International Journal of Social Robotics (2021) 13:863–876
https://doi.org/10.1007/s12369-020-00675-4

# Does Context Matter? Effects of Robot Appearance and Reliability on Social Attention Differs Based on Lifelikeness of Gaze Task

**Abdulaziz Abubshait**[1,2] · **Patrick P. Weis**[1,3] · **Eva Wiese**[1]

Accepted: 26 June 2020 / Published online: 12 July 2020

## Abstract

Social signals, such as changes in gaze direction, are essential cues to predict others' mental states and behaviors (i.e., mentalizing). Studies show that humans can mentalize with nonhuman agents when they perceive a mind in them (i.e., mind perception). Robots that physically and/or behaviorally resemble humans likely trigger mind perception, which enhances the relevance of social cues and improves social-cognitive performance. The current experiments examine whether the effect of physical and behavioral influencers of mind perception on social-cognitive processing is modulated by the lifelikeness of a social interaction. Participants interacted with robots of varying degrees of physical (humanlike vs. robot-like) and behavioral (reliable vs. random) human-likeness while the lifelikeness of a social attention task was manipulated across five experiments. The first four experiments manipulated lifelikeness via the physical realism of the robot images (Study 1 and 2), the biological plausibility of the social signals (Study 3), and the plausibility of the social context (Study 4). They showed that humanlike behavior affected social attention whereas appearance affected mind perception ratings. However, when the lifelikeness of the interaction was increased by using videos of a human and a robot sending the social cues in a realistic environment (Study 5), social attention mechanisms were affected both by physical appearance and behavioral features, while mind perception ratings were mainly affected by physical appearance. This indicates that in order to understand the effect of physical and behavioral features on social cognition, paradigms should be used that adequately simulate the lifelikeness of social interactions.

**Keywords** Gaze-cueing · Social cognition · Human–robot gaze · Mind perception

## 1 Introduction

Humans make inferences based on observing nonverbal social behaviors, such as changes in gaze direction, and make predictions about the intentions underlying these behaviors [1–3]. Reasoning about internal states occurs when an entity is believed to have a mind (i.e., *mind perception*), with the capability of possessing internal states, such as emotions, preferences, and intentions [4]. While there is no doubt that humans can experience internal states, the degree to which nonhuman entities like robots can trigger mind perception can depend on the human-likeness of the entity's physical appearance and displayed behaviors [5]. Previous studies have shown that when an entity is *believed* to "have a mind" (independent of its actual mind status), more social relevance is ascribed to its nonverbal signals [6]. Specifically, it was shown that attentional orienting to changes in gaze direction [7], was more pronounced when gaze signals were believed to be generated by a human (i.e., an entity with a mind) as opposed to a non-intentional machine [8–12]. While these studies show a clear link between beliefs about an agent's mind status and social-cognitive processing, they do not inform about potential effects of physical (e.g., humanlike appearance), behavioral (e.g., biological motion), and contextual (e.g., lifelikeness of interaction) features on mind perception and social cognitive processes. This is crucial for social roboticists, in order to understand how to design robots that trigger social-cognitive processes similar

---

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s12369-020-00675-4) contains supplementary material, which is available to authorized users.

✉ Abdulaziz Abubshait
Abdulaziz.abubshait@iit.it

[1] George Mason University, Fairfax, VA, USA

[2] Italian Institute of Technology, Genoa, Italy

[3] Ulm University, Ulm, Germany

to humans. To address this, the current study manipulated physical and behavioral agent features, as well as the lifelikeness of the social interaction and examined the combined effects of these parameters on mind perception on social cognitive processing.

To investigate the effects of physical, behavioral, and contextual parameters on social cognitive processing, we used a social attention task that measured the extent to which participants orient their attention to a location that is spatially cued by a face's change in gaze direction (i.e., gaze cues) [7]. For this purpose, a face stimulus was presented in the center of a screen that first looked straight and then changed its gaze direction to either the left or right side of the screen, which constitutes the gaze cue. The gaze cue is then followed by a target that participants were asked to respond to as quickly and accurately as possible. Observing gaze cues shifts the observer's attention to the gazed-at location, which results in faster reaction times to targets that are presented at the gazed-at location (i.e., valid trials) than those opposite of the gaze cue (i.e., invalid trials). The difference in reaction times between valid and invalid trials is called the *gaze-cueing effect* and its size is indicative of the extent to which people attend to where an interaction partner is looking [7]. This task was chosen for four reasons: First, attentional orienting to gaze signals is a social-cognitive process that is essential for human development and a prerequisite for higher-order social-cognitive processes, such as mentalizing [9, 13, 14]. Second, prior studies have shown that social attention is sensitive to the perceived social relevance of an interaction [10, 12, 15–19], and specifically to the degree to which the gazer is perceived as having a mind [8, 10, 11, 18]. Third, cognitive modeling of nonverbal signals like gaze cues in nonhuman agents has been a central topic for HRI since robots that display nonverbal signals can evoke natural responses from the interacting human [15, 20, 21]. Fourth, the paradigm allows for the simple manipulation of physical parameters of the gazer (i.e., humanlike vs. robot-like), behavioral parameters of the gaze signal (i.e., predictiveness and biological plausibility), and contextual parameters of the interaction (i.e., presence of reference objects and lifelikeness of the simulation).

## 1.1 Causes and Effects of Mind Perception

Research suggests that mind perception can be manipulated via physical and behavioral agent features, as well as contextual features of an interaction. Agents that physically resemble humans are more likely to be perceived as "having a mind" than actors that appear mechanistic [20, 22–25]. Specifically, when robots have similar physical characteristics as humans (e.g., humanlike head dimensions) or when their human-likeness is increased by adding a high percentage of humanness via morphing a human face into nonhuman faces (e.g., dolls, robots or stuffed animals), people tend to ascribe a higher mind status to them [22, 24, 26, 27]. Likewise, people also perceive "more mind" in other agents when their behavior is predictable, for instance when an agent's gaze signals reliably indicate the location of an upcoming target [28] or when their behaviors generate unexpected outcomes, for instance when playing economic games with entities whose human-likeness is unknown [29]. People are also more likely to attribute mental states to inanimate objects when they move at similar speeds as human agents [30], when they show behavioral patterns reminiscent of human–human interactions [31, 32] (even when the objects are abstract, such as triangles [33]), or when they interact with nonhuman agents that display negative intentions or violate social norms, such as robots that cheat during an interactive game (e.g., rock-paper-scissors; [34]). Finally, studies have shown that contextual features of an interaction can influence the extent of mind perception. For example, when the outcome of an interaction is negative, people attribute more mental capacities to robots [35], and focusing on the body rather than the face of another agent changes the dynamic of mind perception such that it reduces perceptions of the agency component of mind perception (i.e., planning, acting) but increases perceptions of the experience component (i.e., emotion, sensation [36]).

Physical, behavioral and context features not only affect mind perception, but have also been shown to change the social relevance ascribed to others' actions and consequently modulate social-cognitive processing [11, 12, 37]. Increasing an agent's physical human-likeness is associated with enhanced social cognitive processing [20], as well as increased activation in social brain areas [11], but it can also have negative consequences when an agent's appearance is categorically ambiguous and cannot easily be classified as "human" or "nonhuman" [26, 38]. With regard to behavioral factors, robots emulating humanlike behaviors have a positive effect on social-cognitive processes. For example, when robots engage in mutual gaze (as opposed to looking down) with a human interaction partner prior to executing a gaze cue, people follow the signal more strongly resulting in faster responses to gazed-at targets [15]. Likewise, when observed changes in gaze direction are perceived as being predictive of a target's location, attention orienting in response to these cues become spatially more specific resulting in faster reaction times to targets presented at the gazed-at location [28]. Similarly, studies that manipulate the context in which a cue is observed show that participants are more likely to follow a robot's behavioral cue when a deliberate delay is introduced that makes the robot's cues more salient [39] or when adding a reference for where an object can be presented at the time of the gaze shift [17].

### 1.2 Importance of Lifelikeness When Examining Mind Perception and Social Cognition

These studies show that mind perception can be manipulated through physical, behavioral and contextual features [14, 24, 26], and that all features in isolation modulate certain aspects of social cognition [10, 11, 23, 28, 40, 41]. However, in everyday interactions, it is likely that those parameters do not occur in isolation, requiring research to look at the combined effects of these factors on social cognition. Of particular importance for HRI is the question of what happens when robot appearance and behavior are incongruent, for instance when a robot looks humanlike but behaves like a machine (e.g., due to delays or lack of biological motion). As one of the few studies on this topic, Saygin et al. [42] have shown that while activation in the action-perception network of the brain was not sensitive to the appearance or motion of an agent (humanlike vs. machine-like), being exposed to a mismatch between the human-likeness of an agent's appearance and behavior (e.g., agent with robot appearance showing biological motion) was associated with a higher prediction error signals indicating that people expect congruency between physical appearance and behavior and that these two mind perception factors do not work in isolation [42]. Furthermore, Abubshait and Wiese [37] showed that when being examined in combination, physical and behavioral agent features seem to affect different aspects of social cognition than was previously reported: independent of appearance, an agent whose gaze reliably predicted the location of a target induced stronger attentional orienting in response to its gaze signals than an agent whose gaze signals were non-predictive; in contrast, humanlike versus robot-like appearance affected subjective mind perception ratings but did not affect social attention. Taken together, these findings suggest that triggers of mind perception do not work in isolation but interact in more complex ways and thus need to be examined in combination in paradigms that sufficiently simulate the complexity or lifelikeness of social interactions.

### 1.3 Aim of Study

The goal of the current study is to examine (1) how physical and behavioral agent features affect mind perception and social attention when being manipulated within the same paradigm (Experiments 1–4), and (2) whether the effect of these parameters changes as the lifelikeness of the paradigm is increased (Experiment 5). Specifically, we wanted to examine whether effects of physical human-likeness (i.e., human vs. robot appearance of the gazer) on mind perception ratings and behavioral human-likeness (i.e., reliable/predictive vs. random gaze behavior) on social attention [37] would interact in their effect on mind perception ratings and social attention when being presented in more lifelike interaction scenarios. We hypothesized that at a certain level of the paradigm's lifelikeness, both mind perception ratings and gaze cueing effects would be positively affected by physical and behavioral human-likeness, instead of just one of the two parameters. The specific hypotheses can be found below:

- **H1**: In line with previous studies, gaze-cueing effects are expected to be modulated by behavioral triggers of mind perception, such as predictable/reliable gaze behavior compared to random gaze behavior. However, with increasing levels of lifelikeness, we expect physical triggers of mind perception, such as humanlike compared to robot-like appearance of the gazer, to also affect gaze cueing.
- **H2**: In line with previous studies, mind perception ratings are expected to be modulated by physical triggers of mind perception, such as humanlike compared to robot-like appearance. However, with increasing levels of lifelikeness, we expect behavioral triggers of mind perception, such as predictable/reliable gaze behavior compared to random gaze behavior, to also affect mind perception ratings. Since the effect of behavioral cues on mind perception ratings can only take effect after the task, we calculated a pre-post interaction mind perception difference score and examined the effect of both physical and behavioral parameters on this difference score.

## 2 Methods and Materials

### 2.1 Experiments

Five experiments manipulated the physical and behavioral human-likeness of a gazing agent and examined the effects of these manipulations on mind perception and social attention in controlled (Experiments 1–4), and more lifelike (Experiment 5) settings. In the following section, we report the methods and materials that are common to all experiments and then report the specific variants of each experiment separately.

### 2.2 Participants

Participants were recruited from the undergraduate student pool at George Mason University and reimbursed via participation credits. All participants were at least 18 years old and reported normal or corrected to normal vision. The research complies with the APA's code of ethics and was approved by the local Ethics Committee at George Mason University. Participants provided informed consent prior to participation. 375 individuals were recruited for the five experiments (75 per experiment), and the data of 314 participants were included in the final analyses (for details on data rejection, please see the section of the respective experiment).

## 2.3 Stimuli

The target stimuli for the gaze-cueing procedure were black capital letters (F or T), measuring 0.8° in width and 1.3° in height; targets always appeared on the horizontal axis, and were located 6.0° from the center of the screen. The gazing stimuli varied in their degree of human-likeness, but differed between experiments and are described in the *Stimuli* section of the respective experiment.

## 2.4 Apparatus

Stimuli were presented at a distance of about 57 cm on an ASUS VB198T-P 19-inch monitor with a resolution of 1280 × 1024 pixels and a refresh rate of 60 Hz using Experiment Builder ([43]; in Experiment 1) or MATLAB (version R2015b; [44]) in combination with the Psychophysics Toolbox ([45]; in Experiments 2–5). Key press responses were recorded using a USB-connected standard keyboard.

## 2.5 Social Attention Task

Participants were asked to respond as fast and accurately as possible to the identity of target letters (F or T) that appeared either to the left or the right side of a centrally presented face (i.e., the gazer) by pressing one of two response keys ("D" and "K"; marked with stickers "F" and "T"). Prior to the target presentation, a centrally presented face changed its gaze direction (i.e., the gaze cue) to either the left or the right side of the screen, where the target subsequently either would (i.e., valid trial) or would not (i.e., invalid trial) appear. As soon as the target appeared, participants were asked to press the respective key so that reaction times and error rates could be recorded. To avoid spatial compatibility effects, the letter "F" was assigned to the "D" key and the letter "T" to the "K" key for 50% of the participants and vice versa for the other 50% of participants.

Each trial started with the presentation of a fixation cross in the center of the screen for a duration that was jittered between 700 and 1000 ms. Afterwards, the gazer appeared behind the fixation cross, and changed its gaze direction either towards the left or the right side of the screen after a jittered interval of 700–1000 ms. This gaze cue was followed by the presentation of the target letter either at the gazed-at location or opposite of the gazed-at location with a certain stimulus onset asynchrony (SOA), which varied between experiments (500 ms for Experiments 1–4; 1000 ms for Experiment 5). The gazer and target remained on the screen until a response was given or a timeout of 1200 ms was reached, whichever came first. The trial was concluded with the presentation of a blank screen for 680 ms (intertrial interval; ITI). See Fig. 1; for the trial sequences of Experiments 1–5.

For each experiment, physical human-likeness was manipulated within participants (robot vs. human; see Fig. 2), and cue reliability was altered between participants (50% vs. 80%). In the 50% reliability condition, 50% of targets were validly cued and 50% were invalidly cued by the agent, which appeared random. In the 80% reliability condition, 80% of targets were validly cued and 20% were invalidly cued, which appeared predictive.

## 2.6 Procedure

At the beginning of each experiment, participants were welcomed and seated in front of a computer screen. After providing informed consent, they were randomly assigned to either the 50% or 80% reliability condition and subsequently started the gaze cueing task. Participants were told to answer as quickly and as accurately as possible. Participants first completed a training block consisting of 20 trials, followed by an experimental block consisting of 320 trials (160 trials with the humanlike gazer and 160 trials with the robot-like gazer). The gazing stimulus in the training block differed from the agents used in the experimental block (i.e., mechanistic robot), and the order in which the human and the robot agent were presented during the experimental block was counterbalanced across participants. Participants were allowed to take a short break between blocks.

In order to obtain mind perception measures, participants were presented with images of the two gazers before and after the social attention task and asked to rate regarding their potential of having a mind (i.e., "Do you think this agent has a mind?") on 7-point scale (1: definitely not to 7: definitely yes). After completion of the post interaction agent rating, participants took a demographic survey. Each experiment took about 20–25 min to complete.

## 2.7 Analysis

Trials with incorrect answers and reaction times deviating more than 2 standard deviations from the individual mean were excluded from analysis. The gaze cueing effect was calculated for each block and each individual. To do so, the individual reaction time means of invalidly cued trials was subtracted from the individual reaction time means of validly cued trials of the respective block.

To analyze the influence of physical humanness and reliability on participants' gaze cueing effect, a 2 × 2 mixed ANOVA with the within-participants factor physical humanness (human, robot) and the between-participants factor reliability (50%, 80%) was conducted separately for each experiment. A 2 × 2 mixed ANOVA with the within-participants

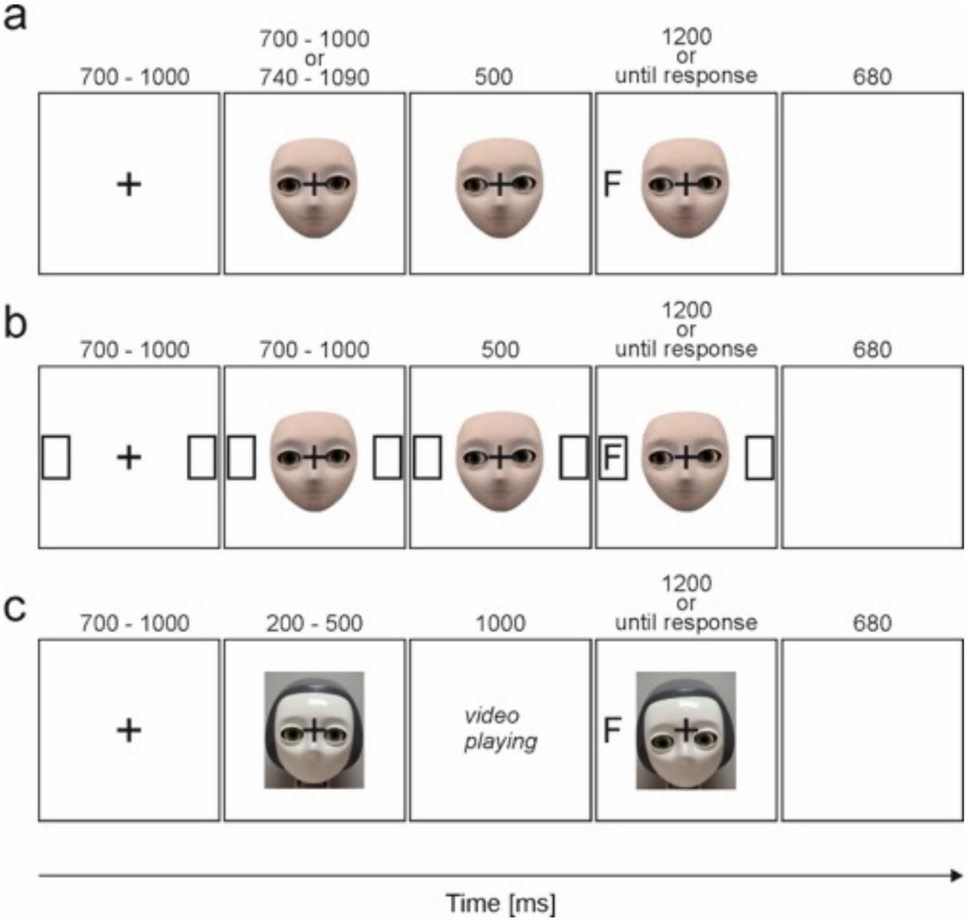

**Fig. 1** Gaze Cueing Paradigm: in all experiments, participants were to identify a target letter that was either validly or invalidly cued by an agent's gaze. In *Experiment 1*, *2* and *3*, the gaze cues consisted of a still image **a**. The time distribution of the straight gaze varied across experiments (see methods of respective experiment). In *Experiment 4*, the gaze cues consisted of a still image, but additionally, possible target locations are indicated with a black frame at the time of the gaze shift **b**. In *Experiment 5,* the gaze cues consisted of a video instead of a still image **c**

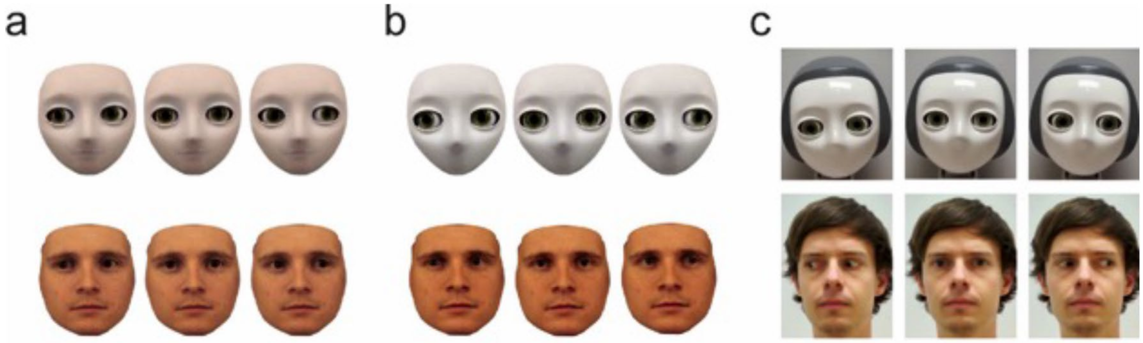

**Fig. 2** Gazing Stimuli: Agents used in *Experiments 1*, *3* and *4* are shown in **a**: the robot agent (top row) is a morphed image that consists of 20% human image and 80% robot image; the human agent (bottom row) is a morphed image that consists of 80% human image and 20% robot image. During the gaze cueing trials, the agents looked either to the left side of the screen (left), straight (middle) or to the right side of the screen (right). gazers **b**. *Experiment 2*, 100% robot (top row) and 100% human (bottom row) images were used as gazers. **c** In *Experiment 5*, videos of 100% robot and 100% human gazers were used instead of pictures. The images presented at the bottom depict the most eccentric gaze (left, right) and straight gaze (middle) shown in the videos

factor Physical Humanness (human, robot) and the between-participants factor Reliability (50%, 80%) was conducted to investigate the influence of physical humanness and reliability on the change in mind ratings of the respective agents (pre-post assessment). With regards to assumptions, it should be noted that (1) outliers had already been removed before conducting the ANOVA, (2) residuals were visually checked for violating normality assumptions, and (3) homogeneity of variance was tested using Levene's test. Residual distributions for all ANOVAs conducted showed no signs of skewness, although some showed signs of platycurtosis. We did not adjust for these signs because platycurtosis will increase the overall variance and thus bias the significance toward a less significant result [46]. The discussed significant results are thus not affected. Violations are reported in the results section of the respective experiment if applicable. In case of violations, we report a nonparametric analogue of the mixed ANOVA using the ezPerm R function (version 4.4-0) to confirm our results [47].

## 3 Experiment 1

In Experiment 1, morphing was used on a 100% human image and a 100% robot image to create one gazing stimulus with a high level of physical human-likeness (i.e., consisting of 80% of the human image and 20% of the robot image) and one gazing stimulus with a low level of physical human-likeness (i.e., consisting of 20% of the human image and 80% of the robot image). This manipulation was chosen to assure that familiarity with human versus robot faces did not bias the results. The reliability of the depicted gaze cues was either low (i.e., random or 50%) or high (i.e., predictive or 80%).

### 3.1 Participants

75 undergraduate students participated in the experiment. Ten participants were excluded due to poor task performance (i.e. answering incorrectly in more than 20% of the trials) or missing data, resulting in a final sample size of 65 participants (49 females; mean age: 20.3; range: 18–33; 56 right-handed). Participants were randomly assigned to either the 80% reliability condition (25 females; mean age: 21.03; range: 18–33; 28 right-handed) or the 50% reliability condition (24 females; mean age: 20; range: 18–28; 30 right-handed).

### 3.2 Stimuli

The human- and robot-like agent images were created by morphing the image of a human face (i.e., male face from the Karolinska Institute database; [48]) into the image of a robot face (i.e., Meka S2 robot head by Meka Robotics) in steps of 10% using the software FantaMorph 5.4.8 (Abrosoft). Out of this spectrum, the morph with 80% physical humanness was used as a humanlike gazer and the morph with 20% physical humanness as a robot-like gazer. The left-and rightward gazing face stimuli were created by shifting irises and pupils of the original 100% human and robot faces until they deviated 0.4° from direct gaze (with Photoshop), followed by another round of morphing as described above for each of the left- and the rightward gazing faces separately. As a last step, GIMP was used for all images to touch up any minor imperfections in the images and to make the sequencing of the images smooth. The face stimuli were 6.4° wide and 10.0° high on the screen, depicted on a white background and presented in full frontal orientation with eyes positioned on the central horizontal axis of the screen; see Fig. 2a.

### 3.3 Results

The mixed $2 \times 2$ ANOVA with gaze cueing effects as dependent variable revealed that *Reliability* ($F(1, 63) = 6.14$, $p = .016$, $\eta_G^2 = .05$), but not *Physical Humanness* ($F(1, 63) = .29$, $p = .593$, $\eta_G^2 < .01$) had a significant impact on social attention, such that gaze cueing effects were significantly larger for reliable than random gaze cues. The *Reliability* x *Physical Humanness* interaction was not significant ($F(1, 63) = .35$, $p = .559$, $\eta_G^2 < .01$); see Fig. 3a. The mixed $2 \times 2$ ANOVA with pre-post difference in mind perception ratings as a dependent variable revealed that *Physical Humanness* ($F(1, 63) = 24.91$, $p < .001$, $\eta_G^2 = .13$), but not *Reliability* ($F(1, 63) = 1.10$, $p = .298$, $\eta_G^2 = .01$) had a significant impact on mind ratings, such that mind ratings generally increased for the robot gazer but decreased for the human gazer after the gaze cueing task. The *Reliability* x *Physical Humanness* interaction did not reach significance ($F(1, 63) = .28$, $p = .600$, $\eta_G^2 < .01$); see Fig. 4a.

Gaze cueing variance between high versus low reliability groups was not equal for the robot level of physical human-likeness, as indicated by a Levene's test ($F(1, 63) = 5.97$, $p = 0.035$).[1] We therefore ran a nonparametric alternative for the mixed ANOVA with gaze cueing effects as dependent variable, which confirmed the significant main effect of *Reliability* ($p = .020$), as well as the insignificant effects of *Physical Humanness* and *Reliability* x *Physical Humanness* (both $p > .5$).

### 3.4 Discussion

The results of this experiment show that physical and behavioral parameters associated with human-likeness exert independent effects on mind perception ratings and social attention: physical human-likeness exclusively affected mind perception ratings, such that mind perception ratings for the robot agent increased after the gaze cueing task and decreased for the human agent, whereas cue reliability exclusively affected social attention, such that reliable gaze behavior induced larger gaze cueing effects than random

---

[1] The $p$ value has been adjusted using the Bonferroni procedure because two Levene's tests—one for each Physical Humanness level—have been conducted.

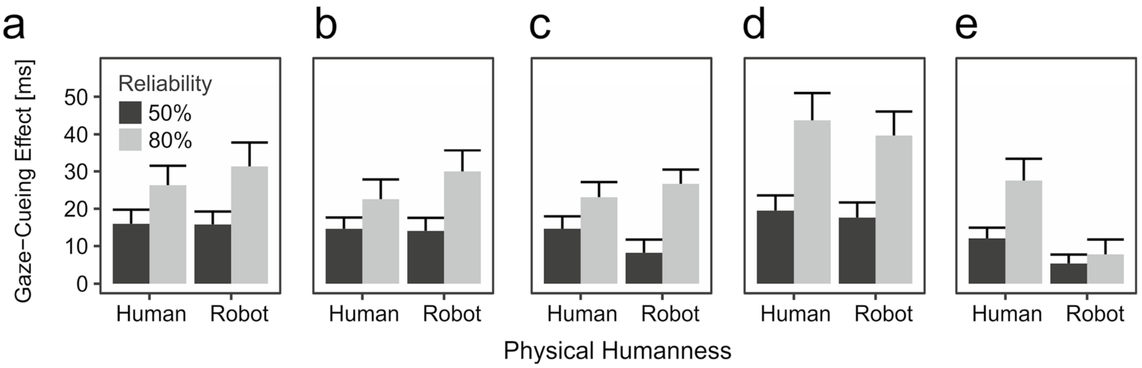

**Fig. 3** Gaze Cueing Effects as a function of physical (human vs. robot) and behavioral features (random vs. reliable): Patterns in gaze cueing were similar for *Experiment 1* (morphed images: 80% robot and 80% human; **a**), *Experiment 2* (original images: 100% robot and 100% human; **b**), *Experiment 3* (recorded human gaze behavior displayed on 80% robot and 80% human morph; **c**) and *Experiment 4* (spatial marker in periphery with 80% robot and 80% human morph; **d**): gaze cueing effects were affected by behavioral features, but not by physical features. In *Experiment 5* (videos of 100% robot and 100% human as gazing stimuli; **e**) an interaction effect between physical and behavioral features was found, such that gaze cueing effects were largest for videos of reliable human gazers and smallest for random robot gazers

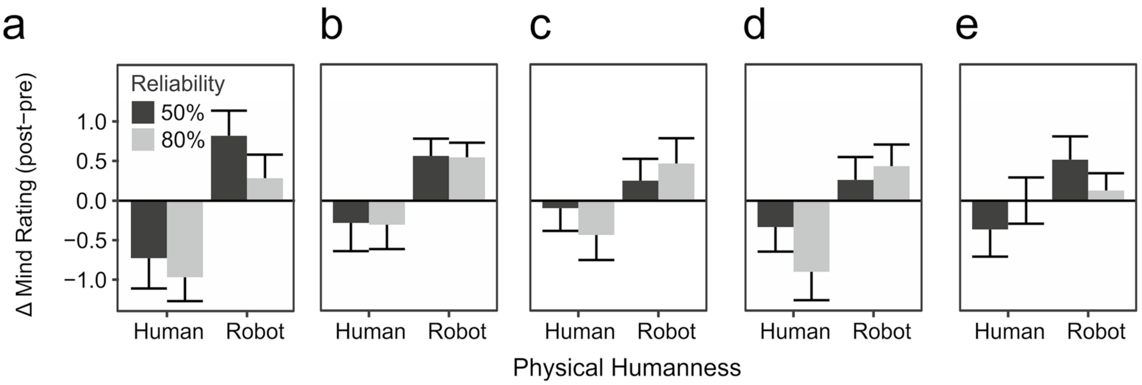

**Fig. 4** Changes in Mind Ratings (pre- vs. post-gaze cueing) as a function of physical (human vs. robot) and behavioral (random vs. reliable) features: Patterns in mind rating differences before and after interacting with the agents were comparable for *Experiment 1* (morphed images: 80% robot and 80% human; **a**), *Experiment 2* (original images: 100% robot and 100% human; **b**), *Experiment 3* (recorded human gaze behavior displayed on 80% robot and 80% human morph; **c**), *Experiment 4* (spatial marker in periphery with 80% robot and 80% human morph; **d**) and *Experiment 5* (videos of 100% robot and 100% human as gazers): mind ratings decreased for all agents with human appearance and increased for all agents with robot appearance; the gazer's reliability during the gaze cueing task did not have an impact on mind rating difference scores

gaze behavior. No interaction between the two parameters was observed in Experiment 1.

## 4 Experiment 2

In Experiment 2, the procedure of Experiment 1 was repeated with the 100% human and 100% robot image to assure that the results in Experiment 1 were not due to the morphed nature of the images, which could reduce their lifelikeness and induce feelings of discomfort associated with the 80% morph (as hypothesized by studies on the Uncanny Valley; see [49]); cue reliability was again set at 50% or 80%.

### 4.1 Participants

75 undergraduate students participated in the experiment. Eight participants were excluded due to poor task performance (i.e. answering incorrectly in more than 20% of the trials) and two due to missing data, resulting in a final sample size of 65 participants (50 females; mean age: 20.3; range: 18–29; 59 right handed). Participants were randomly assigned to either the 80% reliability condition (23 females; mean age: 20.3; range: 18–27; 29 right-handed) or the 50% reliability condition (27 females; mean age: 20.3; range: 18–29; 30 right-handed).

### 4.2 Stimuli

As gazing stimuli, the 100% human and 100% robot base images were used; Fig. 2b.

### 4.3 Results

The mixed $2 \times 2$ ANOVA with gaze cueing effects as dependent variable revealed that *Reliability* ($F(1, 63) = 4.64$, $p = .035$, $\eta_G^2 = .05$), but not *Physical Humanness* ($F(1, 63) = 1.12$, $p = .293$, $\eta_G^2 < .01$) had a significant impact on gaze cueing effects. The *Reliability* x *Physical Humanness* interaction did not reach significance ($F(1, 63) = 1.48$, $p = .229$, $\eta_G^2 < .01$); see Fig. 3b. The mixed $2 \times 2$ ANOVA with pre-post difference in mind perception ratings as a dependent variable revealed that *Physical Humanness* ($F(1, 63) = 8.41$, $p = .005$, $\eta_G^2 = .07$), but not *Reliability* ($F(1, 63) < .01$, $p = .940$, $\eta_G^2 < .01$) had a significant impact on mind perception ratings. The *Reliability* x *Physical Humanness* interaction did not reach significance ($F(1, 63) < .01$, $p = .994$, $\eta_G^2 < .01$); see Fig. 4b.

### 4.4 Discussion

The results of Experiment 2 replicate the findings of Experiment 1, showing that mind perception ratings are exclusively influenced by physical human-likeness and gaze cueing effects are exclusively influenced by behavioral human-likeness. The results also show that the lifelikeness of the gazing stimuli themselves did not impact the results, since the same pattern of results was observed for morphed (i.e., 80% and 20% humanlike morphs; Experiment 1) and realistic (i.e., 100% human and robot images; Experiment 2) images.

## 5 Experiment 3

The goal of Experiment 3 was to examine whether changing the lifelikeness of a gazer's eye movements would modulate the previously reported findings. In order to do so, we recorded eye movement patterns from a human volunteer pretending to take the role of the gazer in the gaze cueing task using an eye tracker and replayed the timing of the eye movements on the gazing stimulus during the experiment.[2] Cue reliability was again set at 50% or 80%.

### 5.1 Participants

75 undergraduate students participated in the experiment. 7 participants were excluded due to poor task performance (i.e., answering incorrectly in more than 20% of the trials) and 6 participants were excluded due to missing data (e.g., because participants used the wrong keys), resulting in a final sample size of 62 participants (46 females; mean age: 20.2; range: 18–38; 57 right handed). Participants were randomly assigned to either the 80% reliability condition (22 females; mean age: 19.4; range: 18–38; 29 right-handed) or the 50% reliability condition (20 females; mean age: 21.0; range: 18–25; 28 right-handed).

### 5.2 Stimuli

The agent images were identical to the ones used in Experiment 1; see Fig. 2a.

### 5.3 Trial Sequence

The trial sequence was identical to Experiment 1, with one exception: the time the agent took from looking straight to looking to the side of the screen was not drawn from a uniform distribution but from a mean-adjusted distribution collected from a human volunteer (the first author of this paper: AA). The distribution was obtained using a MATLAB script that recorded the time needed to shift the gaze from a central fixation cross towards a laterally presented target letter using an EyeLink 1000 eye-tracker [52] sampling at 1000 Hz. 320 trials were collected to mirror the distribution needed for the 320 trials in the experiment. On a descriptive level, the distribution was more similar to a normal distribution than to a uniform distribution (as was the case in Experiment 1 and 2). After centering the distribution on the mean of the uniform distribution used for the robot agent, i.e. on 850 ms, values ranged from 750 to 1090 ms. The gaze response latencies used for the experiment were drawn from this mean-adjusted "human" gaze response distribution and can be inspected in Fig. S1. The trial sequence is depicted in Fig. 1a.

### 5.4 Results

The mixed $2 \times 2$ ANOVA with gaze cueing effects as a dependent variable revealed that *Reliability* ($F(1, 60) = 10.15$, $p = .002$, $\eta_G^2 = .10$), but not *Physical Humanness* ($F(1, 60) = .27$, $p = .603$, $\eta_G^2 < .01$) had a significant impact on gaze cueing effects; the *Reliability* x *Physical Humanness* interaction did not reach significance ($F(1, 60) = 2.70$, $p = .106$, $\eta_G^2 = .02$); see Fig. 3c. The mixed 2 x 2 ANOVA with pre-post differences in mind perception ratings as a dependent variable revealed that *Physical Humanness* ($F(1, 60) = 5.55$, $p = 0.022$, $\eta_G^2 = .03$), but not *Reliability* ($F$(1,

---

[2] This manipulation was chosen based on previous research that has shown that people are highly sensitive in differentiating biological from non-biological motion [50, 51].

$60) = .03$, $p = .855$, $\eta^2_G < .01$) had a significant impact on mind ratings; the *Reliability* x *Physical Humanness* interaction did not reach significance ($F(1, 60) = 1.14$, $p = .290$, $\eta^2_G < .01$); see Fig. 4c.

Gaze cueing variance between high and low reliability groups was not equal for the robot level of physical human-likeness as indicated by a Levene's test ($F(1, 55) = 5.61$, $p = 0.042$).[3] We therefore ran a nonparametric alternative for the mixed ANOVA on gaze cueing effects, which confirmed with a main effect of *Reliability* ($p = .028$), as well as the insignificance of the main effect of *Physical Humanness* and the interaction term (both $p > .2$).

### 5.5 Discussion

The results of Experiment 3 replicate the findings of Experiments 1 and 2, again showing an isolated effect of physical human-likeness on mind ratings and behavioral human-likeness on gaze cueing effects, indicating that the lifelikeness of the observed eye movements does not significantly impact the pattern of results.

## 6 Experiment 4

The goal of Experiment 4 was to examine whether the lifelikeness of the context in which a social exchange takes place potentially modulates previous findings. One known issue with the gaze cueing paradigm that could reduce the perceived lifelikeness of the interaction is that changes in gaze direction are not tied to changes in the environment but are directed at empty space where subsequently a target appears (on valid trials) or not (on invalid trials). In reality, however, changes in gaze direction usually occur *in response* to a triggering event, for instance a loud sound or the appearance of a person or an object. To increase the lifelikeness of the interaction, we added abstract objects in the environment that were already present at the time when the face changed its gaze direction and could serve as spatial markers to which the gaze cue could refer (and which became the location at which the targets appeared later). Cue reliability was again set at 50% or 80%.

### 6.1 Participants

75 undergraduate students participated in the experiment. 12 participants were excluded due to poor task performance (i.e. answering incorrectly in more than 20% of the trials) and 6 because of technical issues (e.g., pressing the wrong response keys), resulting in a final sample size of 57 participants (46 females; mean age: 20.1; range: 18–29; 50 right handed). Participants were randomly assigned to either the 80% reliability condition (24 females; mean age: 20.1; range: 18–29; 27 right-handed) or the 50% reliability condition (22 females; mean age: 20.1; range: 18–29; 23 right handed).

### 6.2 Stimuli

The agent images were identical to the ones used in Experiment 1; see Fig. 2a.

### 6.3 Trial Sequence

The trial sequence was identical to experiment one with one exception: when shifting its gaze, the agent did not look towards empty space but towards a placeholder that indicated the two locations at which the target could subsequently appear. The frames appeared together with the fixation cross at the beginning of each trial and disappeared during the ITI. The trial sequence is depicted in Fig. 1b.

### 6.4 Results

The mixed $2 \times 2$ ANOVA with gaze cueing effects as a dependent variable revealed that *Reliability* ($F(1, 55) = 10.59$, $p = .002$, $\eta^2_G = .13$), but not *Physical Humanness* ($F(1, 55) = .57$, $p = .453$, $\eta^2_G < .01$) had a significant impact on gaze cueing effects; the *Reliability* x *Physical Humanness* interaction did not reach significance ($F(1, 55) = .08$, $p = .784$, $\eta^2_G < .01$); see Fig. 3d. The mixed 2 x 2 ANOVA with pre-post difference in mind perception ratings as a dependent variable revealed that *Physical Humanness* ($F(1, 55) = 13.93$, $p < .001$, $\eta^2_G = .08$), but not *Reliability* ($F(1, 55) = .31$, $p = .582$, $\eta^2_G < .01$) had a significant impact on mind ratings; the Reliability x *Physical Humanness* interaction did not reach significance ($F(1, 55) = 1.97$, $p = .17$, $\eta^2_G = .01$); see Fig. 4d.

### 6.5 Discussion

The results of Experiment 4 replicate the findings of experiments 1-3, again showing an isolated effect of physical human-likeness on mind ratings and behavioral human-likeness on gaze cueing effects, indicating that the lifelikeness of the context in which a social exchange take places does not significantly impact the pattern of previous results.

---

[3] The $p$-value has been adjusted using the Bonferroni procedure because two Levene's tests—one for each Physical Humanness level—have been conducted.

## 7 Experiment 5

Experiments 2–4 showed that increasing lifelikeness of the interaction paradigm by using stimuli that are physically realistic, that move their eyes with humanlike timing or whose gaze cues refer to objects in visual space in a meaningful way was not impactful enough to change the pattern of results. In all previous experiments, observers interacted with static images of human or humanlike gazers, which is very unlike lifelike social interactions with other humans. To increase the perceived lifelikeness of the social attention task as a whole, we used video recordings of a human and a robot agent as gazing stimuli instead of static images. Cue reliability was again set at 50% or 80%.

### 7.1 Participants

75 undergraduate students participated in the experiment. Eight participants were excluded due to poor task performance (i.e., answering incorrectly in more than 20% of the trials) and two due to missing data, resulting in a final sample size of 65 participants (51 females; mean age: 19.9; range: 18–30; 58 right handed). Participants were randomly assigned to either the 80% reliability condition (26 females; mean age: 19.9; range: 18–30; 30 right-handed) or the 50% reliability condition (25 females; mean age: 19.8; range: 18–25; 28 right-handed).

### 7.2 Stimuli

Video sequences simulating gaze cues of a human and a robot agent were recorded: for the robot condition, cues to the left and right were recorded from the humanoid Meka S2 robot head; for the human condition, cues to the left and right were recorded from a human, the second author PPW. All videos were cut such that the first frame showed the gazing agents with straight gaze (Fig. 2c, middle), the gaze shift was completed within 1000 ms and the last frame's gaze was of maximal eccentricity (Fig. 2c, left and right). On top of the gaze cues, both human and robot videos included head cues of comparable strength.

### 7.3 Trial Sequence

The trial sequence was kept as similar as possible to Experiment 1. Each trial started with the presentation of a fixation cross at the center of the screen for a duration drawn from values uniformly distributed between 700 and 1000 ms. Afterwards, the agent as appearing in the first frame of the respective video, appeared behind the fixation cross for a duration drawn from values uniformly distributed between 200 and 500 ms. Subsequently, the video was being played for 1000 ms during which the agent changed its gaze towards either the left or the right side of the screen, thereby either validly or invalidly cueing the location of the subsequently presented target letter. When the video finished playing, the last frame froze and the target letter was presented at the left or the right side of the screen. The last frame and the target remained on the screen until a response was given or 1200 ms had passed. The trial was concluded with a blank screen presented for 680 ms. The trial sequence is depicted in Fig. 1c.

### 7.4 Results

In contrast to previous experiments, the mixed 2 × 2 ANOVA with gaze cueing effects as a dependent variable revealed that *Reliability* ($F(1, 63) = 4.71$, $p = .034$, $\eta_G^2 = .04$) and *Physical Humanness* ($F(1, 63) = 12.05$, $p < .001$, $\eta_G^2 = .08$) had a significant impact on gaze cueing effects; the *Reliability* x *Physical Humanness* interaction was trending towards significance but did not reach significance ($F(1, 63) = 2.96$, $p = .090$, $\eta_G^2 = .02$); see Fig. 3e. Again in contrast to previous findings, the mixed 2 × 2 ANOVA with pre-post differences in mind perception ratings as a dependent variable revealed that neither *Physical Humanness* ($F(1, 63) = 2.37$, $p = .129$, $\eta_G^2 = .02$) nor *Reliability* ($F(1, 63) < .01$, $p = .958$, $\eta_G^2 < .01$) had a significant impact on mind ratings. the *Reliability* x *Physical Humanness* interaction did not reach significance ($F(1, 63) = 1.31$, $p = .257$, $\eta_G^2 = .01$); see Fig. 4e.

Gaze cueing variance between high and low reliability groups was not equal for both levels of physical humanlikeness (human and robot) as indicated by Levene's tests (Human: $F(1, 63) = 6.61$, $p = 0.025$; Robot: $F(1, 63) = 5.41$, $p = 0.047$).[4] We therefore ran a nonparametric alternative for the mixed ANOVA with gaze cueing effects as a dependent variable, which confirmed the main effect of *Reliability* ($p = .030$) and *Physical Humanness* ($p < .001$), as well as a trend for the interaction term ($p = 0.105$).

### 7.5 Discussion

The results of Experiment 5 show that changing the lifelikeness of the interaction scenario as a whole by using dynamic videos instead of static images changes the pattern of results such that physical and behavioral markers of human-likeness now both affect gaze cueing effects independently, with larger cueing effects for the human versus robot gazer, as well as the reliable versus random gaze cues (with no interaction effects between the two components). In contrast,

---

[4] The *p*-values have been adjusted using the Bonferroni procedure because two Levene's tests—one for each Physical Humanness level—have been conducted.

physical human-likeness does not significantly impact pre-post interaction changes in mind perception anymore. The implications of these findings are discussed below.

## 8 General Discussion

This study aimed to investigate how factors that, independent of each other, have been related to mind perception, such as physical human-likeness and predictable behavior, affect mind perception ratings and social attention mechanisms as a function of the interaction's lifelikeness. For that purpose, we manipulated physical, behavioral and contextual parameters that were thought to manipulate the lifelikeness of a social interaction scenario. In Experiment 1, which constituted the baseline, we looked at the influence of physical appearance (human morph vs. robot morph) and gaze predictivity on mind perception ratings and gaze-cueing effects without specifically manipulating lifelikeness. In Experiment 2, the lifelikeness of the gazer was manipulated by using a 100% human face and a 100% robot face as opposed to morphed images. Experiment 3 manipulated the lifelikeness of the gaze signal by modeling the onset of the gaze cues after a real human's cue onsets, thereby incorporating biological eye movements (i.e., right and left gaze changes) into the paradigm. Experiment 4 manipulated the lifelikeness of the context by adding reference objects (i.e., place holders) to the gaze cueing paradigm that were already present at the time of the gaze change, as gaze changes in real life usually are targeted at reference objects in the environment and not at empty space (like in traditional gaze cueing paradigms). Experiments 1–4 revealed similar results, such that the behavioral component (i.e., reliability of the gaze cue) affected social attention but not mind perception, whereas the physical component (i.e., appearance of the gazer) affected mind perception but not social attention.[5] Only when the lifelikeness of the overall interaction was changed by using videos of an actual human and an actual robot first engaging in mutual gaze and then performing gaze cues, the pattern of results changed: both gaze reliability and physical appearance now had an influence on social attention, such that gaze cueing effects were larger for human versus robot gazers and reliable versus random gaze behaviors; in contrast, pre-post mind perception ratings were neither affected by physical appearance nor by gaze reliability.

The experiments outline two important findings with regard to the effects of physical, behavioral and contextual effects on mind perception and social attention: First, behavioral features, such as the reliability of gaze signals, robustly modulated social attention across experiments, whereas physical appearance only had an effect when the interaction seemed sufficiently lifelike (through the use of video sequences). This replicates findings from previous studies showing that even very basic social-cognitive processes like gaze cueing can be top-down modulated by social context information [53], and highlights that certain top-down modulators, such as the physical appearance of an agent, might only exert their effect in relatively lifelike interactions. This observation also provides some clarity regarding the ongoing debate in the literature whether manipulations related to mind perception and/or mentalizing have an effect on social attention [54] or not [55]. The current study suggests that there is an interaction between top-down and bottom-up mechanisms influencing social attention, but that the top-down component might only take effect in sufficiently realistic paradigms (see also [56]). Although the current study does not maximize lifelikeness to the same extent as other studies where the gaze cues are sent by a real human actor sitting opposite of the participant (e.g., [57]), it indicates that a certain level of lifelikeness needs to be reached before various context factors start modulating social attention. Where exactly this level is located and whether different context factors require different levels of lifelikeness should be the focus of future studies.

Second, physical agent features, such as the human-likeness of a gazer's appearance, modulated pre-post mind perception changes in more controlled versions of the paradigm (Experiments 1–4) but not under relatively lifelike interaction conditions using videos (Experiment 5); behavioral parameters, such as the reliability of gaze cues, never modulated mind perception ratings. One explanation as to why reliability did not modulate mind perception ratings could be that completing the gaze-cueing task with very reliable agents diminished participants' need for anthropomorphizing nonhuman agents, which resulted in mind ratings that were not different from those for agents whose gaze behavior was random. In other words, maybe a certain level of uncertainty is needed in order to strongly trigger mind perception. This interpretation is supported by prior work suggesting that agents displaying very predictable actions, decrease our need to understand their behaviors, and consequently trigger less anthropomorphizing/mind perception [58].

The current study is consistent with previous literature illustrating the importance of using ecologically valid paradigms when investigating social cognition [59]. While prior work shows that robots may not be able to reflexively shift human attention in computer-based paradigms [20], face-to face gaze-cueing paradigms using real robots as gazers illustrate that robots can in fact reflexively shift human attention (like human gazers) when the surrounding is sufficiently

---

[5] Interestingly, the results of Experiment 4 show a descriptive difference such that gaze-cueing effects were overall larger increase. This is not surprising as previous studies have shown that including contextual information has a positive effect on gaze-cueing effects [17].

lifelike [15]. Prior literature also shows that different brain regions are activated during social attention depending on whether highly controlled, offline paradigms or face-to-face, online paradigms are employed, that is: traditional fMRI studies identify brain regions in the right hemifield (e.g., STS, ACC, TPJ) as important neural correlates of social attention [60], whereas studies that use dynamic face-to-face paradigms implicate similar structures in the left hemifield [61], suggesting that some social-cognitive processes may not be sufficiently activated in highly controlled experiments (see [59]; for detailed arguments for the necessity to examine social cognition "online"). Other studies using VR-based paradigms showed that joint attention not only consists of directing others' attention to important objects or events in the environment (i.e., other-representations) but also requires another essential mechanism, that is, engaging in mutual gaze to signal the readiness for joint attention (i.e., self-representations) [59, 62, 63] —an insight that traditional gaze cueing paradigms were unable to uncover. Consistent with these observations, the current study shows that the effect of physical and behavioral parameters on social attention might change depending on the lifelikeness of the paradigm. Although "online" social cognition paradigms are more challenging to design and implement than "offline" paradigms (e.g., additional programming requirements, access to embodied robot platforms, more involved study approval processes), it is important to examine social cognitive processes in settings that are similar enough to real interactions in order to draw firm conclusions regarding the impact of potential modulating factors. Future studies should increase the lifelikeness of social attention paradigms in HRI even more, for instance by using embodied robot platforms instead of video recordings; see [64–67].

In conclusion, this study illustrates the importance of using methods to mimic real-life gaze interaction in investigating social gaze whenever possible. This is of the upmost relevance for social roboticists since the goal is to design social robots that are equipped with means to display social human behaviors and evoke both natural and intuitive reactions from the humans that interact with these robots.

**Acknowledgements** We would like to acknowledge the hard-working research assistants that helped collect our sample.

**Author's contribution** AA and EW conceptualized the study. PW programmed the experiments. AA and PW collected and analyzed the data. AA, EW, and PW interpreted the results and wrote the manuscript.

**Funding** This study was not funded by any grants.

## Compliance with Ethical Standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. Adolphs R (1999) Social cognition and the human brain. Trends Cogn Sci 3(12):469–479. https://doi.org/10.1016/s1364-6613(99)01399-6
2. Emery NJ (2000) The eyes have it: the neuroethology, function and evolution of social gaze. Neurosci Biobehav Rev 24(6):581–604. https://doi.org/10.1016/S0149-7634(00)00025-7
3. Gallagher HL, Frith CD (2003) Functional imaging of 'theory of mind'. Trends Cogn Sci 7(2):77–83. https://doi.org/10.1016/S1364-6613(02)00025-6
4. Gray HM, Gray K, Wegner DM (2007) Dimensions of mind perception. Science 315(5812):619. https://doi.org/10.1126/science.1134475
5. Waytz A, Gray K, Epley N, Wegner DM (2010) Causes and consequences of mind perception. Trends Cogn Sci 14(8):383–388. https://doi.org/10.1016/j.tics.2010.05.006
6. Wiese E, Metta G, Wykowska A (2017) Robots as intentional agents: using neuroscientific methods to make robots appear more social. Front Psychol 8:1663. https://doi.org/10.3389/fpsyg.2017.01663
7. Friesen CK, Kingstone A (1998) The eyes have it! Reflexive orienting is triggered by non-predictive gaze. Psychon Bull Rev 5(3):490–495. https://doi.org/10.3758/BF03208827
8. Özdem C, Wiese E, Wykowska A, Müller H, Brass M, Van Overwalle F (2016) Believing androids—fMRI activation in the right temporo-parietal junction is modulated by ascribing intentions to non-human agents. Soc Neurosci 12(5):582–593. https://doi.org/10.1080/17470919.2016.1207702
9. Teufel C, Fletcher PC, Davis G (2010) Seeing other minds: attributed mental states influence perception. Trends Cogn Sci 14(8):376–382. https://doi.org/10.1016/j.tics.2010.05.005
10. Wiese E, Wykowska A, Zwickel J, Müller HJ (2012) I see what you mean: how attentional selection is shaped by ascribing intentions to others. PLoS ONE 7(9):e45391. https://doi.org/10.1371/journal.pone.0045391
11. Wiese E, Buzzell GA, Abubshait A, Beatty PJ (2018) Seeing minds in others: mind perception modulates low-level social-cognitive performance and relates to ventromedial prefrontal structures. Cogn Affect Behav Neurosci 18(5):837–856. https://doi.org/10.3758/s13415-018-0608-2
12. Wykowska A, Wiese E, Prosser A, Müller HJ (2014) Beliefs about the minds of others influence how we process sensory information. PLoS ONE 9(4):e94339. https://doi.org/10.1371/journal.pone.0094339
13. Nummenmaa L, Calder AJ (2009) Neural mechanisms of social attention. Trends Cogn Sci 13(3):135–143. https://doi.org/10.1016/j.tics.2008.12.006
14. Pfeiffer UJ, 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
15. Kompatsiari K, Ciardo F, Tikhanoff V, Metta G, Wykowska A (2018) On the role of eye contact in gaze cueing. Sci Rep 8(1):17842. https://doi.org/10.1038/s41598-018-36136-2
16. Perez-Osorio J, Müller HJ, Wiese E, Wykowska A (2015) Gaze following is modulated by expectations regarding others' action goals. PLoS ONE 10(11):e0143614. https://doi.org/10.1371/journal.pone.0143614
17. Wiese E, Zwickel J, Müller HJ (2013) The importance of context information for the spatial specificity of gaze cueing. Atten Percept Psychophys 75(5):967–982. https://doi.org/10.3758/s13414-013-0444-y
18. Caruana N, de Lissa P, McArthur G (2017) Beliefs about human agency influence the neural processing of gaze

during joint attention. Soc Neurosci 12(2):194–206. https://doi.org/10.1080/17470919.2016.1160953

19. Deaner RO, Shepherd SV, Platt ML (2007) Familiarity accentuates gaze cuing in women but not men. Biol Lett 3(1):64–67. https://doi.org/10.1098/rsbl.2006.0564
20. Admoni H, Bank C, Tan J, Toneva M, Scassellati B (2011) Robot gaze does not reflexively cue human attention. In: Proceedings of the 33rd annual conference of the cognitive science society, 1983–1988
21. Admoni H, Scassellati B (2017) Social eye gaze in human-robot interaction: a review. J. Hum. Robot Interact. 6(1):25. https://doi.org/10.5898/JHRI.6.1.Admoni
22. DiSalvo CF, Gemperle F, Forlizzi J, Kiesler S (2002) All robots are not created equal: the design and perception of humanoid robot heads. In: Conference on designing interactive systems processes practices methods and techniques, pp 321–326. https://doi.org/10.1145/778712.778756
23. Kiesler S, Powers A, Fussell SR, Torrey C (2008) Anthropomorphic interactions with a robot and robot–like agent. Soc Cogn 26(2):169–181. https://doi.org/10.1521/soco.2008.26.2.169
24. Looser CE, Wheatley T (2010) The tipping point of animacy: how, when, and where we perceive life in a face. Psychol Sci 21(12):1854–1862. https://doi.org/10.1177/0956797610388044
25. Tung F (2011) Influence of gender and age on the attitudes of children towards humanoid robots. Hum Comput Interact IV. https://doi.org/10.1007/978-3-642-21619-0_76
26. Martini M, Buzzell G, Wiese E (2015) Agent appearance modulates mind attribution and social attention in human-robot interaction. Soc Robot 1:431–439. https://doi.org/10.1007/978-3-319-25554-5
27. Wiese E, Weis PP (2020) It matters to me if you are human—examining categorical perception in human and nonhuman agents. Int J Hum Comput Stud 133:1–12. https://doi.org/10.1016/j.ijhcs.2019.08.002
28. Wiese E, Wykowska A, Müller HJ (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
29. Morewedge CK (2009) Negativity bias in attribution of external agency. J Exp Psychol Gen 138(4):535–545. https://doi.org/10.1037/a0016796
30. Morewedge CK, Preston J, Wegner DM (2007) Timescale bias in the attribution of mind. J Pers Soc Psychol 93(1):1–11. https://doi.org/10.1037/0022-3514.93.1.1
31. Abell F, Happe F, Frith U (2000) Do triangles play tricks? attribution of mental states to animated shapes in normal and abnormal development. Cogn Dev 15(1):1–16. https://doi.org/10.1016/S0885-2014(00)00014-9
32. Castelli F, Happé F, Frith U, Frith C (2013) Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. In: Social neuroscience: key readings, pp 155–170. https://doi.org/10.4324/9780203496190
33. Heider F, Simmel M (1944) An experimental study of apparent behavior. Am J Psychol 57(2):243–259. https://doi.org/10.1017/CBO9781107415324.004
34. Short E, Hart J, Vu M, Scassellati B (2010) No fair!! an interaction with a cheating robot. In 2010 5th ACM/IEEE international conference on human–robot interaction (HRI), pp 219–226. https://doi.org/10.1109/HRI.2010.5453193
35. Ciardo F, Beyer F, De Tommaso D, Wykowska A (2020) Attribution of intentional agency towards robots reduces one's own sense of agency. Cognition 194:104109. https://doi.org/10.1016/j.cognition.2019.104109
36. Gray K, Knobe J, Sheskin M, Bloom P, Barrett LF (2011) More than a body: mind perception and the nature of objectification. J Pers Soc Psychol 101(6):1207–1220. https://doi.org/10.1037/a0025883
37. Abubshait A, Wiese E (2017) You look human, but act like a machine: agent appearance and behavior modulate different aspects of human–robot interaction. Front. Psychol. 8:1393. https://doi.org/10.3389/fpsyg.2017.01393
38. Mandell AR, Smith MA, Martini MC, Shaw TH, Wiese E (2015) Does the presence of social agents improve cognitive performance on a vigilance task? Int. Conf. Soc. Robot. 1:421–430
39. Admoni H, Dragan A, Srinivasa SS, Scassellati B (2014) Deliberate delays during robot-to-human handovers improve compliance with gaze communication. In: Proceedings of the 2014 ACM/IEEE international conference on human–robot interaction - HRI'14, pp 49–56. https://doi.org/10.1145/2559636.2559682
40. Hungr CJ, Hunt AR (2012) Physical self-similarity enhances the gaze-cueing effect. Q J Exp Psychol 65(7):1250–1259. https://doi.org/10.1080/17470218.2012.690769
41. Waytz A, Heafner J, Epley N (2014) The mind in the machine: anthropomorphism increases trust in an autonomous vehicle. J Exp Soc Psychol 52:113–117. https://doi.org/10.1016/j.jesp.2014.01.005
42. Saygin AP, 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. SCAN 7:413–422. https://doi.org/10.1093/scan/nsr025
43. SR Research (2004) Experiment builder
44. The Mathworks Inc (2015) Matlab
45. Brainard DH (1997) The psychophysics toolbox. Spat Vis 10(4):433–436. https://doi.org/10.1163/156856897X00357
46. Field AP (2005) Discovering statistics using SPSS: (and sex, drugs and rock'n'roll). SAGE, Thousand Oaks
47. Feys J (2016) Nonparametric tests for the interaction in two-way factorial designs using R. R J 8(1):367. https://doi.org/10.32614/RJ-2016-027
48. Lundqvist D, Flykt A, Ohman A (1988) Karolinska directed emotional faces. In: Psychology section. Department of Clinical Neuroscience, Karolinska Hospital, S-171, 76
49. Kätsyri J, Förger K, Mäkäräinen M, Takala T (2015) A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. Front Psychol 6(MAR):1–16. https://doi.org/10.3389/fpsyg.2015.00390
50. Giese MA, Poggio T (2003) Neural mechanisms for the recognition of biological movements. Nat Rev Neurosci 4(3):179–192. https://doi.org/10.1038/nrn1057
51. Grossman ED, Blake R (2002) Brain areas active during visual perception of biological motion. Neuron 35:9
52. SR Research (2010) Eyelink 1000
53. Ristic J, Kingstone A (2005) Taking control of reflexive social attention. Cognition 94(3):B55–B65. https://doi.org/10.1016/j.cognition.2004.04.005
54. Gobel MS, Tufft MRA, Richardson DC (2017) Social beliefs and visual attention: how the social relevance of a cue influences spatial orienting. Cogn Sci 42:161–185. https://doi.org/10.1111/cogs.12529
55. Kingstone A, Kachkovski G, Vasilyev D, Kuk M, Welsh TN (2019) Mental attribution is not sufficient or necessary to trigger attentional orienting to gaze. Cognition 189:35–40. https://doi.org/10.1016/j.cognition.2019.03.010
56. Capozzi F, Ristic J (2020) Attention and mentalizing? reframing a debate on social orienting of attention. Vis Cogn 28:97–105. https://doi.org/10.1080/13506285.2020.1725206
57. Cole GG, Skarratt PA, Kuhn G (2016) Real person interaction in visual attention research. Eur Psychol 21(2):141–149. https://doi.org/10.1027/1016-9040/a000243

58. Waytz A, Morewedge CK, Epley N, Monteleone G, Gao J-H, Cacioppo JT (2010) Making sense by making sentient: effectance motivation increases anthropomorphism. J Pers Soc Psychol 99(3):410–435. https://doi.org/10.1037/a0020240
59. Schilbach L, Timmermans B, Reddy V, Costall A, Bente G, Schlicht T, Vogeley K (2013) Toward a second-person neuroscience. Behav Brain Sci 36(4):393–414. https://doi.org/10.1017/S0140525X12000660
60. Redcay E, Kleiner M, Saxe R (2012) Look at this: the neural correlates of initiating and responding to bids for joint attention. Front Hum Neurosci 6:169. https://doi.org/10.3389/fnhum.2012.00169
61. Lachat F, Conty L, Hugueville L, George N (2012) Gaze cueing effect in a face-to-face situation. J Nonverbal Behav 36(3):177–190. https://doi.org/10.1007/s10919-012-0133-x
62. Caruana N, Brock J, Woolgar A (2015) A frontotemporoparietal network common to initiating and responding to joint attention bids. NeuroImage 108:34–46. https://doi.org/10.1016/j.neuroimage.2014.12.041
63. Schilbach L, Wilms M, Eickhoff SB, Romanzetti S, Tepest R, Bente G, Shah NJ, Fink GR, Vogeley K (2010) Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. J Cogn Neurosci 22(12):2702–2715. https://doi.org/10.1162/jocn.2009.21401
64. Thellman S, Silvervarg A, Gulz A, Ziemke T (2016) Physical vs. virtual agent embodiment and effects on social interaction. In: Traum D, Swartout W, Khooshabeh P, Kopp S, Scherer S, Leuski A (eds) Intelligent virtual agents, vol 10011. Springer, Berlin, pp 412–415. https://doi.org/10.1007/978-3-319-47665-0_44
65. Mollahosseini A, Abdollahi H, Sweeny TD, Cole R, Mahoor MH (2018) Role of embodiment and presence in human perception of robots' facial cues. Int J Hum Comput Stud 116:25–39. https://doi.org/10.1016/j.ijhcs.2018.04.005
66. Wainer J, Feil-Seifer DJ, Shell DA, Mataric MJ (2007) Embodiment and human–robot interaction: a task-based perspective. In: RO-MAN 2007—the 16th IEEE international symposium on robot and human interactive communication, pp 872–877. https://doi.org/10.1109/ROMAN.2007.4415207
67. Lee KM, Jung Y, Kim J, Kim SR (2006) Are physically embodied social agents better than disembodied social agents?: the effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction. Int J Hum Comput Stud 64(10):962–973. https://doi.org/10.1016/j.ijhcs.2006.05.002

**Abdulaziz Abubshait** is a postdoc at the Italian Institute of Technology in Genova, Italy. He received his PhD in Human Factors and Applied Cognition from George Mason University in 2019. His research interests investigates the dynamics of human-robot social interactions.

**Patrick Weis** is a postdoc at the Nicolaus Copernicus University in Torun. In 2019, he received his PhD in Human Factors and Applied Cognition at George Mason University. He also received an MS in Neuroscience from the University of Tubingen in 2014.

**Eva Wiese** is an Associate Professor in Human Factors and Applied Cognition and the head of the Social and Cognitive Interactions Lab at George Mason University. She has a PhD in Neuroscience from Ludwig Maximilian University Munich and a MS in Psychology from Otto-Friedrich University Bamberg. Eva's research interests focus on mind perception and embodied cognition and their application to social robotics and cognitive offloading.