Cognition 224 (2022) 105076

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Spontaneous perspective taking toward robots: The unique impact of humanlike appearance

Xuan Zhao [a,*], Bertram F. Malle [b]

[a] Department of Psychology, Stanford University, United States of America
[b] Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, United States of America

ARTICLE INFO

Keywords:
Perspective taking
Human-robot interaction
Social cognition
Artificial intelligence
Anthropomorphism
Theory of Mind

ABSTRACT

As robots rapidly enter society, how does human social cognition respond to their novel presence? Focusing on one foundational social-cognitive capacity—visual perspective taking—seven studies reveal that people spontaneously adopt a robot's unique perspective and do so with patterns of variation that mirror perspective taking toward humans. As they do with humans, people take a robot's visual perspective when it displays goal-directed actions. Moreover, perspective taking is absent when the agent lacks human appearance, increases when the agent looks highly humanlike, and persists even when the humanlike agent is perceived as eerie or as obviously lacking a mind. These results suggest that visual perspective taking toward robots is consistent with a "mere appearance hypothesis"—a form of stimulus generalization based on humanlike appearance—rather than following an "uncanny valley" pattern or arising from mind perception. Robots' superficial human resemblance may trigger and modulate social-cognitive responses in human observers originally developed for human interaction.

1. Introduction

Consider a world where robots drive our cars, teach our children, and assist our aging parents. Futuristic as these scenarios would have appeared a decade ago, fast-developing technologies in robotics and artificial intelligence have moved such possibilities closer to reality (Belpaeme, Kennedy, Ramachandran, Scassellati, & Tanaka, 2018; Rahwan et al., 2019). Yet when encountering robots, people have no uniquely adapted psychological mechanisms to make sense of those machines; what people rely on, instead, are the social-cognitive capacities evolved over hundreds of thousands of years for interacting with other humans (Cosmides & Tooby, 1997; Herrmann, Call, Hernàndez-Lloreda, Hare, & Tomasello, 2007; Heyes & Frith, 2014). Therefore, understanding whether, when, and to what extent people employ their social-cognitive capacities toward the newly arrived robots requires systematic research and promises new insights into human psychology (Broadbent, 2017).

One social-cognitive capacity particularly relevant to collaboration and communication with other humans—and likely robots—is visual perspective taking. In human-human interaction, disparities in visual perspectives frequently pose an inherent challenge for human minds: Because every person perceives the world from their unique vantage point, overlooking other people's distinct viewpoints can result in confusion and misunderstanding, whereas considering and adopting their distinct viewpoints facilitates successful social engagement (Brennan, Galati, & Kuhlen, 2010; Keysar, Barr, Balin, & Brauner, 2000). Therefore, the ability to judge whether someone else can see an object (typically referred to as "Level-1" perspective taking), as well as how they see it differently from oneself (typically referred to as "Level-2" perspective taking), constitutes a developmental milestone during the first years of life (Flavell, 2004; Moll & Meltzoff, 2011; Piaget & Inhelder, 1956) and is often linked to other important social-cognitive abilities such as theory of mind and empathy (Erle & Topolinski, 2017; Hamilton, Brindley, & Frith, 2009).

Accumulating evidence suggests that the human mind is surprisingly attuned to other people's visual perspectives: For instance, merely seeing another person—even without actual interaction—often leads people to infer and even adopt that person's distinct viewpoints (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010; Surtees, Apperly, & Samson, 2016; Ward, Ganis, & Bach, 2019; Zhao, Cusimano, & Malle, 2015). As people work alongside robot workers and share roads with robot drivers, perspective disparities with robots will likely become

* Corresponding author at: Department of Psychology, Stanford University, 450 Jane Stanford Way, Building 420, Stanford, CA 94305, United States of America.
E-mail address: xuanzhao@stanford.edu (X. Zhao).

https://doi.org/10.1016/j.cognition.2022.105076
Received 5 May 2021; Received in revised form 25 February 2022; Accepted 28 February 2022
Available online 29 March 2022

unavoidable. While researchers have attempted to design robots that can take people's perspectives (e.g., Trafton et al., 2005), we know very little about whether people spontaneously consider how the world would appear from a robot's viewpoint.

Granted, robots are unlikely to have humanlike visual experiences, nor are humans able to have robot-like vision. We therefore define perspective taking toward a robot not as a simulation of what it is *experiencing*, but as an inference of what the robot *could see* from its vantage point—that is, how objects in a three-dimensional world would appear to the robot given its physical location. This functional definition avoids the much harder question of what it would be like to experience the world as a different "species" (Nagel, 1974) and can help solve the practical challenge of resolving perspective disparities with robots in a shared space. To begin identifying the scope and limits of such possible perspective taking, our research focuses on the role that robot appearance and action play in eliciting spontaneous (i.e., unprompted) perspective-taking responses in human observers.

In the present paper, we test the hypothesis that humanlike appearance is a powerful trigger for visual perspective taking toward robots. Research on human-robot interaction has revealed that humanlike robots can stimulate a number of social-cognitive processes, ranging from lower-order processes such as following a robot's gaze direction and representing its movements as goal-directed (Krach et al., 2008; Meltzoff, Brooks, Shon, & Rao, 2010; Urgen, Plank, Ishiguro, Poizner, & Saygin, 2013), to higher-order processes such as ascribing humanlike traits and mental states, thereby anthropomorphizing these machines agents (de Graaf & Malle, 2019; Epley, Waytz, & Cacioppo, 2007). Likewise, we hypothesize that people may spontaneously (i.e., without explicit prompts) infer a robot's unique visual perspective and may do so with patterns of variation that mirror perspective-taking responses toward humans. Specifically, because observing another person's goal-directed actions (e.g., gaze or reaching) invites people to adopt that person's physical viewpoint (Tversky & Hard, 2009; Zhao et al., 2015), observing goal-directed actions by a humanlike robot may similarly scaffold social cognition toward the robot and increase people's tendency to adopt its distinct perspective.

Importantly, we predict that, as robots appear physically more humanlike, the likelihood of perspective taking increases, which we term the *"mere-appearance hypothesis."* This hypothesis is informed by decades of research on stimulus generalization, where organisms extend highly practiced stimulus responses to new stimuli if they resemble the original (Guttman & Kalish, 1956; Shepard, 1987). While generalization underlies the human brain's remarkable learning abilities, generalization from appearance can prompt surprisingly social reactions to clearly inanimate objects—such as face-like drawings or eye-like shapes (Dear, Dutton, & Fox, 2019; Johnson, Slaughter, & Carey, 1998)—and may similarly underlie how people respond to robots.

An alternative, prominent theory of the relationship between robots' humanlike appearance and human responses is the *"uncanny valley" hypothesis*. Primarily focusing on people's affective responses to robots, this theory posits that people's increasingly positive response to increasingly humanlike robots plummets into a deep "valley" of repulsion as the robots become highly, yet imperfectly, human-looking (Mori, 1970). Supporting this hypothesis, recent studies have documented that people (at least in the U.S.) not only find highly humanlike robots to be eerie (e.g., Kim, Bruce, Brown, de Visser, & Phillips, 2020; Wang, Lilienfeld, & Rochat, 2015), but they may even trust such robots to a lesser extent (Mathur & Reichling, 2016). Applying the uncanny valley hypothesis to the topic of taking a robot's perspective, one should expect that extremely humanlike appearance will similarly backfire when taking the visual perspective of robots. Thus, while both the mere-appearance hypothesis and the uncanny valley hypothesis predict that people should increasingly take robots' perspective as they look more humanlike, the uncanny valley hypothesis uniquely predicts a sharp decline of perspective taking when the robot becomes eerily humanlike.

A third prominent framework of how people conceive of nonhuman agents is the *mind perception hypothesis*, which emphasizes the central role of attributing a humanlike mind in people's reactions to non-human agents (Gray & Wegner, 2012; Waytz & Norton, 2014). Mind perception is arguably relevant to perspective taking—after all, visual perspective taking is often regarded as one aspect of Theory of Mind (Samson et al., 2010), which presumes the presence of another "mind" to begin with. Like both previous hypotheses, this account also predicts that people engage in more visual perspective taking as robots appear increasingly humanlike, but it posits that such responses result from perceiving human mental capacities in those agents (Looser & Wheatley, 2010; Zhao, Phillips, & Malle, 2019). By contrast, the mere-appearance hypothesis claims a direct impact of humanlike appearance on perspective taking, unmediated by mind perception and persisting even when a highly humanlike agent is clearly lacking a mind.

In what follows, we examine whether, when, and why people spontaneously take a robot's visual perspective. Studies 1A and 1B test if people take the perspectives of moderately humanlike robots and are more inclined to do so upon observing the robots' goal-directed actions. In Studies 2 to 5, we manipulate the agents' appearance and mental capacities to examine to what extent people's perspective-taking responses vary as a function of the agent's physical appearance (the mere-appearance hypothesis), affinity and likability (uncanny valley hypothesis), or the amount of mind attributed to the agent (mind perception hypothesis). Stimuli, data, analysis (R code), and output files for all studies are available on Open Science Framework: https://osf.io/ymkqp/.

2. Study 1A: Humanoid robots

Would people ever take a robot's visual perspective? As an initial step to address this question, we compared whether people would adopt the perspective of a humanoid robot (Nao or Baxter), compared to that of an actual human, when seeing it display goal-directed actions.

2.1. Method

2.1.1. Participants

Participants were U.S. residents recruited via the crowdsourcing platform Amazon Mechanical Turk (MTurk). Because the effect sizes of observing robots' goal-directed actions were unknown, we first collected data featuring a human agent to guide our decision on the sample size for sufficiently powered robot conditions. Specifically, we first administered the three conditions featuring a human male (and the corresponding control condition), aiming at approximately $n = 60$ per condition prior to exclusion, which was consistent with previous research that employed similar open-ended questions (Tversky & Hard, 2009). After examining people's perspective taking rates and anticipating that those rates would be smaller for robots, we set sample sizes for conditions featuring Nao and Baxter to $n = 100$ per condition. Further, we replicated the initial results in the human conditions by featuring a human female agent and targeting the same sample size of $n = 100$ as in the robot conditions—testing the robustness of findings across agent gender.[1]

A total of 1729 participants completed Study 1A ($M_{age} = 33.56$, $SD_{age} = 12.18$; 56.4% female; $n = 59$–69 per condition for those with a human male, $n = 89$–103 per condition for all other conditions). Participant counts varied slightly across conditions due to unequal numbers of participants being randomly assigned to each condition.

[1] As a result of these procedures, participants were not recruited all at once and were not randomly assigned to agents. To address this issue, we conducted a pre-registered replication study (i.e., Study 5) where all conditions were administered simultaneously to ensure random assignment, and we replicated the results from Study 1A.

2.1.2. Design

To compare people's tendency to engage in perspective taking across agents and actions, we employed a 3 (agent type: Nao, Baxter, human) × 3 (action: side-look, gaze, reach) between-subjects factorial design. In addition, we included three control conditions and four exploratory conditions, as detailed below.

2.1.2.1. Paradigm. To capture spontaneous visual perspective taking, we created a single-trial task where each participant saw a photograph depicting an agent (either a robot or a human) engage in one specific action behind a table. On the table was a "9" from participants' own perspective, which would appear as a "6" from the agent's vantage point (see Fig. 1). Below the photograph was an open-ended question, "*What number is on the table*?", to which each participant provided one response. With no time constraint and no explicit demand on how to respond, participants were free to write down responses that seemed most intuitive to them. Similar open-ended verbal tasks have been introduced in previous research to investigate people's sensitivity to multiple spatial frameworks or visual perspectives (Tversky & Hard, 2009; Zhao et al., 2015).

To analyze these responses, we calculated the proportion of participants who adopted the other agent's viewpoint (i.e., responding "6" or "six") in each condition—hereafter, the "*VPT rate*." Hence, comparing VPT rates across agents (robots vs. humans) reveals whether humanoid robots can trigger perspective taking to the same *extent* as humans can. Furthermore, comparing participants' VPT rates for each agent across actions reveals whether robots and humans elicit the same *pattern* of perspective taking. Therefore, this paradigm allows face validity and tight experimental control in measuring and comparing spontaneous perspective taking when observing an agent and its action.

2.1.2.2. Humanoid robots and human agents. We created photographs featuring either a humanoid robot of Nao or Baxter or an actual human male or female actor (see Fig. 1) in the scene. The Nao robot, a 58-cm-tall red-colored humanoid by *Aldebaran Robotics*, is the most widely used social robot for education (Belpaeme et al., 2018). In all stimulus photos, similar to the human agents, the robot was shown sitting in a black office chair, adjusted to reach-optimal height behind the table. In addition, we edited its pupils in the photographs using Adobe Photoshop CS6 to make its eyes clearly indicative of gaze direction. The Baxter robot is a 6-ft-tall red-colored humanoid robot by *Rethink Robotics*. This robot has a small screen toward the top that can present images of its eyes to indicate gaze direction. We used Adobe Illustrator to create pictures of Baxter's face with various gaze directions and projected them on its screen. The Baxter robot also has an upright posture with two sturdy arms and two gripper "hands." According to the Anthropomorphic Robot Database (ABOT)—a collection of real-world robots with humanlike features—Baxter's appearance is judged as slightly less humanlike than that of Nao, and removing its face and head would likely make the robot appear even less humanlike (Phillips, Zhao, Ullman, & Malle, 2018).

The human agents were a male and a female, both in their 20s. The male was Caucasian, and the female was Asian.

2.1.2.3. Actions. Within the 3 × 3 factorial design, each agent displayed one of three actions: looking to the side ("side-look" condition), looking at the number ("gaze" condition), or reaching a hand toward the number while looking at it ("reach" condition) (see Fig. 1). Such gaze and reach actions are routinely perceived as goal-directed when performed by human agents (Phillips & Wellman, 2005; Sodian & Thoermer, 2004).

2.1.2.4. Control conditions. We included three control conditions to rule out the influence of low-level perceptual features in the physical setup. In the *novelty-control* condition, we replaced Nao with a red electric guitar "sitting" in a chair, and in two *absence-control* conditions, we presented the same physical setups, yet without the agent or the chair (see Fig. S1 in the Supplemental Materials, panels a–c).

2.1.2.5. Exploratory conditions. To explore the potential impact of a robot's degree of humanlikeness, we took advantage of Baxter's head-like structure and manipulated its appearance in another set of four conditions: While Baxter was shown to be standing still or reaching for the object, we either turned its screen off (the *"presence-no-face"* and *"reach-no-face"* conditions) or simply cropped its head out of the scene (*"presence-no-head"* and *"reach-no-head"* conditions) (see Fig. S1 in the Supplemental Materials, panels d–g). In light of prior research (Phillips et al., 2018), we expected this manipulation to create only a mild impact on Baxter's overall degree of humanlikeness, but it nevertheless offered tight control on the robot model and provided an intriguing opportunity to probe the potential impact of humanlike features on perspective taking.

2.1.3. Procedure

All studies were approved by the Institutional Review Board at Brown University. All participants read and provided informed consent before completing the experiments. After typing in their identification codes, participants saw the experimental webpage, which simultaneously displayed the stimulus photo, the open-ended question "*What number is on the table?*" underneath the photo, and a textbox underneath the question, where participants typed their responses (with no constraints on word limit). After typing, participants clicked "continue" to submit their answer, filled out three demographic questions (gender, age, and English proficiency), and received their payment codes.

2.1.4. Data analysis

In all studies, we applied the same two-step data exclusion procedure, which had been defined prior to data collection. First, in order to gain some control of the online research environment, we measured how much time participants spent on the experimental webpage and used these "page time" data to exclude those who took an excessively long time to generate responses (i.e., three standard deviations beyond the mean in their respective condition).[2] This excluded 6.99% of all responses, but even analyzing all participants regardless of how long they spent on the webpage led to the same findings (see Supplementary Materials). Second, because we were interested in whether people took either a robot's perspective or their own, and very few participants provided responses from both perspectives, we excluded nine participants who provided multiple perspectives (e.g., "6 for the bot 9 for me"). This accounted for 0.52% of all responses)[3], but including those responses by coding them as "perspective taking" did not change our findings either (see Supplemental Materials). In addition, we excluded one participant who did not respond with a number (i.e., "robot"). This data exclusion procedure yielded a total of 1599 participants for final analysis (M_{age} = 33.43, SD_{age} = 12.20; 56.6% female; n = 57–64 for conditions with a human male, n = 86–99 for all other conditions).

In this and all subsequent studies, we conducted logistic regression analyses with participants' perspective-taking responses as the dichotomous outcome variable (0 = responding "9" from self-perspective; 1 = responding "6" from other-perspective) and action as a predictor. Given that action had three levels, we consistently tested two a priori orthogonal contrasts—also known as Helmert contrasts: The first

[2] We also considered excluding participants who responded too quickly; however, three standard deviations below the means yielded negative page time values.

[3] Combining those nine participants with the 15 participants who also provided multiple perspectives but were already excluded due to excessively long page times, a total of 24 participants responded with multiple perspectives across all 1729 who completed the study (1.39%), again suggesting that such responses were relatively infrequent.

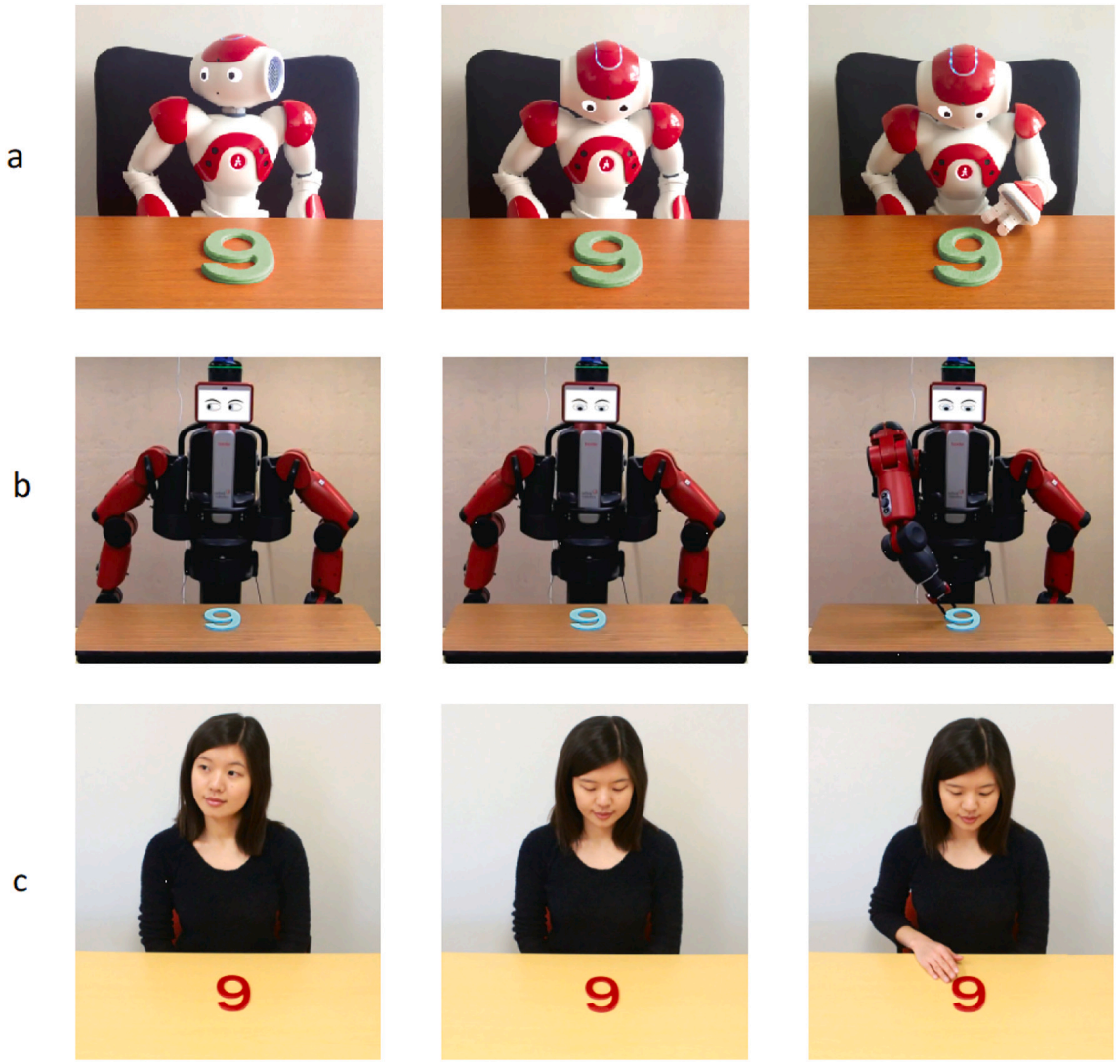

Fig. 1. Humanoid Robots of Nao (a) and Baxter (b), and Human Female (c) in the Side-Look (left panel), Gaze (middle panel), and Reach (right panel) Conditions in Study 1A.

compared the two goal-directed actions (*gaze, reach*) with the *side-look* condition to examine whether observing goal-directed actions led to a higher VPT rate than observing no particular goal-directed actions. The second compared the *reach* condition with the *gaze* condition to test whether observing the agent reach toward the number made people become more inclined to adopt the agent's perspective than merely observing the agent's gaze alone. Agent type (human, Nao, Baxter) was the second predictor, and we also represented this three-level factor by two Helmert contrasts in the regression model: The first compared human against the humanoid robots (both Nao and Baxter), and the second compared Nao against Baxter. All *p*-values reported are two-tailed.

2.2. Results

Because participants were equally inclined to take the perspectives of a human male and a human female ($p = .70$), we combined those data as representing a single agent type (i.e., human). VPT rate and cell size for each agent in each action condition and each study can be found in Table S1 in the Supplemental Material.

2.2.1. Primary analysis

Would people ever take a robot's perspective? When there was no agent in the scene, people rarely reported the number "6"—the VPT rate was 0.7% (1 out of 142) in two absence-control conditions combined. Yet seeing the Nao robot in a photograph prompted a notable proportion of people to adopt its perspective—an average of 24.5% of participants across side-look, gaze, and reach conditions, which is considerably higher than VPT rates in the absence-control conditions, $z = 3.77$, $p < .001$, $d = 0.39$. Similarly, including Baxter in the photographs (side-look, gaze, reach) also significantly increased VPT rates to 21.4%, which was again considerably higher than VPT rates in the absence-control conditions, $z = 3.60$, $p < .001$, $d = 0.37$. Moreover, both Nao and Baxter elicited significantly higher VPT rates than the closely matched novelty-control condition, in which a vibrantly colored electric guitar was placed behind the table (8.3%), $zs = 3.22$ and 2.77, $ps = 0.001$ and 0.006, $ds = 0.38$ and 0.33, respectively.

How does perspective taking elicited by humanoid robots compare to that elicited by humans? A logistic regression analysis of 3 agent types (Helmert contrasts: human vs. robots; Nao vs. Baxter) × 3 actions (Helmert contrasts as above) revealed that humanoid robots Nao and Baxter elicited lower VPT rates overall (24.5% and 21.4%, respectively) than the human agents (33.8%), $z = 3.52$, $p < .001$, $d = 0.22$, and between the two humanoid robots, Baxter and Nao elicited comparable VPT rates, $z = 0.65$, $p = .52$. Critically, participants' responses to human and robot agents were characterized by similar patterns across actions,

as no interaction terms between action and agent type reached statistical significance, $ps > 0.16$: Across all three agent types, VPT rates were higher in the gaze and reach conditions than in the side-look condition (human agents: 41.7% and 43.9% vs. 16.2%; Nao robot: 28.9% and 30.4% vs. 13.2%; Baxter robot: 23.4% and 26.5% vs. 14.0%), $z = 5.54$, $p < .001$, $d = 0.37$; furthermore, VPT rates between gaze and reach conditions were indistinguishable among all agent types, $z = 0.64$, $p = .52$.

Taken together, these results suggest that participants exhibited strikingly similar patterns of VPT responses toward humanoid robots and humans, but they were less inclined overall to adopt the perspective of a humanoid robot than that of another human (see Fig. 2).

2.2.2. Exploratory analysis

Next, we examined the four exploratory conditions that manipulated the presence of facial features in the same robot model (Baxter). First, when Baxter's screen was either turned off (i.e., the "no-face" conditions) or cropped out of the photographs (i.e., the "no-head" conditions) such that it was merely standing still, a small fraction of participants took Baxter's perspective (9.4% and 12.2%, respectively), but not significantly more than those in the novelty-control condition who saw an electric guitar, $z = 0.64$, $p = .52$. When the image showed Baxter reaching out for the number, participants were marginally more inclined to adopt its perspective compared to the novelty-control condition, despite its lack of a face or a head (16.0% and 17.4%, respectively), $z = 1.88$, $p = .06$, $d = 0.24$. This result suggests that even in the absence of gaze, a robot's goal-directed reaching can trigger perspective taking to some extent.

Examining how the specific humanlike features of head and face may have moderated perspective taking, we next combined the above four conditions with the gaze and reach conditions in the primary analysis to conduct a 3 appearance (Helmert contrasts: with face/head vs. no face/head; no-face vs. no-head) × 2 action (reach vs. no reach) logistic regression analysis. Results showed that having a head and a face with a pair of gazing eyes elicited more perspective-taking responses (25.0%) than having no face (12.8%) or no head (14.9%), $z = 2.67$, $p = .003$, $d = 0.26$, whereas VPT rates in the latter two conditions were indistinguishable, $p = .56$. Finally, given that the presence of face and head was confounded with the presence of gazing eyes in the above analysis, yet having a head and a face does not necessitate the goal-directed action of gaze, we also measured the "pure" impact of those appearance features on perspective taking. To this end, we compared VPT rate in the side-look condition (14.0%), where Baxter had a pair of eyes but looked away from the number, against the two exploratory conditions where Baxter similarly stood still but had no face or head (9.4% and 12.2%, as mentioned above). However, VPT rate in the side-look condition was comparable to those in the two presence conditions, $z = 0.94$, $p = .35$, suggesting that the impact of humanlike appearance on perspective taking lies more in the goal-directed behaviors that it enables—such as gaze and reaching—than in the mere possession of the physical features per se.

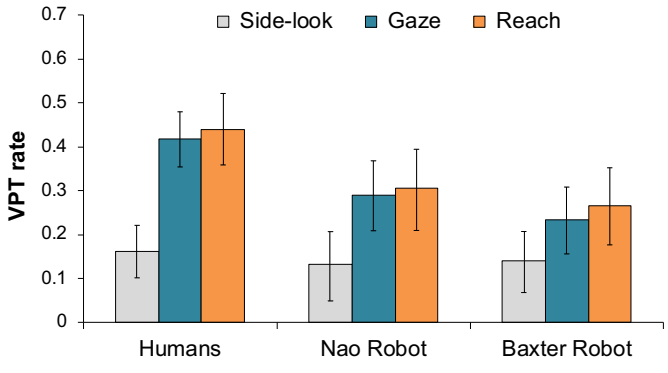

Fig. 2. Results from Study 1A: Visual perspective taking (VPT) rates toward human agents, a humanoid robot Nao, and a humanoid robot Baxter across action conditions. All actions were presented in photographs. Error bars represent 95% confidence intervals (CIs) with 5000 resamples.

3. Study 1B: Dynamic actions of humanoid robots

A sizable portion of participants in Study 1A spontaneously adopted the perspective of humanoid robots and exhibited similar response patterns as when observing another human. Study 1B sought to replicate these results and to further test whether conditions that modify perspective taking toward humans (namely dynamic visual display) similarly modify perspective taking toward robots.

3.1. Method

3.1.1. Participants

Similar to Study 1A, we first recruited participants for the conditions featuring a human male (side-look, gaze, reach) and aimed at $n = 60$ participants per condition, and then recruited for the conditions featuring the humanoid robots Nao and Baxter, as well as for a human female, with an increased target sample size of $n = 100$ participants per condition before exclusion. A total of 1291 MTurk participants with IP addresses located within the U.S. completed this study ($M_{age} = 33.76$, $SD_{age} = 11.47$; 51.3% female; $n = 69$–71 per condition for those with a human male; $n = 96$–102 per condition for all other conditions). Among them, 1095 were assigned to the 3 (agent types) × 3 (action) design, and 196 participants were assigned to the two exploratory conditions that removed Baxter's face or head. Based on the predetermined two-step data exclusion procedure, 72 participants were then excluded—63 participants (4.9% of all responses) spent an excessively long time on the experimental page, and nine more provided responses containing multiple perspectives or invalid responses (0.7% of all responses), yielding a total of 1219 participants for final analysis ($M_{age} = 33.58$, $SD_{age} = 11.36$; 51.3% female; $n = 66$–69 for conditions with a human male, $n = 90$–94 for all other conditions). All findings remained the same when we analyzed the full data set without participant exclusions (see Supplementary Materials).

3.1.2. Design

3.1.2.1. Actions and agents. A dynamic display of the human reach action tends to reveal the actor's goal more vividly than a mere gaze (Sodian & Thoermer, 2004), so when shown in motion, the reaching action should increase perspective taking *relative to* gaze when performed by both humans and humanoid robots. Similar to Study 1A, we featured Nao and Baxter robots as the humanoid robots and a Caucasian male and an Asian female as the human agents. Different from Study 1A, however, all actions were presented as five-second videos. Hence, this study again employed an agent type (humanoid robots, human) × action (side-look, gaze, reach) between-subjects design.

For the robots, we employed Nao's and Baxter's movement-recording software to make their head-turning and hand-reaching movements relatively smooth and steady and finished with the same poses as in the photographs. To ensure that videos of the same condition were maximally similar across agents, we produced all videos under similar protocols of movement sequences and followed similar timelines (see the Supplementary Materials). However, due to mechanical limitations of the robots, their actions still look somewhat jerky—or "robotic"—compared to human actions.

3.1.2.2. Exploratory conditions. Similar to Study 1A, we included two exploratory conditions that manipulated Baxter's screen and head to further test how the degree of humanlikeness impacted perspective taking. While Baxter performed the same reach action, we either turned its screen off (the "*reach-no-face*" condition) or cropped its head out of

the scene (the "*reach-no-head*" condition). We did not include the "presence" conditions where it merely stood still without the face or the head, because such videos would be identical to the photographs in Study 1A.

3.1.3. Procedure

The procedure was similar to that in Study 1A except that, in every condition, the video appeared first on the experimental webpage, played for five seconds, and then "froze" on its last frame. Then, the question, "*What number is on the table*?" and the textbox immediately appeared below the final frame of the video, where participants freely entered their open-ended responses before submitting. Once again, responding "6" or "six" from the agent's vantage point was considered as perspective taking.

3.2. Results

3.2.1. Primary analysis

As in Study 1A, we first confirmed that participants were equally inclined to take the perspectives of a human male and a human female ($p = .56$) and combined those conditions as representing the same agent type. Then, we conducted logistic regression analyses with the same Helmert contrasts on agent type (human vs. robots; Nao vs. Baxter) and action (side-look vs. goal-directed actions; gaze vs. reach) as in Study 1A. Once again, we found a main effect for the human vs. robot contrast. As shown in Fig. 3, a sizable portion of participants adopted the humanoid robots' perspectives (29.7%), but the VPT rate toward humanoid robots was lower than that toward human agents (42.7%), $z = 4.19$, $p < .001$, $d = 0.26$, and VPT rates did not significantly differ between Nao and Baxter, $p = .49$. Furthermore, we found main effects of action for both contrasts: Not only did the gaze and reach videos elicit higher VPT rates on average (42.7%) than the side-look videos (21.8%), $z = 5.71$, $p < .001$, $d = 0.37$, but the reach videos also evoked higher VPT rates (47.8%) than the gaze videos (37.5%), $z = 2.84$, $p = .005$, $d = 0.22$ (see Table S1 for VPT rates in each condition). Critically, participants exhibited strikingly similar patterns of VPT responses for both humanoid robots and humans, with no interaction between action and agent type on any contrasts, $ps > 0.19$.

3.2.2. Exploratory analysis

When a video showed the Baxter robot with a dark screen gradually moving its arm to reach for the number, participants were somewhat inclined to adopt its perspective despite the lack of a face (33.0%), which was significantly higher than taking the perspective of an identically-looking Baxter merely standing still in Study 1A (9.4%), $z = 3.61$, $p < .001$, $d = 0.54$. Similarly, when Baxter's screen was entirely cropped out, participants were still inclined to adopt the reaching robot's perspective

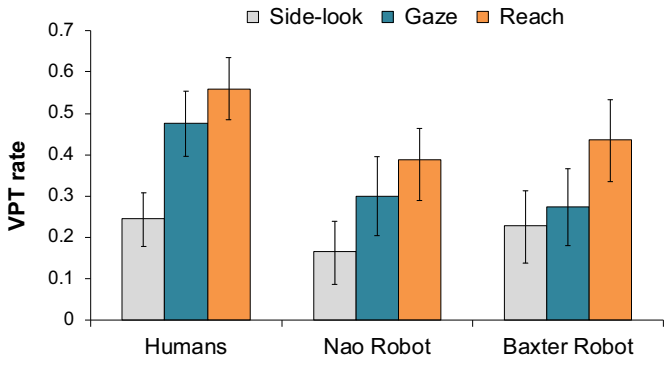

Fig. 3. Results from Study 1B (video stimuli): Visual perspective taking (VPT) rates toward human agents, a humanoid robot Nao, and a humanoid robot Baxter across action conditions. All actions were presented in five-second videos. Error bars represent 95% CIs with 5000 resamples.

despite the lack of a head (26.9%), more so than they adopted the perspective of the same Baxter standing still in Study 1A (12.2%), $z = 2.37$, $p = .018$, $d = 0.36$. Thus, seeing Baxter's goal-directed reaching behavior increased perspective taking relative to seeing its mere presence. Next, we combined the two exploratory reach conditions reported above (no face/head), the gaze and reach conditions (both with gazing eyes), and the two mere presence conditions in Study 1A for an exploratory 3 appearance (Helmert contrasts: with face/head vs. no face/head; no-face vs. no-head) × 2 action (reach vs. no reach) logistic regression analysis. Once again, having a face/head led to significantly higher VPT rates (an average of 35.3%) than having either no face (an average of 21.8%) or no head (an average of 20.0%) given the same robot model, $z = 3.34$, $p < .001$, $d = 0.30$. A robot's humanlike face, which enables it to gaze at the number, considerably increased people's tendency to adopt the perspective of that robot.

4. Study 2A: Machine-like robot

Study 1B replicated Study 1A's finding that humanoid robots can indeed elicit spontaneous perspective taking, albeit to a lesser extent than human actors. Furthermore, with dynamic videos, the response pattern across robots' goal-directed actions paralleled that elicited by human actors. In Studies 2A and 2B, we tested whether an agent that lacks humanlike appearance (i.e., mechanical robot or cat) would fail to trigger visual perspective taking.

4.1. Method

4.1.1. Participants

To ensure a sample size of at least 60 participants per condition after exclusion, we opened recruitment to 210 participants ($n = 70$ per condition) for a total of three conditions, and MTurk returned 209 participants. Nine participants were excluded after page time screening, and no participants responded from multiple perspectives, resulting in 200 participants for final analysis ($M_{age} = 33.93$, $SD_{age} = 11.93$; 56.3% female; $n = 65$–69 per condition).

4.1.2. Design

Study 2A featured "Thymio"—a mechanical robot in a box shape that can move autonomously and blink its lights yet completely lacks characteristic human appearance features (see Fig. 4): It does not have eyes or limbs and is therefore incapable of anything resembling biological gaze or reaching. However, it can still perform self-propelled movements such as approaching an object, and it can blink its light to "communicate" with the user. According to previous research, these behaviors can appear goal-directed and make people ascribe agency and intentionality to a nonhuman agent (Heider & Simmel, 1944; Mandler, 1992; Premack, 1990).

Leveraging Thymio's capacity for self-propelled motion, we created five-second videos that conceptually implemented side-look, gaze, and reach actions by showing it, respectively, driving away from the number, blinking its light toward the number, or driving toward the number after blinking (see Fig. 4).

4.1.3. Procedure

The procedure was identical to that in Study 1B. Participants were randomly assigned to one of the action conditions (side-look, gaze, reach), watched a five-second video on the experimental webpage, responded to the open-ended question (i.e., "What number is on the table?") in a textbox, and then completed basic demographic information on the final page.

4.2. Results

Among all participants who observed Thymio, only two in the side-look condition responded "6" from its "perspective" (3.1%), and none in

Fig. 4. Screenshots of the Final Frame of the Video Stimuli in Study 2A: the side-look (left), gaze (middle), and reach (right) conditions of Thymio, a non-humanlike robot. White arrows indicate the movement trajectory of the robot over five-second videos.

the gaze or reach conditions did. A cross-study comparison between VPT rates toward Thymio and the humanoid robots of Nao and Baxter revealed that, when the robots' actions were all dynamically displayed in videos, people were much more inclined to take the perspective of the humanlike robots (29.7%) than that of the mechanical Thymio (1.0%), $z = 5.21$, $p < .001$, $d = 0.43$.

One alternative explanation for such little perspective taking toward Thymio is that participants might have found its autonomous movements too abstract to make sense. Hence, we introduced an agent that is decidedly non-human-looking but can display goal-directed actions that are biologically similar to humans: a cat.

5. Study 2B: Biological cat

5.1. Method

5.1.1. Participants

As in Study 2A, we recruited 210 participants ($n = 70$ per condition) for a total of three conditions. Among them, 25 participants were excluded after page time screening, and no participants responded from multiple perspectives,[4] resulting in 185 participants for data analysis ($M_{age} = 32.65$, $SD_{age} = 10.47$; 46.5% females; $n = 60$–63 per condition).

5.1.2. Design

We created photographs of a cat engaging in similar side-look or goal-directed actions of gaze and reach as the agents in Study 1A. In three photographs (see Fig. 5), the cat either stared into space toward no particular goal object (side-look condition), gazed at the number (gaze condition), or gazed at the number while reaching for it (reach condition). We also attempted to create short videos for the cat but later aborted that plan, because we could not precisely control its movement pace and trajectory to make the cat's actions comparable to those by humans and robots.

5.1.3. Procedure

The study procedure was similar to that in Study 1A. Participants were randomly assigned to the side-look, gaze, or reach condition, saw one of the photographs along with the open-ended question, "*What number is on the floor?*", and typed their response in a textbox. They ended the study by providing basic demographic information and reporting whether they currently owned pets and, if so, what species.

5.2. Results

VPT rates toward the cat were low across conditions (an average of 5.4%), and the difference in VPT rates between the side-look (1.7%) and the gaze and reach conditions (6.4% and 7.9%, respectively) was not significant, $z = 1.42$, $p = .17$, $d = 0.22$.[5] In addition, VPT rates between cat owners and non-owners was negligible (6.2% vs. 5.0%, respectively), $p > .65$, ruling out potential effects of familiarity or affection. Importantly, in a cross-study comparison, we contrasted VPT rates toward the cat against those toward the humanoid robots Nao and Baxter in Study 1A and found that people were more inclined to take the perspective of the humanlike robots (22.9%) than that of the cat, $z = 4.84$, $p < .001$, $d = 0.41$. In fact, the highest VPT rate elicited by the cat in the present study (7.9% in the reach condition) was no higher than the VPT rate when people merely saw an electric guitar behind the number (8.3% in the novelty-control condition in Study 1A).

6. Study 3: Android robot

Studies 2A and 2B showed that, in contrast to humanoid robots, agents lacking in humanlike appearance triggered relatively little perspective taking, even when they engaged in goal-directed actions that were conceptually similar—or even biologically analogous—to human actions. Study 3 examined whether a highly humanlike robot would significantly enhance perspective taking, as the mere-appearance hypothesis suggests. Alternatively, extreme humanlikeness could push people into an "uncanny valley," thus resulting in a sharp decline in perspective taking.

6.1. Method

6.1.1. Participants

We aimed at recruiting 100 participants per condition for a total of three conditions and ended up with responses from 298 participants. Among them, 43 participants were excluded due to failing the comprehension check question (i.e., identified the robot as other agent types, see below); eight participants were excluded after page time screening; and three were excluded because their responses contained multiple perspectives, yielding 234 participants for data analysis ($M_{age} = 36.5$, 40.6% female; $n = 76$–81 per condition). All findings remained the same when we analyzed the full dataset without participant exclusions.

6.1.2. Design

In Study 3, we featured one of the most humanlike android robots in the world (Erica), had it display the same actions as the humanoid robots in Study 1A, and tested whether the level of perspective taking toward

[4] The only participant who provided multiple perspectives was already excluded during page time screening.

[5] When all 210 participants were included in the analysis, difference in VPT rates between the side-look condition (1.4%) and the goal-directed actions of gaze and reach (10.4% and 9.7%, respectively) reached statistical significance, $z = 1.97$, $p = .049$, $d = 0.29$.

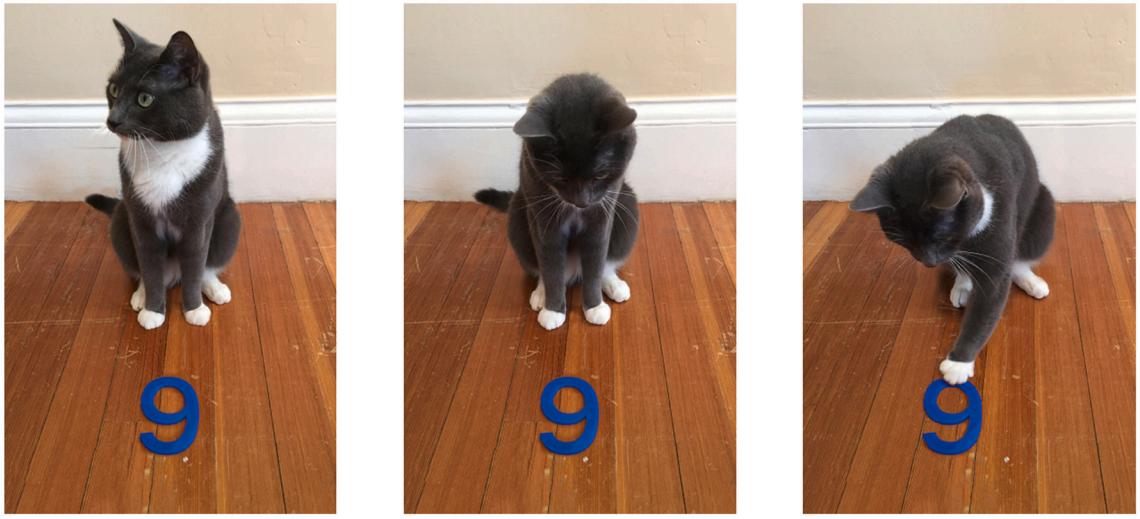

Fig. 5. Photo Stimuli in Study 2B: the side-look (left), gaze (middle), and reach (right) conditions of a biological cat.

the android would exceed that toward the humanoid robots.

Erica is a state-of-the-art android robot developed by the Japan Science and Technology Agency (JST) at Osaka University and Advanced Telecommunications Research Institute International (ATR) at Kyoto University. It is acclaimed as one of the most humanlike robots in the world, with its skin made out of silicone resin and its face modeled after 30 images of Japanese and European females in their 20s (McCurry, 2015). In an additional survey ($N = 180$; U.S. residents), we measured how likable (likable, friendly, pleasant) and uncanny (eerie, creepy, unnerving) people found Erica and all the other agents in the present article. Erica was perceived as highly uncanny and the least likable among all agents (see Survey 1 in the Supplemental Materials).

Across three photographs, Erica displayed the same side-look, gaze, or reach actions as agents in Study 1A (see Fig. 6).

6.1.3. Procedure

In a pilot survey ($N = 60$), we found that a considerable number of participants who viewed an unlabeled photograph of Erica mistook this robot for either a non-intelligent artifact (46.8%, such as a mannequin, wax figure, or doll) or a human (35.5%), and only 11.3% explicitly identified it as a robot. In the actual study, we therefore provided an instruction page before the experimental webpage that stated, "*On the next page, you are going to see a picture containing an android—a robot designed to look like a human—and a question below the picture.*" The experimental webpage was identical to that in Study 1, which included a photograph above the question ("*What number is on the table?*") and a textbox below. After this page, we also introduced a manipulation check question, "*Recall the picture you just saw. The figure sitting behind the table was a …*" and included five options: human, robot, doll, mannequin, and wax figure. We excluded participants who failed the manipulation check before applying the standard two-step data exclusion procedure.

6.2. Results

Once again, participants exhibited the same response pattern across actions displayed by the highly humanlike android robot as they did to actions by humans—an increase from side-look (21.0%) to gaze and reach actions (41.6% and 48.1%, respectively), $z = 3.44$, $p < .001$, $d = 0.48$, with the latter two indistinguishable from one another, $p = .41$. Furthermore, comparing VPT rates across studies revealed that people were more inclined to adopt the perspective of the strikingly human-looking android robot (37.2%) compared to the humanoid robots Nao and Baxter (22.9%), $z = 4.07$, $p < .001$ $d = 0.32$. In fact, people were just as inclined to take the android's perspective as that of another human (33.8%), $p = .38$, even though this highly humanlike robot falls into the "uncanny valley."

7. Study 4: Mindless agents

Our results so far are consistent with the mere-appearance hypothesis for visual perspective taking. One alternative explanation is that physical cues led participants to attribute a more humanlike mind to an agent, and as a result of mind perception, participants felt more compelled to further infer the content of that mind, such as its distinct visual experience. To test this mind perception hypothesis, we displayed the same highly humanlike figure as in Study 3 but introduced it as an artifact that clearly lacks a mind. If mind perception is necessary for visual perspective taking, then we should observe little perspective taking toward a mind-less agent.

7.1. Method

7.1.1. Participants

We aimed to recruit 100 participants per condition for a total of six conditions and ended up with 602 participants. Among them, 83 participants (37 in the mannequin conditions and 46 in the wax figure conditions) were excluded due to failing the comprehension check question; 24 participants (14 in the mannequin conditions and 10 in the wax figure conditions) were excluded after page time screening; and three were excluded because their responses contained multiple perspectives, yielding 492 participants for data analysis ($M_{age} = 36.8$, 58.0% female; $n = 76$–85 per condition).

7.1.2. Design and procedure

We employed a 2 (agent type: mannequin, wax figure) × 3 (action: side-look, gaze, reach) between-subjects factorial design. To test the robustness of perspective taking even for agents that patently lack a mind, we informed participants that they were going to see either a mannequin or a wax figure and then presented the same side-look, gaze, and reach photographs of the android robot as in Study 3. Although we expected people to respond similarly to these two labels, we included both as a between-subjects factor to ensure that any pattern we observed was not specific to a particular type of artifact.

The procedure in Study 4 was identical to that in Study 3, except that before the experimental webpage, the agent was introduced as either a mannequin or a wax figure rather than an android robot.

Fig. 6. Photo Stimuli in Studies 3 and 4: the side-look (left), gaze (middle), and reach (right) conditions of the Erica robot, a highly humanlike android.

7.2. Results

As shown in Fig. 7, the VPT rates were high and indistinguishable when participants believed that they saw a mannequin or a wax figure (41.4% and 39.8%, respectively), $p = .87$. Moreover, people's VPT rates were higher when the agent displayed goal-directed actions (51.8%) than when it stared into space toward no particular goal object (18.7%), $z = 4.60$, $p < .001$, $d = 0.44$. In addition, the VPT rate in the reach condition (65.2%) was higher than that in the gaze condition (38.8%), $z = 3.63$, $p < .001$, $d = 0.40$.

Cross-study comparisons contrasting people's VPT rate toward the non-intelligent artifacts in the current study (40.6%) against that toward the android robot in Study 3 (37.2%) revealed no statistical difference, $z = 0.89$, $p = .37$. However, VPT rate toward the non-intelligent artifacts was even slightly higher than that toward human actors in Study 1A (33.8%), $z = 2.16$, $p = .030$, $d = 0.14$, which is a surprising difference but should be treated with caution because participants were not randomly assigned to studies/agents—an issue we addressed in Study 5.

Finally, to confirm that participants viewed the human-looking artifacts as clearly lacking mental capacities, we conducted a separate survey ($N = 300$; see Supplemental Survey 2) that measured people's inferences about two mental capacities for each agent in the current study: being capable of "looking at" the number and being capable of "making sense of" the number. Even though people may attribute a variety of mental capacities when anthropomorphizing nonhuman agents (Gray, Gray, & Wegner, 2007; Malle, 2019; Weisman et al., 2021), we featured these two mental capacities because they captured the distinct mental states of vision and knowledge (Dretske, 1969). In this survey, participants viewed each agent (e.g., the Nao robot, the cat, the android robot, the mannequin) and responded to the above items on a scale of 1 (*not at all likely*) to 7 (*very likely*). This survey confirmed that, while people indicated that the human agent had both capacities, they believed the mannequin had neither (see Table S2 in the Supplemental Materials).

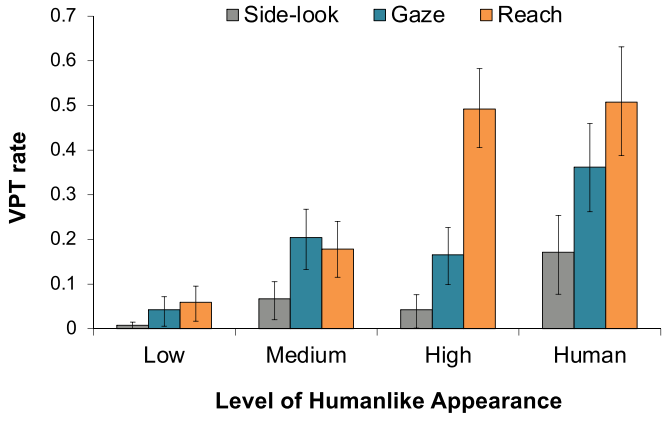

Fig. 7. Results from Study 5: Visual perspective taking (VPT) rates toward nonhuman agents of three levels of humanlikeness and human agents in the side-look, gaze, and reach conditions. Error bars represent 95% CIs with 5000 resamples.

8. Study 5: Comparisons across seven agents

So far, we have compared across studies how agents' levels of human resemblance and perceived mental capacities influenced perspective taking. One limitation of such cross-study comparisons is that participants were not randomly assigned to conditions across studies. Thus, in a preregistered study, we randomly assigned participants to one of the seven agents from previous studies and one of the three action conditions and compared their responses in the same task. We also measured each participant's perceptions of their respective agent's physical humanlikeness and mental capacities so we could assess which factor—a humanlike body or a humanlike mind—made a larger contribution to each participant's likelihood of taking the robot's perspective.

8.1. Method

8.1.1. Pre-registration

This study was pre-registered at AsPredicted.org (#47051). Data, materials, analysis, and the preregistration form are available in the OSF folder.

8.1.2. Participants

To identify a proper target sample size, we conducted power analyses based on the effect sizes in previous studies to ensure that all effects of critical interest could reach a power of 0.80 at an alpha level of 0.05. The analyses indicated that we needed at least 63 participants per cell to detect one of the smaller effects found in previous studies—a significant difference in VPT rates between side-look and the two goal-directed actions when the agent was a humanoid robot (either Nao or Baxter). Accounting for a 10% exclusion rate based on a predetermined screening procedure/attention check and a 5% exclusion rate given our standard exclusion criteria across studies, we targeted 75 participants per cell for a total of 21 cells (see below), thus 1575 participants in total. In the end, 1690 participants were recruited on Prolific and completed our study—a number slightly higher than our target due to a small error in our Qualtrics survey flow. Nonetheless, participants were randomly assigned to agent and action conditions as intended.

In line with the preregistration, we conducted a four-step data exclusion procedure: First, prior to examining participants' survey responses, we excluded 99 participants (5.9%) who failed our screener/

attention check procedure (see Procedure). Second, consistent with Studies 3 and 4, we excluded 43 participants who failed the agent identity check in conditions featuring the highly humanlike "Erica" figure (2.5% of the entire sample, or 9.5% of those in the Erica conditions). Next, we applied the same data exclusion procedure as in previous studies and excluded 100 participants (5.9%) after page times screening and 17 participants (1.0%) who responded with multiple perspectives ($n = 16$) or an irrelevant answer ($n = 1$). Hence, a total of 1431 participants remained for analysis ($M_{age} = 32.14$, $SD = 12.40$; 54.9% female, 43.6% male, 1.5% other or undisclosed; $n = 56$–73 per cell, except for $n = 93$ in the human side-look condition due to the survey flow error). Analyses that included all participants without data exclusion supported all primary findings (see Footnote 6 for the only exception).

8.1.3. Design

This study employed a between-subjects design with seven agent types (six nonhuman agents and a human) crossed with three actions (side-look, gaze, reach). The six non-human agents, featured in the previous studies, represented three *levels of humanlikeness*—low (cat, Thymio robot), medium (Nao robot, Baxter robot), and high (wax figure, android robot). All agents were displayed in photos except for the feature-less, minimally humanlike Thymio robot, which was displayed in five-second videos as in Study 2A to allow a fair chance of appearing plausibly "goal-directed." In the human agent condition, participants were randomly presented with a photo of either the female or the male actor.

8.1.4. Procedure

The study procedure was similar to that in Studies 3 and 4. Prior to the study, participants responded to two simple screening questions intended to identify suspicious Prolific accounts or inattentive participants. Next, participants were randomly assigned to one of the 21 agent/action combinations and saw a brief instruction page describing the task and the agent's identity, which read, "*On the next page, you are going to see a picture containing [a human/a cat/a robot/a wax figure/an android—a robot designed to look like a human—] and a question below the picture.*" Next, participants proceeded to the experimental page as in previous studies. After submitting their responses, those who saw one of the Erica photos also indicated whether the figure they just saw was a human, a robot, or a wax figure.

In the next part of the study, all participants saw a picture of the previously shown agent looking to the side and reported their assessment of the agent's mental capacities and degree of physical humanlikeness. For mental capacity, they responded to the same two items as in Supplemental Survey 2 (i.e., capable of "looking at" or "making sense of" the number). Participants then responded to the physical humanlikeness item, "In your opinion, to what degree does the figure in this picture look like a human?", on a 100-point slider from 0 (*not at all*) to 100 (*perfectly*). This finely grained scale was consistent with prior research on robot humanlikeness to allow accurate assessment along a wide range of appearances (Phillips et al., 2018). We preregistered this part of the survey as exploratory.

Finally, participants answered two demographic questions (gender and age) and received their payment codes for a compensation of $0.25. The entire study took 1 to 2 min to finish.

8.2. Results

Our analyses focused on three primary predictions of the mere-appearance hypothesis: First, we conducted a factorial analysis to test whether participants spontaneously took robots' visual perspectives in patterns that largely mirrored perspective taking toward other humans (i.e., triggered by goal-directed actions). Second, we examined whether VPT rates increased monotonically with increasingly humanlike appearance instead of descending into an "uncanny valley." Third, we tested whether it was superficial human resemblance, or the perception of a humanlike mind, that drove people's perspective-taking responses.

8.2.1. Appearance and action

As preregistered, we conceptualized the six non-human agents as falling into low, medium, and high levels of humanlikeness. Examining people's humanlikeness ratings on these agents confirmed that Thymio and the cat were indeed perceived to be the least humanlike (Thymio: $M = 3.54$, $SD = 10.99$; cat: $M = 8.47$, $SD = 15.16$). By contrast, Nao and Baxter were perceived to be somewhat more humanlike (Nao: $M = 17.91$, $SD = 18.24$; Baxter: $M = 24.36$, $SD = 22.21$). And the "Erica" figure, when presented as either an android robot or a wax figure, was indeed perceived to have the highest level of physical humanlikeness among all non-human agents (android: $M = 53.64$, $SD = 23.20$; wax figure: $M = 58.46$, $SD = 22.99$). Therefore, we collapsed pairs of agents at the same level of humanlikeness to construct a 4 (level of humanlikeness: low/medium/high humanlikeness, and actual human) $\times$ 3 (actions: side-look, gaze, reach) factorial analysis on participants' perspective-taking responses.

The impact of humanlike appearance on visual perspective taking was in line with the mere-appearance hypothesis: Collapsed across action conditions, the moderately humanlike robots of Nao and Baxter elicited a higher VPT rate (15.1%) than the minimally humanlike agents of Thymio robot and the cat (3.6%), $z = 5.60$, $p < .001$, $d = 0.39$, yet lower VPT rates than the highly humanlike android robot and the wax figure (23.7%), $z = 6.52$, $p < .001$, $d = 0.46$. Furthermore, the VPT rate toward a human agent (34.2%) was higher than that toward the android robot and the wax figure (i.e., 23.7%), $z = 7.44$, $p < .001$, $d = 0.62$.

Examining the effect of action, we once again observed that across all agents, goal-directed actions of gaze and reaching elicited higher VPT rates (an average of 22.5%) than did looking to the side (5.7%), $z = 5.41$, $p < .001$, $d = 0.31$. In addition, seeing the agent reaching toward the number object also elicited higher VPT rates (27.6%) than seeing the gaze alone (17.7%), $z = 3.10$, $p = .002$, $d = 0.20$. Examining the interaction terms revealed that the contrast of medium vs. high humanlikeness significantly interacted with the gaze vs. reach contrast, $z = 2.65$, $p = .008$, $d = 0.23$, indicating that people responded to the reach action with an even higher VPT rate than gaze when the agent's appearance became highly humanlike.[6] None of the other interaction terms reached statistical significance, $ps > 0.42$.

8.2.2. Correlating VPT rates with physical appearance and mind perception

The factorial analysis above compared VPT rates by first combining pairs of agents into levels of humanlikeness. In an exploratory analysis, we treated each agent as a distinct data point and explored the relationship between the VPT rate each agent elicited and its average ratings of humanlikeness and mental capacities (see VPT rates and mean ratings by each agent in Table S2 in Supplemental Materials). As shown in Fig. 8 (a), VPT rates exhibited a monotonic, even linear increase as the agent's physical appearance became more humanlike, $r = 0.96$, $p < .001$. By contrast, Figs. 8(b) and 8(c) indicate that VPT rates were barely correlated with the extent to which an agent was perceived to be capable of either looking at ("vision capacity") or making sense of ("representation capacity") the number ($rs = -0.04$ and 0.34, $ps = 0.93$ and 0.45, respectively). Despite being conducted on a small group of seven agents, such results are clearly consistent with the mere-appearance hypothesis and inconsistent with the uncanny valley or the mind perception hypothesis.

8.2.3. Comparing android robot versus wax figure

With random assignment across agents, we also replicated the cross-

[6] This unexpected interaction effect became only marginally significant when we included all responses without data exclusion, $z = 1.62$, $p = .10$ (see Supplemental Materials).

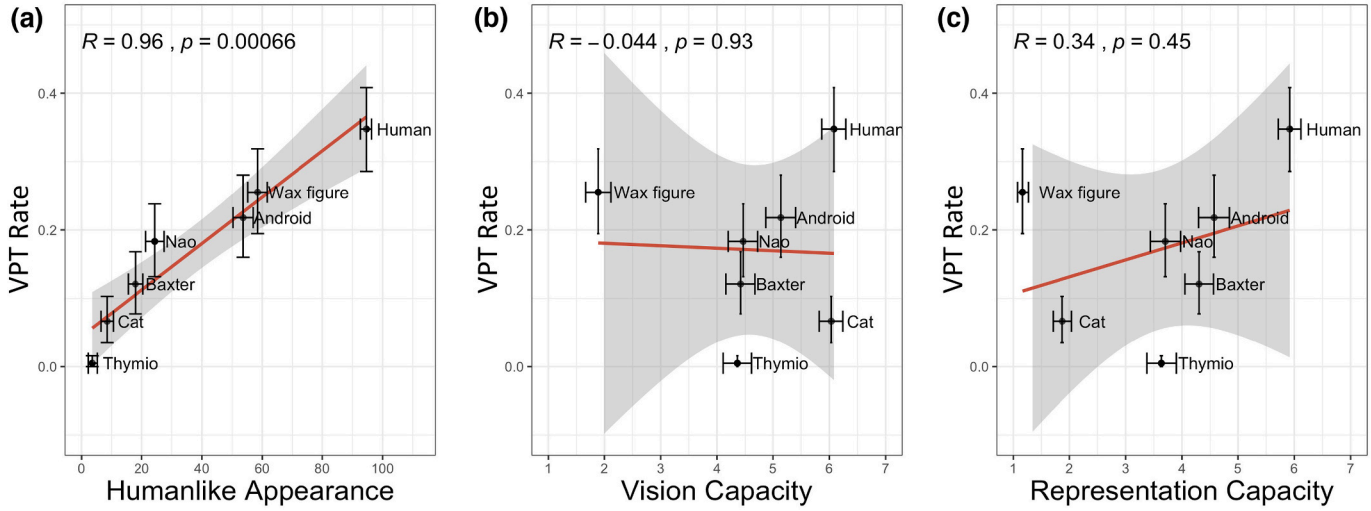

Fig. 8. VPT rates elicited by seven agents (Y-axis) plotted against each agent's mean ratings of (a) physical humanlikeness, (b) mental capacity of vision ("looking at the number"), and (c) mental capacity of forming representations ("making sense of the number"). Error bars indicate 95% CIs (with 5000 resamples) around the means of VPT rates (vertical bars) and item ratings (horizontal bars). Red lines indicate the best-fitting linear regression lines. Shaded regions represent 95% CIs around regression lines. Correlation statistics are shown in the upper left corner. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

study comparison between Studies 3 and 4, finding that people took the perspective of a mindless wax figure and a state-of-the-art artificial intelligent robot at similar rates (25.5% and 21.8%, respectively), $z = 0.84$, $p = .40$. In addition, confirming the validity of our manipulation, people expected the wax figure to have much less mental capacity to look at the number ($M = 1.89$, $SD = 1.62$) than the android robot ($M = 5.14$, $SD = 1.79$), $t = -18.28$, $p < .001$, $d = -1.91$, and they expected the wax figure to have virtually no capacity of making sense of the number ($M = 1.16$, $SD = 0.69$) compared to the android robot ($M = 4.57$, $SD = 1.86$), $t = -23.06$, $p < .001$, $d = -2.46$. In fact, the wax figure received the lowest mental capacity ratings among all seven agents (see Table S3).

8.2.4. Predicting perspective taking in a participant-level analysis

Finally, we examined to what extent physical appearance and mind perception could predict individual participants' perspective-taking responses. To this end, we preregistered two multiple logistic regression analyses with participants' ratings on their agent's physical humanlikeness and either its looking capacity or representation capacity as predictors. (We did not include all three items in the same model given that we expected people's ratings on the two mental capacities to be highly correlated—which was supported by the data, $r = 0.53$, $p < .001$—thus making either capacity even less likely to show a unique contribution in the regression model.) In light of different scales, we standardized participants' ratings on all three items prior to data analysis.

Our results showed that people's perceptions of an agent's humanlike appearance strongly predicted their perspective-taking responses to that agent ($B = 0.64$, $z = 8.98$, $p < .001$), whereas beliefs in its capacity to look at the number showed no predictive power ($B = 0.03$, $z = 0.45$, $p = .66$). Similarly, the second model showed that perceptions of humanlikeness again strongly predicted the same agent's perspective taking ($B = 0.63$, $z = 8.25$, $p < .001$), whereas belief in the agent's representation capacity did not ($B = 0.06$, $z = 0.74$, $p = .46$). Even though regression analyses are correlational in nature, perception of humanlike appearance is rooted in the agent's physical properties and is unlikely to have been caused by perspective taking. These results therefore support the hypothesis that it was humanlike appearance, rather than mind perception, that largely drove participants' inclination to take an agent's unique vantage point.

8.3. Discussion

In a preregistered study with full random assignment and a sufficiently large sample size, we observed converging evidence supporting the mere-appearance hypothesis of perspective taking toward robots: First, people are more inclined to take a humanlike robot's visual perspective when it displays goal-directed actions; second, such responses increase monotonically with humanlike appearance as opposed to plummeting into an uncanny valley; third, perspective taking is strongly driven by superficial human resemblance rather than by perceiving a humanlike mind.

9. General discussion

9.1. The unique impact of humanlike appearance

As impending members of modern society, robots can trigger social-cognitive mechanisms foundational to human social interaction. Specifically, a robot's humanlike appearance can lead people to spontaneously (i.e., unprompted) consider how an object would appear from the robot's vantage point, and people do so in patterns of variation that mirror perspective taking toward other humans (Studies 1A and 1B; Study 5). More strikingly, perspective taking is minimal when the agent lacks human appearance (Studies 2A and 2B), increases as the agent appears increasingly humanlike, and persists even when the agent seems eerie (android robot; Study 3) or obviously lacking a mind (mannequin or wax figure; Study 4). These results suggest that visual perspective taking toward robots is consistent with the mere appearance hypothesis rather than following an uncanny valley pattern, and that perspective taking can be triggered even without explicitly attributing a mind to the agent. These results update our first documentation of people's spontaneous perspective taking toward robots (Zhao, Malle, & Gweon, 2016), where we assumed that perspective taking revealed people's attribution of mind-like qualities to robots—an interpretation that has since been cited in other articles. But we now revise our stance in light of the finding that mere appearance has a unique impact on visual perspective taking and can still occur even when people do not explicitly believe that an agent has a mind.

Given that people often need extra effort to overcome their own egocentric viewpoint (Epley, Keysar, Van Boven, & Gilovich, 2004), it

may seem counterintuitive that participants would ever take the perspective of a robot, let alone a mind-less mannequin. Thus, one might wonder if such responses could have resulted from "experimenter demand"—from participants' attempts to see the number through the eyes of the experimenter who staged the mannequin, created the photographs, and designed the study. However, this alternative interpretation is not consistent with our results: The experimenter staged all the agents, including the novel guitar (Study 1A), the Thymio robot (Study 2A), and the cat (Study 2B), yet perspective taking was considerably lower in those cases. If perspective-taking responses were primarily driven by people's inferences of what the experimenter wanted the participant to say, then we would expect people to infer the perspective of the novel agents, such as a cat reaching for a number, which is such an unusual sight and must "mean" something. While it is certainly possible to conceive of elaborate explanations that may fit just the patterns we have observed (e.g., "participants think the experimenter thinks participants should take the perspective of a mannequin but not a cat"), we believe that positing the impact of humanlike appearance is both more parsimonious and generates novel predictions rather than fitting the data post hoc.

Hence, a humanlike robot can jump-start certain basic social-cognitive mechanisms without being seen as human-minded. This finding contributes to a growing literature indicating that people sometimes extend familiar patterns of social behaviors to novel technologies even if those technologies do not have humanlike appearance. For instance, the computer as social actor (CASA) theory, pioneered in the human-computer interaction literature, suggests that people readily apply social rules, scripts, and expectations when engaging with novel technologies even while fully recognizing that they are void of humanlike mental qualities (Nass & Moon, 2000; Reeves & Nass, 1996). Understanding when physical appearance, mind perception, and affective responses influence people's responses to robots is critical for achieving an integrative framework that can predict how people will engage with robots as they arrive in our society.

The unique impact of mere appearance does not imply, however, that mind perception or the uncanny valley are irrelevant to visual perspective taking. One central tenet of the "mind perception" hypothesis is that people attribute different amounts of mental capacities to different agents—a core mechanism that is believed to underlie a wide range of phenomena from dehumanizing other people to anthropomorphizing inanimate objects (Epley & Waytz, 2010; Gray, Young, & Waytz, 2012; Haslam, 2006; Schroeder & Epley, 2020; Wheatley, Kang, Parkinson, & Looser, 2012). Our results suggest that mind attribution may not be the only route to responding socially toward robots, as a "skin-deep" resemblance in appearance can also have a powerful impact, but they do not preclude the possibility that mind perception may exert a top-down influence on perspective taking in more dynamic, communicative encounters. For instance, after seeing a mechanical-looking robot navigate its surroundings in a thoughtful manner, or hearing that it does not know numbers and wants to learn the number in front of it, it is possible that people might become more inclined to describe the number from its perspective due to ascribing a mind and certain intentions to the agent (Zhao et al., 2016). Likewise, extended interactions with a robot are likely to generate affective impressions (e.g., of comfort or eeriness), and if a robot is at the edge of the uncanny valley, prolonged feelings of eeriness may well hinder perspective taking responses and prevent people from treating the robot as a social partner.

9.2. Limitations of the paradigm

Our research was able to isolate mind perception from physical appearance by introducing the mannequin and by offering statistical dissociation in Study 5's participant-level regression analysis. However, we could not entirely disentangle mind and appearance across all agents because appearance often influences the amount of mind people perceive in an agent (Zhao et al., 2019). Nonetheless, other recent studies have also found that people's social-cognitive responses to robots are dissociable from their explicit mind attribution to the agents (Banks, 2020; Złotowski et al., 2018). So future research will need to clarify how top-down and bottom-up processes may jointly influence the activation of perspective taking toward robots of various appearances.

The current paradigm is unable to pinpoint by what mechanism appearance induces perspective taking. On one end of possibilities, responses to humanlike appearance may be innate, automatic, and reflexive—similar to how a goose seeks to protect itself upon seeing a hawk-shaped silhouette (Schleidt, Shalter, & Moura-Neto, 2011). A form of generalization (Shepard, 1987) from truly human appearance to human*like* appearance may serve as one important mechanism underlying social responses toward robots. On the other end of possibilities, such responses may arise from repeated experience with artifacts of humanlike appearance (e.g., barbie dolls, cartoon characters) in which certain social practices and expectations have been overlearned—just like how people have learned to associate colors at a traffic light with certain appropriate behaviors. And of course the truth may lie somewhere in between. Understanding which mechanisms drive people's perspective-taking response to robots will not only illuminate the social-cognitive underpinnings of human-robot interaction but may also avert the danger of inferring capacities in highly humanlike robots that they simply do not have (Malle, Fischer, Young, Moon, & Collins, 2020).

The present paradigm also does not clarify what specific psychological state the observer is engaged in when taking the robot's perspective. It seems clear that, when answering "6" from the other agent's vantage point, participants are not actually *having* the other agent's visual experience (Cole & Millett, 2019; Samuel, Hagspiel, Eacott, & Cole, 2021); after all, robots do not have visual "experiences" to begin with, and a mannequin certainly cannot see. But people may be projecting themselves into the agent's physical position and imagining how the world would appear from the other agent's vantage point (Kessler & Rutherford, 2010; Ward et al., 2019). Whatever the proximal psychological process is, it normally encourages people to take a step toward engaging with the other agent, perhaps treating it as worthy of collaboration.

Finally, like other researchers who study similar phenomena (e.g., Freundlieb, Kovács, & Sebanz, 2018), we have called the kind of perspective taking induced by the current paradigm "spontaneous" in its conventional meaning of "unprompted," which clearly differentiates the present phenomenon from experimenter-instructed perspective taking (e.g., Michelon & Zacks, 2006; Myers, Laurent, & Hodges, 2014). We believe unprompted perspective taking occurs naturally in social interaction and appears early in life (Southgate, 2020), but we do not suggest it is automatic, effortless, or uncontrollable. In fact, participants had to choose the other agent's perspective over their own when offering their verbal responses, and the process of selecting and verbalizing another person's perspective is known to be effortful and deliberate (Epley et al., 2004; Qureshi, Apperly, & Samson, 2010).

9.3. Implications for the perspective taking literature

Our research primarily aims to understand how people make sense of humanlike machines. However, we believe it can also inform ongoing debates in the perspective taking literature more broadly.

First, our finding suggests that the debate on whether cognitive processes underlying perspective taking are social or nonsocial may be a simplification. Some forms of perspective taking, which researchers take to be nonsocial, can be triggered by superficial cues such as arrows and chairs (e.g., Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014; Westra, Terrizzi, van Baal, Beier, & Michael, 2021). Participants in our studies also responded to a "superficial" characteristic of humanlike appearance (even in the case of a mind-less mannequin), but this characteristic may still count as social because it reflects people's responsiveness to a reliable indicator of human social agents — their appearance.

Second, our finding supports the distinction between two questions about perspective taking: under what conditions it is initiated, and whether—if initiated—it is accurate. In the classic "Three Mountain Task" (Piaget & Inhelder, 1956), for example, children at a certain age take the "perspective" of a lifeless doll, but they do not have an accurate understanding of what the other is seeing or thinking, because the doll has no perceptual experiences. Likewise, many people in our study took the perspective of a mannequin, which also cannot see. And in recent studies, too, adult participants tried to adopt other people's perspectives when prompted by study instructions, but such efforts did not reliably increase people's accuracy in predicting others' thoughts, feelings, attitudes, or other mental states (e.g., Eyal, Steffel, & Epley, 2018). Therefore, perspective taking can involve the inferential process of projecting oneself into another agent's unique position and predicting how the world would appear from their vantage point; it does not guarantee, however, an *accurate* understanding of their actual mental operations.

Finally, our research highlights that perspective taking is not an on-or-off phenomenon and supports studying the *extent* to which certain stimuli properties can elicit perspective taking. Clearly, arrows and empty chairs may have a statistically above-zero chance to elicit *some* perspective taking (e.g., Millett, D'Souza, & Cole, 2020; Santiesteban et al., 2014), but our research shows that a more humanlike figure has a higher likelihood of triggering perspective taking than a less humanlike figure. Therefore, future research needs to go beyond offering existence proofs of perspective taking and examine instead whether different circumstances trigger people's perspective-taking responses to different degrees, and what cognitive processes may underlie those responses.

9.4. Concluding thoughts

The generalization of social-cognitive processes originally acquired for human social interaction to humanlike robots raises many theoretically intriguing and practically important questions: What other social-cognitive mechanisms can a robot's humanlike appearance trigger? Are there limits to this triggering power? And what consequences will such responses have for human-robot interaction and for our society? Robots' human resemblance is desirable when it facilitates coordination and communication. However, if humanlike appearance also elicits other patterns of social behavior, such as emotional attachments, this might leave people vulnerable to potential manipulation in relationships (Scheutz, 2011). Therefore, knowledge of the psychological impact of humanlike appearance must lead to more informed decisions about what kind of robots—with what kind of appearance—we should usher into our society.

Open science practices

All stimuli, data, analysis (R code), and output files for all studies have been made available via the Open Science Framework (OSF) and can be accessed at https://osf.io/ymkqp/.

Author contributions

X.Z. and B.F.M. conceptualized and designed the study; X.Z. collected and analyzed the data; X.Z. and B.F.M. interpreted the data; B.F.M. acquired research funding; X.Z. and B.F.M wrote the manuscript.

Declaration of Competing Interest

The authors report no conflicts of interest.

Acknowledgments

This project was supported in part by a grant from the Office of Naval Research, No. N00014-14-l-0144. The opinions expressed here are our own and do not necessarily reflect the views of ONR.

We thank the following people and laboratories for generously contributing their robots for stimulus production: Matthias Scheutz's Human-Robot Interaction Laboratory at Tufts University for the Nao robot; Stefanie Tellex's Humans to Robots Laboratory at Brown University for the Baxter robot; and Hiroshi Ishiguro Laboratories for the Erica robot. We thank Megan Strait, Emily Wu, Hidenobu Sumioka, and Daniel Ullman for assisting with producing photo and video stimuli featuring Nao, Baxter, Erica, and Thymio, respectively. And we thank Daniel Platt for introducing the research team to his adorable cat, Koji.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2022.105076.

References

Banks, J. (2020). Theory of mind in social robots: Replication of five established human tests. *International Journal of Social Robotics, 12*(2), 403–414.
Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics, 3*, Article eaat5954.
Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding. In B. H. Ross (Ed.), *Vol. 53. The psychology of learning and motivation* (pp. 301–344). Burlington: Academic Press.
Broadbent, E. (2017). Interactions with robots: The truths we reveal about ourselves. *Annual Review of Psychology, 68*, 627–652.
Cole, G. G., & Millett, A. C. (2019). The closing of the theory of mind: A critique of perspective-taking. *Psychonomic Bulletin & Review, 26*(6), 1787–1802.
Cosmides, L., & Tooby, J. (1997). Evolutionary psychology: A primer. Retrieved June 13, 2019, the University of California, Santa Barbara, Center for Evolutionary Psychology website: http://www.psych.ucsb.edu/research/cep/primer.html.
Dear, K., Dutton, K., & Fox, E. (2019). Do 'watching eyes' influence antisocial behavior? A systematic review & meta-analysis. *Evolution and Human Behavior, 40*(3), 269–280.
Dretske, F. (1969). *Seeing and knowing*. London: Routledge.
Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*(3), 327–339.
Epley, N., & Waytz, A. (2010). Mind perception. In S. T. Fiske, & D. T. Gilbert (Eds.), *The handbook of social psychology* (pp. 498–541). John Wiley & Sons, Inc.
Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886.
Erle, T. M., & Topolinski, S. (2017). The grounded nature of psychological perspective-taking. *Journal of Personality and Social Psychology, 112*(5), 683–695.
Eyal, T., Steffel, M., & Epley, N. (2018). Perspective mistaking: Accurately understanding the mind of another requires getting perspective, not taking perspective. *Journal of Personality and Social Psychology, 114*(4), 547–571.
Flavell, J. H. (2004). Development of knowledge about vision. In D. T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children* (pp. 13–36). Cambridge, MA: MIT Press.
Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2018). Reading your mind while you are reading—Evidence for spontaneous visuospatial perspective taking during a semantic categorization task. *Psychological Science, 29*(4), 614–622.
de Graaf, M. M. A., & Malle, B. F. (2019). People's explanations of robot behavior subtly reveal mental state inferences. In *Proceedings of the international conference on human-robot interaction, HRI '19* (pp. 239–248). New York, NY: IEEE Press.
Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619.
Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition, 125*(1), 125–130.
Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry, 23*(2), 101–124.
Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology, 51*(1), 79–88.
Hamilton, A. F., Brindley, R., & Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition, 113*, 37–44.
Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10*(3), 252–264.
Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57*(2), 243–259.
Herrmann, E., Call, J., Hernàndez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science, 317*, 1360–1366.
Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science, 344*(6190).
Johnson, S. C., Slaughter, V., & Carey, S. (1998). Whose gaze will infants follow? The elicitation of gaze-following in 12-month-olds. *Developmental Science, 1*(2), 233–238.
Kessler, K., & Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology, 1*(December), 1–12.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*(1), 32–38.

Kim, B., Bruce, M., Brown, L., de Visser, E., & Phillips, E. (2020). A comprehensive approach to validating the uncanny valley using the Anthropomorphic roBOT (ABOT) database. In *2020 systems and information engineering design symposium (SIEDS)* (pp. 1–6). IEEE.

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), Article e2597.

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.

Malle, B. F. (2019). How many dimensions of mind perception really are there? In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st annual meeting of the cognitive science society* (pp. 2268–2274). Montreal, QB: Cognitive Science Society.

Malle, B. F., Fischer, K., Young, J. E., Moon, A., & Collins, E. C. (2020). Trust and the discrepancy between expectations and actual capabilities of social robots. In D. Zhang, & B. Wei (Eds.), *Human-robot interaction: Control, analysis, and design* (pp. 1–23). Cambridge Scholars Publishing.

Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review, 99*(4), 587–604.

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition, 146*, 22–32.

McCurry, J. (2015, December 31). *Erica, the "most beautiful and intelligent" android, leads Japan's robot revolution.* The Guardian. Retrieved from https://www.theguardian.com/technology/2015/dec/31/erica-the-most-beautiful-and-intelligent-android-ever-leads-japans-robot-revolution.

Meltzoff, A. N., Brooks, R., Shon, A. P., & Rao, R. P. N. (2010). "Social" robots are psychological agents for infants: A test of gaze following. *Neural Networks, 23*(8–9), 966–972.

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & Psychophysics, 68*(2), 327–337.

Millett, A. C., D'Souza, A., & Cole, G. G. (2020). Attribution of vision and knowledge in 'spontaneous perspective taking. *Psychological research, 84*(6), 1758–1765.

Moll, H., & Meltzoff, A. N. (2011). How does it look? Level 2 perspective-taking at 36 months of age. *Child Development, 82*(2), 661–673.

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

Myers, M. W., Laurent, S. M., & Hodges, S. D. (2014). Perspective taking instructions and self-other overlap: Different motives for helping. *Motivation and Emotion, 38*(2), 224–234.

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*(4), 435–450.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103.

Phillips, A. T., & Wellman, H. M. (2005). Infants' understanding of object-directed action. *Cognition, 98*(2), 137–155.

Phillips, E., Zhao, X., Ullman, D., & Malle, B. F. (2018). What is human-like?: Decomposing robots' human-like appearance using the anthropomorphic roBOT (ABOT) database. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction (HRI'18)* (pp. 105–113).

Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London, UK: Routledge & Kegan Paul.

Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition, 36*, 1–16.

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition, 117*(2), 230–236.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., … Wellman, M. (2019). Machine behaviour. *Nature, 568*(7753), 477–486.

Reeves, B., & Nass, C. (1996). *The media equation*. Cambridge, UK: Cambridge University Press.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5), 1255–1266.

Samuel, S., Hagspiel, K., Eacott, M. J., & Cole, G. G. (2021). Visual perspective-taking and image-like representations: We don't see it. *Cognition, 210*, Article 104607.

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance, 40*(3), 929–937.

Scheutz, M. (2011). The inherent dangers of unidirectional emotional bonds between humans and social robots. In P. Lin, G. Bekey, & K. Abney (Eds.), *Robot ethics: The ethical and social implications of robotics* (pp. 205–221). Cambridge, MA: MIT Press.

Schleidt, W., Shalter, M. D., & Moura-Neto, H. (2011). The hawk/goose story: The classical ethological experiments of Lorenz and Tinbergen, Revisited. *Journal of Comparative Psychology, 125*(2), 121–133.

Schroeder, J., & Epley, N. (2020). Demeaning: Dehumanizing others by minimizing the importance of their psychological needs. *Journal of Personality and Social Psychology, 119*(4), 765–791.

Shepard, R. N. (1987). Towards a universal theory of generalization for psychological science. *Science, 237*(4820), 1317–1323.

Sodian, B., & Thoermer, C. (2004). Infants' understanding of looking, pointing, and reaching as cues to goal-directed action. *Journal of Cognition and Development, 5*(3), 289–316.

Southgate, V. (2020). Are infants altercentric? The other and the self in early social cognition. *Psychological Review, 127*(4), 505.

Surtees, A., Apperly, I., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition, 150*, 43–52.

Trafton, J. G., Cassimatis, N. L., Bugajska, M. D., Brock, D. P., Mintz, F. E., & Schultz, A. C. (2005). Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics —Part A: Systems and Humans, 35*(4), 460–470.

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition, 110*(1), 124–129.

Urgen, B. A., Plank, M., Ishiguro, H., Poizner, H., & Saygin, A. P. (2013). EEG theta and Mu oscillations during perception of human and robot actions. *Frontiers in Neurorobotics, 7*, 1–13.

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology, 19*(4), 393–407.

Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. *Current Biology, 29*, 874–880.

Waytz, A., & Norton, M. I. (2014). Botsourcing and outsourcing: Robot, British, Chinese, and German workers are for thinking-not feeling-jobs. *Emotion, 14*(2), 434–444.

Weisman, K., Legare, C. H., Smith, R. E., Dzokoto, V. A., Aulino, F., Ng, E., & Luhrmann, T. M. (2021). Similarities and differences in concepts of mental life among adults and children in five cultures. *Nature Human Behaviour*, 1–11.

Westra, E., Terrizzi, B. F., van Baal, Beier, J. S., & Michael, J. (2021). Beyond avatars and arrows: Testing the mentalising and submentalising hypotheses with a novel entity paradigm. *Quarterly Journal of Experimental Psychology, 74*(10), 1709–1723.

Wheatley, T., Kang, O., Parkinson, C., & Looser, C. E. (2012). From mind perception to mental connection: Synchrony as a mechanism for social understanding. *Social and Personality Psychology Compass, 6*(3), 589–606.

Zhao, X., Cusimano, C., & Malle, B. F. (2015). In search of triggering conditions for spontaneous visual perspective taking. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 2811–2816).

Zhao, X., Malle, B. F., & Gweon, H. (2016). Is it a nine, or a six? Prosocial and selective perspective taking in four-year-olds. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 924–929). Austin, TX: Cognitive Science Society.

Zhao, X., Phillips, E., & Malle, B. F. (2019). How people infer a humanlike mind from a robot body. PsyArXiv. https://doi.org/10.31234/osf.io/w6r24

Złotowski, J., Ishiguro, H., Sumioka, H., Eyssel, F., Nishio, S., & Bartneck, C. (2018). Model of dual anthropomorphism: The relationship between the media equation effect and implicit anthropomorphism. *International Journal of Social Robotics, 10*(5), 701–714.