	4	24	−90	34	3.35	
Supracalcarine cortex	4	4	−78	18	3.27	
Occipital pole	4	18	−88	28	3.12	
Fusiform cortex	4	24	−46	−22	3.04	
Inferior lateral occipital cortex	3	−54	−64	12	3.34	1,573
Superior lateral occipital cortex	3	−38	−76	20	3.3	
Posterior superior temporal gyrus	3	−56	−42	6	3.22	
Angular gyrus (TPJ)	3	−48	−56	16	3.18	
Posterior superior temporal gyrus	3	−58	−38	4	3.13	
Angular gyrus (TPJ)	3	−52	−58	26	3.09	
Insular cortex	2	−38	6	−2	3.38	1,217
Putamen	2	−32	−12	6	3.33	
Middle frontal gyrus	2	−26	24	36	3.29	
Putamen	2	−28	−8	10	3.23	
Superior frontal gyrus	2	−24	24	42	3.12	
Middle frontal gyrus	2	−26	10	44	3.02	
Frontal pole	1	−46	38	−2	3.59	982
Orbitofrontal cortex	1	−50	30	−12	3.57	
Orbitofrontal cortex	1	−24	28	0	3.16	
Inferior frontal gyrus	1	−60	30	4	3.11	
Inferior frontal gyrus	1	−56	34	−2	3.09	
White matter	1	−22	28	12	3	

1998). It suggests that although increasing physical humanness enhances the reflexive component of gaze cueing, it also leads to higher levels of mind perception, which seems to facilitate voluntary shifts of attention to uncued, but likely target locations (in counterpredictive cueing paradigms). In particular, it is possible that participants who ascribe higher levels of intentionality to the gazing stimulus might pay more attention to contingencies in its behavior, making it more likely that they pick up on the counterpredictivity of the gaze signal (which can potentially be interpreted as negative intention; e.g. "The agent wants to trick me and make me miss the target"), and adjust attentional orienting accordingly (for reports of top-down modulation of gaze cueing, see Bonifacci, Ricciardelli, Lugli, & Pellicano, 2008; Cazzato, Liuzza, Caprara, Macaluso, & Aglioti, 2015; Dalmaso, Edwards, & Bayliss, 2016; Fox, Mathews, Calder, & Yiend, 2007; Graham, Friesen, Fichtenholtz, & LaBar, 2010; Hungr & Hunt, 2012; Tipples, 2006; Wiese, Wykowska, & Müller, 2014; Wykowska et al.,

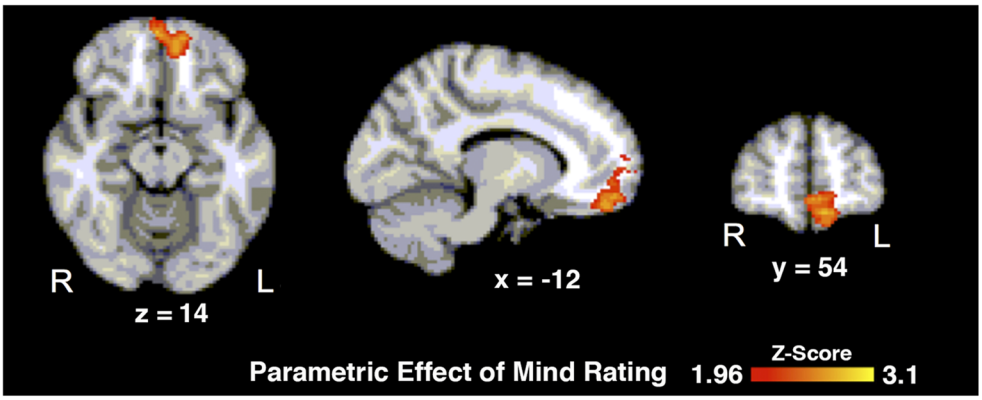

Fig. 10 Post hoc parametric analysis of fMRI activations based on mind ratings. *Z* maps reflecting onset of the morph images, using mind ratings to weight the parametric regressor. From left to right: coronal ($y = 54$), sagittal ($x = -12$), and axial ($z = -14$) slices; cluster corrected ($Z = 1.96$, $p < .05$) at the whole-brain level. (Color figure online)

2014). Although this finding is interesting, since it points at a possible dissociation between perception of intentionality and perception of human appearance, it needs to be interpreted with caution, due to potential issues with multicollinearity in the current experiment, and warrants further investigation.

Analysis of the fMRI data provided insight into the neural regions involved in mind perception and explored how brain activation related to mind perception is related to subsequent gaze-cueing performance (as a proxy for social-cognitive performance). We found that mind ratings were associated with vmPFC activation, a finding that is consistent with prior investigations linking perceptions of intentionality to ventromedial prefrontal areas (Gallagher et al., 2002; Pfeiffer et al., 2014; Sanfey et al., 2003). Activity within vmPFC was also related to low-level social-cognitive performance during gaze cueing, together with a set of other regions including the left TPJ and insula, right medial temporal gyrus and fusiform cortex, and bilateral occipital cortex. Thus, while social attention was associated with a set of regions involved in gaze perception (Nummenmaa & Calder, 2009) and mentalizing (Van Overwalle, 2009), an overlapping region of vmPFC was associated with both mind perception and social attention, suggesting that the vmPFC might play an important role in linking higher-order social-cognitive processes (like mind perception) and performance on lower-level social-cognitive tasks (like gaze cueing). However, additional research that measures neural activity not only during mind perception but also during social-cognitive tasks within the same study will be required to substantiate claims surrounding the link between mind perception and social interaction within the vmPFC.

While prior work has investigated relations between mind perception and social attention (Özdem et al., 2016; Teufel et al., 2009; Wiese et al., 2012; Wykowska et al., 2014), the current study adds to these findings by showing that both mind perception and mechanisms of social attention are related to activation within the vmPFC. This neural region has been associated with mentalizing (Amodio & Frith, 2006; Frith & Frith, 1999; Frith & Frith, 2003; Gallagher et al., 2002), and is involved in impression formation in social situations by providing access to general social knowledge (Mitchell et al., 2006; Szczepanski & Knight, 2014), and retrieving script-based social knowledge (Ghosh et al., 2014; van Kesteren et al., 2012). Moreover, lesions to vmPFC result

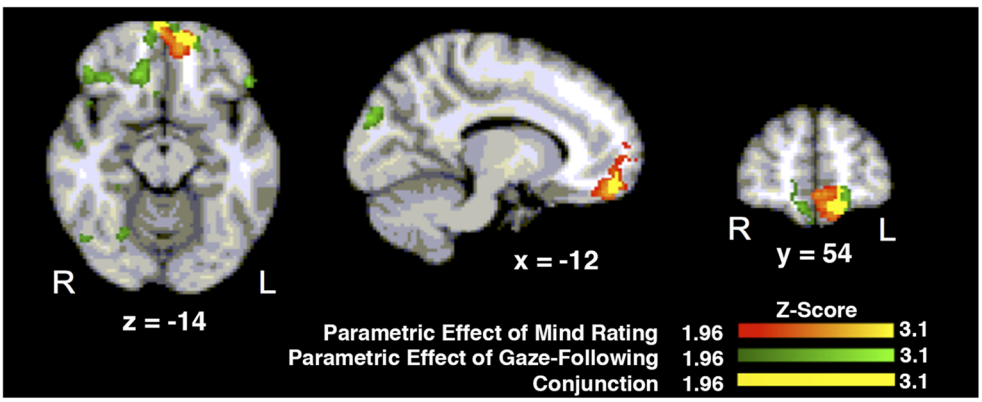

Fig. 11 Parametric analysis of fMRI activations based on gaze cueing effects, mind ratings, and their conjunction. *Z* maps reflecting onset of the morph images, using either mind ratings (orange) or gaze cueing (green) to weight the parametric regressor, along with their conjunction (yellow). From left to right: coronal ($y = 54$), sagittal ($x = -12$), and axial ($z = -14$) slices; cluster corrected ($Z = 1.96$, $p < .05$) at the whole-brain level. (Color figure online)

in impaired mental state understanding (Beer, Heerey, Keltner, Scabini, & Knight, 2003; Stone, Baron-Cohen, & Knight, 1998), emotion recognition (Hornak et al., 2003; Tsuchida & Fellows, 2012), social and moral reasoning (Anderson, Bechara, Damasio, Tranel, & Damasio, 1999), and cognitive empathy (Shamay-Tsoory, Aharon-Peretz, & Perry, 2009). While the vmPFC has previously been shown to modulate higher-order social-cognitive processes involved in economic or strategic decision-making (Gallagher et al., 2002; Sanfey et al., 2003), associations between this neural region and low-level social-cognitive processes like gaze cueing have not previously been reported (to the best of our knowledge).

In addition to the vmPFC literature reviewed above, it is important to note that this neural region has also been shown to track feelings of eeriness toward nonhuman agents in a parametric fashion (Wang & Quadflieg, 2015), and has been suggested as a potential neural correlate of the *uncanny valley* (i.e., nonhuman agents with human-like appearance induce feelings of eeriness when being not perfectly human; Mori, 1970). Based on this research, variation of vmPFC activation in the current experiment could be driven by feelings of eeriness toward agents that are ambiguous in terms of their physical human-likeness. This could lead to a general disengagement from gaze cues as the agents' physical human-likeness increases (this is less likely since increases in physical humanness were associated with stronger reflexive gaze cueing in the current study), or an impaired ability to control attentional orienting in a top-down manner, since processing "uncanny" stimuli has been shown to consume additional cognitive resources to resolve the conflict of whether one is looking at a human or a nonhuman agent (Weis & Wiese, 2017). Previous studies have also shown that activation in medial prefrontal areas, and in particular bilateral vmPFC, is related to evaluating the predictability of stimuli, leading to higher levels of anthropomorphism if a stimulus is hard to predict (Waytz, Morewedge, et al., 2010). Thus, it is possible that activation within vmPFC reflects one's sensitivity to the predictability of gaze cues, and that the level of vmPFC activation is modulated by the degree to which mind is perceived in the agents. This interpretation is in line with previous neurophysiological studies showing that evaluations of predictability are associated with additional neural effort in frontocentral areas for human versus robot agents (Caruana et al., 2016).

In addition to activation in bilateral vmPFC, gaze cueing performance was also associated with activation in left TPJ and insula, right medial temporal gyrus and fusiform cortex, and bilateral occipital cortex. This is in line with previous studies showing that both prefrontal and temporo-occipital areas like bilateral TPJ and STS are implicated in social attention (Nummenmaa & Calder, 2009). Previous studies have also related TPJ activation to judgments about another's intentionality, with stronger activation for agents with versus without a mind (Cavanna & Trimble, 2006; Chaminade et al., 2012; Gallagher et al., 2002; Krach et al., 2008), and shown that TPJ serves as convergence point for social and nonsocial processes (Chang et al., 2013; Krall et al., 2015; Krall et al., 2016; Mitchell, 2008; Scholz et al., 2009). Notably, Özdem et al. (2016) have shown that attentional orienting in response to nonpredictive gaze cues is sensitive to the perceived intentionality underlying these cues (i.e., human controlled vs. preprogrammed) and is associated with activation in bilateral TPJ. The question remains, however, why only relations between gaze cueing performance and the left TPJ reached statistical significance in the current experiment, although both left and right TPJ are activated during mentalizing and social judgment tasks. First, it is possible that the lack of significant activation within the right TPJ could simply arise as a result of issues with statistical power. However, we might also suggest that social functions of the TPJ are lateralized and that the functionalities subserved by the left TPJ (i.e., attribution of human-likeness; Perner et al., 2006) might be more important for the current task than the functionalities of the right TPJ (i.e., mentalizing; Costa et al., 2008; Gallagher et al., 2000; Saxe, 2006; Saxe & Kanwisher, 2003). Specifically, right TPJ activation is found during classic mentalizing tasks (Frith & Frith, 2003; Saxe & Wexler, 2005), while left TPJ activation is related to perspective taking (Samson et al., 2004), anthropomorphism (Chaminade et al., 2007; Cullen et al., 2013; Zink et al., 2011), and processing of agent identity from visual information (Van Overwalle, 2009). Cullen et al. (2013) also showed that gray-matter volume in the left TPJ is related to individual differences in one's willingness to treat nonhuman entities as human-like, and Chaminade et al. (2007) showed that activation in the left TPJ is positively correlated with one's tendency to perceive humanness in motion patterns of nonhuman agents. Both the mind perception task and gaze-cueing task employed in the current study required reasoning about the agents' human-likeness based on visual features, which is expected to trigger different degrees of anthropomorphism (Cheetham, Suter, & Jäncke, 2011, 2014; Martini et al., 2016) and might explain why specifically left TPJ activation was found to be related to social attention performance. Nonetheless, the lateralized function of the TPJ observed in the current experiment will require replication in future work.

Conclusions

In sum, the present study provides evidence that variation in bilateral vmPFC activation, when perceiving the mind of a novel agent, is related to variation in subsequent low-level social-cognitive performance when interacting with that agent, as measured in gaze-cueing performance. Critically, this relationship was identified by recording neural activity upon initial exposure to a set of novel agents, followed by

engaging in a separate, orthogonal low-level social-cognitive task with the same agents. The current study adds to previous research by (a) showing that the degree to which an agent is perceived to have a mind is significantly related to low-level social-cognitive performance on an orthogonal task with the respective agent (with the advantage that measuring brain activation related to mind perception is not confounded by behavioral performance during the low-level social-cognitive task), and (b) identifying a potential neural substrate associated with both mind perception and low-level social-cognitive performance: the bilateral vmPFC. This finding also adds to a growing body of evidence suggesting that mind perception constitutes a source of top-down modulation on attentional orienting, ensuring that more attentional resources are devoted to interactions with agents who are believed to have a mind compared to machine agents without a mind (Krall et al., 2015; Mitchell, 2008; Özdem et al., 2016; Scholz et al., 2009; Wiese et al., 2012; Wykowska et al., 2014). Future research could build on the present results by employing a network perspective, probing the functional or structural connectivity of the vmPFC with other neural regions involved in social cognition and attention.

Acknowledgements We would like to dedicate this paper to the late Raja Parasuraman, who provided insight into the design and implementation of this study.

Funding This work was supported by the Air Force Office of Scientific Research, Grant Number FA9550-10-1-0385, the Center of Excellence in Neuroergonomics, Technology, and Cognition. The authors declare no competing financial interests.

References

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, *15*(1), 1–16. https://doi.org/10.1016/S0885-2014(00)00014-9

Adolphs, R. (2009). The social brain: Neural basis of social knowledge. *Annual Review of Psychology*, *60*, 693–716. https://doi.org/10.1146/annurev.psych.60.110707.163514

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268–277.

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neuroscience*, *2*(11), 1032–1037. https://doi.org/10.1038/14833

Baron-Cohen, S. (1997). Mindblindness: An essay on autism and theory of mind. Cambridge, MA: MIT Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ? *Cognition*, *21*(1), 37–46. doi:10.1016/0010-0277(85)90022-8

Bartneck, C. (2003). Interacting with an embodied emotional character. In *Proceedings of the 2003 International Conference on Designing Pleasurable Products and Interfaces* (pp. 55–60). New York, NY: ACM. https://doi.org/10.1145/782896.782911

Becchio, C., Adenzato, M., & Bara, B. G. (2006). How the brain understands intention: Different neural circuits identify the componential features of motor and prior intentions. *Consciousness and Cognition*, *15*(1), 64–74. https://doi.org/10.1016/j.concog.2005.03.006

Beer, J. S., Heerey, E. A., Keltner, D., Scabini, D., & Knight, R. T. (2003). The regulatory function of self-conscious emotion: Insights from patients with orbitofrontal damage. *Journal of Personality and Social Psychology*, *85*(4), 594–604. https://doi.org/10.1037/0022-3514.85.4.594

Bennewitz, M., Faber, F., Joho, D., Schreiber, M., & Behnke, S. (2005). Towards a humanoid museum guide robot that interacts with multiple persons. In *5th IEEE-RAS International Conference on Humanoid Robots, 2005,* (pp. 418–423). https://doi.org/10.1109/ICHR.2005.1573603

Bering, J., & Johnson, D. (2005). "O lord you perceive my thoughts from afar": Recursiveness and the evolution of supernatural agency. *Journal of Cognition and Culture*, *5*, 118–142. https://doi.org/10.1163/1568537054068679

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex (New York, NY)*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Bisio, A., Sciutti, A., Nori, F., Metta, G., Fadiga, L., Sandini, G., & Pozzo, T. (2014). Motor contagion during human-human and human–robot interaction. *PLOS ONE*, *9*(8), e106172. https://doi.org/10.1371/journal.pone.0106172

Bonifacci, P., Ricciardelli, P., Lugli, L., & Pellicano, A. (2008). Emotional attention: effects of emotion and gaze direction on overt orienting of visual attention. *Cognitive Processing*, *9*(2), 127–135. https://doi.org/10.1007/s10339-007-0198-3

Brothers, L. (2002). The social brain: A project for integrating primate behavior and neurophysiology in a new domain. In J. T. Cacioppo (Ed.), Foundations in social neuroscience (pp. 367–385). Cambridge, MA: MIT Press.

Bzdok, D., Langner, R., Schilbach, L., Engemann, D. A., Laird, A. R., Fox, P. T., & Eickhoff, S. (2013). Segregation of the human medial prefrontal cortex in social cognition. *Frontiers in Human Neuroscience*, *7*, 232.

Caruana, N., McArthur, G., Woolgar, A., & Brock, J. (2016). Simulating social interactions for the experimental investigation of joint attention. *Neuroscience & Biobehavioral Reviews*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0149763416304778

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage*, *12*(3), 314–325. https://doi.org/10.1006/nimg.2000.0612

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain*, *129*(3), 564–583.

Cazzato, V., Liuzza, M. T., Caprara, G. V., Macaluso, E., & Aglioti, S. M. (2015). The attracting power of the gaze of politicians is modulated by the personality and ideological attitude of their voters: A functional magnetic resonance imaging study. *The European Journal of Neuroscience*, *42*(8), 2534–2545. https://doi.org/10.1111/ejn.13038

Chaminade, T., & Decety, J. (2002). Leader or follower? Involvement of the inferior parietal lobule in agency. *Neuroreport*, *13*(15), 1975–1978.

Chaminade, T., Hodgins, J., & Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience*, *2*(3), 206–216. https://doi.org/10.1093/scan/nsm017

Chaminade, T., Rosset, D., Da Fonseca, D., Nazarian, B., Lutscher, E., Cheng, G., & Deruelle, C. (2012). How do we think machines think? An fMRI study of alleged competition with an artificial

intelligence. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00103

Chang, C.-F., Hsu, T.-Y., Tseng, P., Liang, W.-K., Tzeng, O. J. L., Hung, D. L., & Juan, C.-H. (2013). Right temporoparietal junction and attentional reorienting. *Human Brain Mapping*, *34*(4), 869–877. https://doi.org/10.1002/hbm.21476

Cheetham, M., Suter, P., & Jäncke, L. (2011). The human likeness dimension of the "uncanny valley hypothesis": Behavioral and functional MRI findings. *Frontiers in Human Neuroscience*, *5*, 126. https://doi.org/10.3389/fnhum.2011.00126

Cheetham, M., Suter, P., & Jancke, L. (2014). Perceptual discrimination difficulty and familiarity in the uncanny valley: More like a "Happy Valley." *Frontiers in Psychology*, *5*, 1219. https://doi.org/10.3389/fpsyg.2014.01219

Chong, T. T.-J., Cunnington, R., Williams, M. A., Kanwisher, N., & Mattingley, J. B. (2008). fMRI adaptation reveals mirror neurons in human inferior parietal cortex. *Current Biology: CB*, *18*(20), 1576–1580. https://doi.org/10.1016/j.cub.2008.08.068

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S., (2003). Applied multiple regression/correlation analysis for the behavioral sciences. New York, NY: Routledge.

Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2012). Dissociable neural correlates of stereotypes and other forms of semantic knowledge. *Social Cognitive and Affective Neuroscience*, *7*(7), 764–770. https://doi.org/10.1093/scan/nsr053

Costa, A., Torriero, S., Oliveri, M., & Caltagirone, C. (2008). Prefrontal and temporo-parietal involvement in taking others' perspective: TMS evidence. *Behavioural Neurology*, *19*(1/2), 71–74.

Cullen, H., Kanai, R., Bahrami, B., & Rees, G. (2013). Individual differences in anthropomorphic attributions and human brain structure. *Social Cognitive and Affective Neuroscience*, *9*(9),1276–1280. doi: 10.1093/scan/nst10

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353–380. https://doi.org/10.1016/j.cognition.2008.03.006

Dalmaso, M., Edwards, S. G., & Bayliss, A. P. (2016). Re-encountering individuals who previously engaged in joint gaze modulates subsequent gaze cueing. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *42*(2), 271–284. https://doi.org/10.1037/xlm0000159

Decety, J., & Chaminade, T. (2003). When the self represents the other: A new cognitive neuroscience view on psychological identification. *Consciousness and Cognition*, *12*(4), 577–596. https://doi.org/10.1016/S1053-8100(03)00076-X

Dinstein, I., Hasson, U., Rubin, N., & Heeger, D. J. (2007). Brain areas selective for both observed and executed movements. *Journal of Neurophysiology*, *98*(3), 1415–1427. https://doi.org/10.1152/jn.00238.2007

DiSalvo, C. F., Gemperle, F., Forlizzi, J., & Kiesler, S. (2002). All robots are not created equal: The design and perception of humanoid robot heads. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (pp. 321–326). New York, NY: ACM. https://doi.org/10.1145/778712.778756

Epley, N., Waytz, A., Akalis, S., & Cacioppo, J. T. (2008). When we need a human: Motivational determinants of anthropomorphism. *Social Cognition*, *26*(2), 143–155. https://doi.org/10.1521/soco.2008.26.2.143

Fairhall, S. L., Anzellotti, S., Ubaldi, S., & Caramazza, A. (2014). Person- and place-selective neural substrates for entity-specific semantic access. Cerebral Cortex (New York, N.Y.: *1991)*, *24*(7), 1687–1696. https://doi.org/10.1093/cercor/bht039

Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J., & Jeannerod, M. (2003). Modulating the experience of agency: A positron emission tomography study. *Neuro Image*, *18*(2), 324–333.

Fox, E., Mathews, A., Calder, A. J., & Yiend, J. (2007). Anxiety and sensitivity to gaze direction in emotionally expressive faces. *Emotion (Washington, D.C.)*, *7*(3), 478–486. https://doi.org/10.1037/1528-3542.7.3.478

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495. https://doi.org/10.3758/BF03208827

Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(2), 319–329. https://doi.org/10.1037/0096-1523.30.2.319

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*(4), 694–724. https://doi.org/10.1037/0033-2909.133.4.694

Frith, C. D., & Frith, U. (1999). Interacting minds—A biological basis. *Science (New York, N.Y.)*, *286*(5445), 1692–1695.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534. https://doi.org/10.1016/j.neuron.2006.05.001

Frith, U., & Frith, C. (2001). The biological basis of social interaction. *Current Directions in Psychological Science*, *10*(5), 151–155.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*(1431), 459–473. https://doi.org/10.1098/rstb.2002.1218

Fussell, S. R., Kiesler, S., Setlock, L. D., & Yew, V. (2008). How people anthropomorphize robots. In *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 145–152). https://doi.org/10.1145/1349822.1349842

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of "theory of mind." *Trends in Cognitive Sciences*, *7*(2), 77–83.

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of "theory of mind" in verbal and nonverbal tasks. *Neuropsychologia*, *38*(1), 11–21.

Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage*, *16*(3, Pt. 1), 814–821.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain: A Journal of Neurology*, *119*(Pt. 2), 593–609.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, *8*(9), 396–403. https://doi.org/10.1016/j.tics.2004.07.002

Gao, T., McCarthy, G., & Scholl, B. J. (2010). The wolfpack effect: Perception of animacy irresistibly influences interactive behavior. *Psychological Science*, *21*(12), 1845–1853. https://doi.org/10.1177/0956797610388814

Ghosh, V. E., Moscovitch, M., Melo Colella, B., & Gilboa, A. (2014). Schema representation in patients with ventromedial PFC lesions. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(36), 12057–12070. https://doi.org/10.1523/JNEUROSCI.0740-14.2014

Graham, J., & Haidt, J. (2010). Beyond beliefs: Religions bind individuals into moral communities. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology*, *14*(1), 140–150. https://doi.org/10.1177/1088868309353415

Graham, R., Friesen, C. K., Fichtenholtz, H. M., & LaBar, K. S. (2010). Modulation of reflexive orienting to gaze direction by facial expressions. *Visual Cognition*, *18*(3), 331–368. https://doi.org/10.1080/13506280802689281

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619–619. https://doi.org/10.1126/science.1134475

Gray, K., & Wegner, D. M. (2008). The sting of intentional pain. *Psychological Science*, *19*(12), 1260–1262. https://doi.org/10.1111/j.1467-9280.2008.02208.x

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*(2), 101–124. https://doi.org/10.1080/1047840X.2012.651387

Grèzes, J., Berthoz, S., & Passingham, R. E. (2006). Amygdala activation when one is the target of deceit: Did he lie to you or to someone else? *NeuroImage*, *30*(2), 601–608. https://doi.org/10.1016/j.neuroimage.2005.09.038

Grèzes, J., Frith, C., & Passingham, R. E. (2004). Brain mechanisms for inferring deceit in the actions of others. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *24*(24), 5500–5505. https://doi.org/10.1523/JNEUROSCI.0219-04.2004

Hackel, L. M., Looser, C. E., & Van Bavel, J. J. (2014). Group membership alters the threshold for mind perception: The role of social identity, collective identification, and intergroup threat. *Journal of Experimental Social Psychology*, *52*, 15–23.

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, *26*(3), 245–256. https://doi.org/10.1016/j.evolhumbehav.2005.01.002

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science*, *17*(10), 847–853. https://doi.org/10.1111/j.1467-9280.2006.01793.x

Hornak, J., Bramham, J., Rolls, E. T., Morris, R. G., O'Doherty, J., Bullock, P. R., & Polkey, C. E. (2003). Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. *Brain: A Journal of Neurology*, *126*(Pt. 7), 1691–1712. https://doi.org/10.1093/brain/awg168

Huang, C. M., & Thomaz, A. L. (2011). Effects of responding to, initiating and ensuring joint attention in human-robot interaction. Paper published in *2011 RO-MAN* (pp. 65–71). https://doi.org/10.1109/ROMAN.2011.6005230

Huey, E. D., Krueger, F., & Grafman, J. (2006). Representations in the human prefrontal cortex. *Current Directions in Psychological Science*, *15*(4), 167–171.

Hungr, C. J., & Hunt, A. R. (2012). Physical self-similarity enhances the gaze-cueing effect. *Quarterly Journal of Experimental Psychology (2006)*, *65*(7), 1250–1259. https://doi.org/10.1080/17470218.2012.690769

Hynes, C. A., Baird, A. A., & Grafton, S. T. (2006). Differential role of the orbital frontal lobe in emotional versus cognitive perspective-taking. *Neuropsychologia*, *44*(3), 374–383. https://doi.org/10.1016/j.neuropsychologia.2005.06.011

Iacoboni, M. (2005). Neural mechanisms of imitation. *Current Opinion in Neurobiology*, *15*(6), 632–637. https://doi.org/10.1016/j.conb.2005.10.010

Jenkins, A. C., Macrae, C. N., & Mitchell, J. P. (2008). Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(11), 4507–4512. https://doi.org/10.1073/pnas.0708785105

Keysers, C., & Perrett, D. I. (2004). Demystifying social cognition: A Hebbian perspective. *Trends in Cognitive Sciences*, *8*(11), 501–507. https://doi.org/10.1016/j.tics.2004.09.005

Kiesler, S., Powers, A., Fussell, S. R., & Torrey, C. (2008). Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition*, *26*(2), 169.

Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J., & Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(32), 10153–10159. https://doi.org/10.1523/JNEUROSCI.2668-09.2009

Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological movement on action. *Current Biology: CB*, *13*(6), 522–525.

Konishi, S., & Kitagawa, G. (2008). Information criteria and statistical modeling. New York, NY: Springer Science & Business Media. Retrieved from https://books.google.com/books?hl=en&lr=&id=3I9ZJusaYh0C&oi=fnd&pg=PA1&dq=Konishi+%26+Kitagawa,+2008&ots=Y_S1YyHleU&sig=6HHmYgfK5xgLLn0FrOqLohteqOA

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*(7), e2597.

Krall, S. C., Rottschy, C., Oberwelland, E., Bzdok, D., Fox, P. T., Eickhoff, S. B., . . . Konrad, K. (2015). The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Structure and Function*, *220*(2), 587–604.

Krall, S. C., Volz, L. J., Oberwelland, E., Grefkes, C., Fink, G. R., & Konrad, K. (2016). The right temporoparietal junction in attention and social interaction: A transcranial magnetic stimulation study. *Human Brain Mapping*, *37*(2), 796–807.

Kupferberg, A., Huber, M., Helfer, B., Lenz, C., Knoll, A., & Glasauer, S. (2012). Moving just like you: Motor interference depends on similar motility of agent and observer. *PLOS ONE*, *7*(6), e39637. https://doi.org/10.1371/journal.pone.0039637

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind." *Trends in Cognitive Sciences*, *8*(12), 528–533. https://doi.org/10.1016/j.tics.2004.10.001

Looser, C. E., & Wheatley, T. (2010). The tipping point of animacy how, when, and where we perceive life in a face. *Psychological Science*, *21*(12), 1854–1862.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF) (CD-ROM, 91–630). Solna, Sweden: Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet.

Mackinnon, D., Krull, J., & Lockwood, C. (2000). Equivalence of the mediation, confounding, and suppression effect. *Prevention Science, 1*(4), 173–181. https://doi.org/10.1023/A1026595011371

Martini, M. C., Gonzalez, C. A., & Wiese, E. (2016). Seeing minds in others–Can agents with robotic appearance have human-like preferences? *PLOS ONE*, *11*(1), e0146310.

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 Sciences of the United States of America*, *98*(20), 11832–11835. https://doi.org/10.1073/pnas.211415698

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex*, *18*(2), 262–271.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2004). Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *Journal of Neuroscience*, *24*(21), 4912–4917. https://doi.org/10.1523/JNEUROSCI.0481-04.2004

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, *50*(4), 655–663. https://doi.org/10.1016/j.neuron.2006.03.040

Mori, M. (1970). The uncanny valley. *Energy*, *7*(4), 33–35.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology: CB*, *20*(8), 750–756. https://doi.org/10.1016/j.cub.2010.02.045

Mutlu, B., Forlizzi, J., & Hodgins, J. (2006). A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *2006 6th IEEE-RAS International Conference on Humanoid Robots* (pp. 518–523). https://doi.org/10.1109/ICHR.2006.321322

Mutlu, B., Kanda, T., Forlizzi, J., Hodgins, J., & Ishiguro, H. (2012). Conversational gaze mechanisms for humanlike robots. *ACM*

Transactions on Interactive Intelligent Systems, *1*(2), 12:1–12:33. https://doi.org/10.1145/2070719.2070725

Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences*, *13*(3), 135–143. https://doi.org/10.1016/j.tics.2008.12.006

Oberman, L. M., McCleery, J. P., Ramachandran, V. S., & Pineda, J. A. (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*(13), 2194–2203. https://doi.org/10.1016/j.neucom.2006.02.024

Ohnishi, T., Moriguchi, Y., Matsuda, H., Mori, T., Hirakata, M., Imabayashi, E., . . . Uno, A. (2004). The neural network for the mirror system and mentalizing in normally developed children: An fMRI study. *Neuroreport*, *15*(9), 1483–1487.

Ohtsubo, Y. (2007). Perceived intentionality intensifies blameworthiness of negative behaviors: Blame-praise asymmetry in intensification effect. *Japanese Psychological Research*, *49*(2), 100–110. https://doi.org/10.1111/j.1468-5884.2007.00337.x

Özdem, C., Wiese, E., Wykowska, A., Müller, H., Brass, M., & Overwalle, F. V. (2016). Believing androids—fMRI activation in the right temporo-parietal junction is modulated by ascribing intentions to non-human agents. *Social Neuroscience*, *2*(5), 582–593. https://doi.org/10.1080/17470919.2016.1207702

Oztop, E., Franklin, D. W., Chaminade, T., & Cheng, G. (2005). Human–humanoid interaction: Is a humanoid robot perceived as a human? *International Journal of Humanoid Robotics*, *2*(4), 537–559. https://doi.org/10.1142/S0219843605000582

Park, S., & Catrambone, R. (2007). Social facilitation effects of virtual humans. *Human Factors*, *49*(6), 1054–1060. https://doi.org/10.1518/001872007X249910

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neuroscience*, *1*(3/4), 245–258. https://doi.org/10.1080/17470910600989896

Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., & Vogeley, K. (2014). Why we interact: On the functional role of the striatum in the subjective experience of social interaction. *NeuroImage*, *101*, 124–137. https://doi.org/10.1016/j.neuroimage.2014.06.061

Pfeiffer, U. J., Timmermans, B., Bente, G., Vogeley, K., & Schilbach, L. (2011). A non-verbal Turing test: Differentiating mind from machine in gaze-based social interaction. *PLOS ONE*, *6*(11), e27591. https://doi.org/10.1371/journal.pone.0027591

Pfeiffer-Leßmann, N., Pfeiffer, T., & Wachsmuth, I. (2018). An operational model of joint attention—Timing of gaze patterns in interactions between humans and a virtual human. http://mindmodeling.org/cogsci2012/

Pobric, G., & de Hamilton, A. F. C. (2006). Action understanding requires the left inferior frontal cortex. *Current Biology: CB*, *16*(5), 524–529. https://doi.org/10.1016/j.cub.2006.01.033

Press, C., Bird, G., Flach, R., & Heyes, C. (2005). Robotic movement elicits automatic imitation. *Brain Research. Cognitive Brain Research*, *25*(3), 632–640. https://doi.org/10.1016/j.cogbrainres.2005.08.020

Press, C., Gillmeister, H., & Heyes, C. (2007). Sensorimotor experience enhances automatic imitation of robotic action. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1625), 2509–2514. https://doi.org/10.1098/rspb.2007.0774

Riether, N., Hegel, F., Wrede, B., & Horstmann, G. (2012). Social facilitation with social robots? In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 41–47). https://doi.org/10.1145/2157689.2157697

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nature Neuroscience*, *4*(5), 546–550.

Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, *7*(5), 499–500.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science (New York, N.Y.)*, *300*(5626), 1755–1758. https://doi.org/10.1126/science.1082976

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*(2), 235–239.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." *NeuroImage*, *19*(4), 1835–1842.

Saxe, R., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science*, *17*(8), 692–699. https://doi.org/10.1111/j.1467-9280.2006.01768.x

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia*, *43*(10), 1391–1399.

Saygin, A. P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain: A Journal of Neurology*, *130*(Pt. 9), 2452–2461. https://doi.org/10.1093/brain/awm162

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, *7*(4), 413–422. https://doi.org/10.1093/scan/nsr025

Saygin, A. P., Wilson, S. M., Hagler, D. J., Bates, E., & Sereno, M. I. (2004). Point-light biological motion perception activates human premotor cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *24*(27), 6181–6188. https://doi.org/10.1523/JNEUROSCI.0504-04.2004

Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E. N., & Saxe, R. (2009). Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLOS ONE*, *4*(3), e4869.

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain: A Journal of Neurology*, *132*(Pt. 3), 617–627. https://doi.org/10.1093/brain/awn279

Shariff, A. F., & Norenzayan, A. (2007). God is watching you: Priming God concepts increases prosocial behavior in an anonymous economic game. *Psychological Science*, *18*(9), 803–809. https://doi.org/10.1111/j.1467-9280.2007.01983.x

Sidner, C. L., Kidd, C. D., Lee, C., & Lesh, N. (2004). Where to look: A study of human-robot engagement. In *In Proceedings of In TELligent User Interfaces* (pp. 78–84). New York, NY: ACM Press.

Spunt, R. P., Meyer, M. L., & Lieberman, M. D. (2015). The default mode of human brain function primes the intentional stance. *Journal of Cognitive Neuroscience*, *27*(6), 1116–1124. https://doi.org/10.1162/jocn_a_00785

Staudte, M., & Crocker, M. W. (2011). Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition*, *120*(2), 268–291. doi:10.1016/j.cognition.2011.05.005

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, *10*(5), 640–656.

Szczepanski, S. M., & Knight, R. T. (2014). Insights into human behavior from lesions to the prefrontal cortex. *Neuron*, *83*(5), 1002–1018. https://doi.org/10.1016/j.neuron.2014.08.011

Teufel, C., Alexis, D. M., Todd, H., Lawrance-Owen, A. J., Clayton, N. S., & Davis, G. (2009). Social cognition modulates the sensory coding of observed gaze direction. *Current Biology*, *19*(15), 1274–1277.

Tipples, J. (2006). Fear and fearfulness potentiate automatic orienting to eye gaze. *Cognition and Emotion*, *20*(2), 309–320. https://doi.org/10.1080/02699930500405550

Tsuchida, A., & Fellows, L. K. (2012). Are you upset? Distinct roles for orbitofrontal and lateral prefrontal cortex in detecting and distinguishing facial expressions of emotion. Cerebral Cortex (New York, N.Y.: 1991), *22*(12), 2904–2912. https://doi.org/10.1093/cercor/bhr370

Umiltà, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., & Rizzolatti, G. (2001). I know what you are doing. a neurophysiological study. *Neuron*, *31*(1), 155–165.

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, *30*(3), 829–858. https://doi.org/10.1002/hbm.20547

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, *48*(3), 564–584.

Völlm, B. A., Taylor, A. N. W., Richardson, P., Corcoran, R., Stirling, J., McKie, S., . . . Elliott, R. (2006). Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *NeuroImage*, *29*(1), 90–98. https://doi.org/10.1016/j.neuroimage.2005.07.022

Wagner, D. D., Kelley, W. M., & Heatherton, T. F. (2011). Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. *Cerebral Cortex*, *21*(12), 2788–2796. https://doi.org/10.1093/cercor/bhr074

Wang, Y., & Quadflieg, S. (2015). In our own image? Emotional and neural processing differences when observing human-human vs human-robot interactions. *Social Cognitive and Affective Neuroscience*, *10*(11), 1515–1524. https://doi.org/10.1093/scan/nsv043

Weis, P. P., & Wiese, E. (2017). Cognitive conflict as possible origin of the Uncanny Valley. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 1599–1603. https://doi.org/10.1177/1541931213601763

Wheatley, T., Weinberg, A., Looser, C., Moran, T., & Hajcak, G. (2011). Mind perception: Real but not artificial faces sustain neural activity beyond the N170/VPP. *PLOS ONE*, *6*(3), e17960.

Wiese, E., Metta, G., & Wykowska, A. (2017). Robots as intentional agents: Using neuroscientific methods to make robots appear more social. *Frontiers in Psychology*, *8*, 1663. https://doi.org/10.3389/fpsyg.2017.01663

Wiese, E., Wykowska, A., & Müller, H. J. (2014). What we observe is biased by what other people tell us: Beliefs about the reliability of gaze behavior modulate attentional orienting to gaze cues. *PLOS ONE*, *9*(4), e94529. https://doi.org/10.1371/journal.pone.0094529

Wiese, E., Wykowska, A., Zwickel, J., & Müller, H. J. (2012). I see what you mean: How attentional selection is shaped by ascribing intentions to others. *PLOS ONE*, *7*(9), e45391.

Wood, J. N., & Grafman, J. (2003). Human prefrontal cortex: processing and representational perspectives. *Nature Reviews Neuroscience*, *4*(2), 139–147.

Woods, S., Dautenhahn, K., & Kaouri, C. (2005). Is someone watching me? Consideration of social facilitation effects in human-robot interaction experiments. In *2005 International Symposium on Computational Intelligence in Robotics and Automation* (pp. 53–60). https://doi.org/10.1109/CIRA.2005.1554254

Wykowska, A., Wiese, E., Prosser, A., & Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. *PLOS ONE*, *9*(4), e94339.

Yamazaki, A., Yamazaki, K., Burdelski, M., Kuno, Y., & Fukushima, M. (2010). Coordination of verbal and non-verbal actions in human–robot interaction at museums and exhibitions. *Journal of Pragmatics*, *42*(9), 2398–2414. https://doi.org/10.1016/j.pragma.2009.12.023

Zink, C. F., Kempf, L., Hakimi, S., Rainey, C. A., Stein, J. L., & Meyer-Lindenberg, A. (2011). Vasopressin modulates social recognition-related activity in the left temporoparietal junction in humans. *Translational Psychiatry*, *1*(4), e3.