More Than Meets the Eye? An Experimental Design to Test Robot Visual Perspective-Taking Facilitators Beyond Mere-Appearance

Joel Currie
j.currie.22@abdn.ac.uk
University of Aberdeen
Aberdeen, United Kingdon

Katrina Louise McDonough
Katrina.Mcdonough@uea.ac.uk
University of East Anglia
Norwich, United Kingdom

Agnieszka Wykowska
agnieszka.wykowska@iit.it
Istituto Italiano di Tecnologia
Genova, Italy

Maria Elena Giannaccini
elena.giannaccini@abdn.ac.uk
University of Aberdeen
Aberdeen, United Kingdon

Patric Bach
patric.bach@abdn.ac.uk
University of Aberdeen
Aberdeen, United Kingdon

ABSTRACT

Visual Perspective Taking (VPT) underpins human social interaction, from joint action to predicting others' future actions and mentalizing about their goals and affective/mental states. Substantial progress has been made in developing artificial VPT capabilities in robots. However, as conventional VPT tasks rely on the (non-situated, disembodied) presentation of robots on computer screens, it is unclear how a robot's socially reactive and goal-directed behaviours prompt people to take its perspective. We provide a novel experimental paradigm that robustly measures the extent to which human interaction partners take a robot's visual perspective during face-to-face human-robot-interactions, by measuring how much a robot's visual perspective is spontaneously integrated with one's own. The experimental task design of our upcoming user study allows us to investigate the role of robot features beyond its human-like appearance, which have driven research so far, targeting instead its socially reactive behaviour and task engagement with the human interaction partner.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; • **Applied computing** → **Psychology**; • **Computing methodologies** → **Cognitive science**.

KEYWORDS

Human-Robot Interaction; Perspective-Taking; non-verbal behaviours; humanoid robot; mind perception

1 INTRODUCTION

Successful human-robot collaboration relies on the complex interplay of multiple mechanisms [4, 14, 22]. Imagine you are building a bookshelf with your robot colleague and they ask you to pass them the screwdriver on the left. For this simple request to be fluid, both agents need to not only be capable of knowing what the other can see, but how they see it. One needs to take the robot's viewpoint to discern which screwdriver is being requested, and that what is to the left or the robot is to the right of yourself. This ability for visual perspective taking (VPT) helps us navigate our social environment [15], participate in joint action [10] and predict the next actions of others [2, 8]. Accumulating evidence suggests that VPT could also underpin more complex abilities to reason about others' goals [11] and affective/mental states [3, 19]. While impressive progress has been made developing VPT abilities in artificial agents, such that robots can take their human partners' perspective [1, 7, 9, 16, 17, 33], not much is known about how people take a robot's perspective, and which robot features would help them to do so. Humans do not have specialised social-cognitive mechanisms for interacting with robots. Instead the existing mechanisms for interaction with other humans are most likely repurposed [40], [25]. Understanding the feature space that triggers these repurposed mechanisms [12, 13] is critical for designing understandable robots that humans can interact with fluidly.

A problem for research on VPT is that most available tasks rely on measuring whether people take the perspective of robots on the computer screen - either as cartoons, photos, or short movie clips - instead of real-life interactions. These studies have suggested that whether people spontaneously take a robot's perspective primarily depends on its human-like appearance. Prior research has shown, for example, that people often describe numbers on table how they appear to another person sitting opposite to them, rather than to themselves (e.g., they denote the number 6 as a 9, if it looks like this to the other person, [31]). For robots, this tendency to perspective-take increased the more the robot's appearance was human-like, but was surprisingly unrelated to the extent of that it was attributed a "mind" or an ability to represent visual input [45]. Other studies [35, 36] came to similar conclusions. They relied on the phenomenon that people find it more difficult to report a number of dots in a room, if another agent within the room can, from their perspective, not see all of the dots [27]. For robot agents, this interference from another's perspective only occurred when the robot had a human-like body, but not a cat-like body. However, it was independent of whether the

robot's head was human-like or not (a camera head), and independent of whether the robot was on or off (indicated by a red light).

It is striking that the "mere appearance" [35, 36, 45] of a robot as human-like suffices to cause people to spontaneously take its perspective, and that higher-order factors such as the ability to represent the environment or its "mind" contribute little, if at all. We hypothesize that this remarkable state of research might reflect, to a large extent, the artificial and dis-embodied mode of presentation on a computer screen, in which robot and human do neither interact nor share a common task space. Indeed, there is evidence that objective measures of anthropomorphism increase for physically embodied robots [26], and that the tendency to attribute mental states to robots is modulated by socially interactive behaviour [32]. Moreover, there is suggestive evidence that action matters for perspective taking. Using the same 6/9 task as in their study above, Zhao [44] reported that people more likely take the robot's perspective when they saw the robot reach out towards the numbers on the table (instead of merely gazing at it), but only if this action was shown as video (instead of a photo). Thus, while this still falls short of full-scale interaction, perspective taking seems to increase the more people perceive the robot as a realistic agent that acts within its environment.

To the best of our knowledge, only one study investigated perspective taking in a real human-robot interaction [42], but was not able to demonstrate any perspective-taking. Two robots (telenoids [23]) were seated around a table, either side of the participant. One telenoid was inanimate; the other looked and waved to the participant at the beginning of the experiment but remained otherwise passive. Participants had to indicate whether letters that were oriented either towards the animate or inanimate telenoid were mirrored or not. However, instead of the predicted overall advantage when letters were facing the animate robot, they only found that, for the previously animate agent, the response time difference between mirrored and non-mirrored letters increased - a difference that was not pre-registered, inconsistent with the usual perspective-taking results in such paradigms [37–39], and therefore hard to interpret. In essence, it may only reflect that participants are biased to call a letter non-mirrored if it faces the animate agent, irrespective of whether they take its perspective.

Together, these studies make clear that the study of human-to-robot perspective taking requires real-life tasks that are not decoupled from the intricate behavioural dynamics when interacting with a robot embodied in physical space. Moreover, it requires tasks that are not affected by confounds and alternative explanations. For example, in the 6/9 task, just *knowing* that a 6 looks like a 9 to others could be enough to induce seemingly altercentric responses, without the participant visualizing how the robot perceives the environment [28]. Similarly, as the authors acknowledge [36], the results in the dot perspective task could simply reflect participants' attention being cued to the dots the robot is "looking" at, instead of any representation of what it can see [5].

Here, we present a task that overcomes the limitations of prior work on VPT towards robots [35, 36, 44, 45] and can reliably measure perspective taking in face-to-face interactions. It relies on the well-established phenomenon that response times to identify a letter increase linearly the more the letter is oriented away from the participant (i.e., from 0 degrees angular disparity to 180 degrees), as participants first have to "mentally rotate" turned away letters

into their canonical orientation [29]. Our previous work [37–39] shows that this pattern is disrupted by another agent in the scene. People recognize rotated-away letters more quickly if the letters are oriented towards the other agent and can therefore more easily be identified from their perspective. In contrast, they recognize rotated-away letters more slowly if they are rotated even further away from the other agent's perspective. These findings show both, that people spontaneously "borrow" another's perspective if it provides better visual access to a stimulus than ones own, and that this perspective also involuntarily intrudes into their own and disrupts visual processing.

Our task is well-validated for human-to-human interaction when presented as short photo animations on a computer screen [37–39, 43], or in virtual reality [41]. After extensive piloting for human-human VPT, here we adapt it to investigate face-to-face interactions with a humanoid robot to test how factors beyond mere-appearance, such as goal-driven and autonomous socially reactive behaviours, facilitate perspective taking. Importantly, our task measures both, the facilitation and disruption of perceptual processing through VPT; unlike other tasks [5, 28], its findings can therefore not be explained through simple heuristics or attentional cuing.

2 METHODS AND MEASURES

2.1 Task and Experimental Conditions

Participants sit at a table, with a robot to their left or right. In each trial, an alphanumeric character is projected onto the table in-front of the participant in either its normal, or mirror inverted form. The character is shown at differing orientations from upright to upside down relative to the participants viewpoint. Participants simply judge with a button press whether the stimulus is in its canonical or mirror-inverted form ("R" vs. "Я"). We vary between experimental blocks whether the participants' partner (a robot with different behaviours - see Figure 1) is positioned 90° to the right, or 90° to the left of them. In our prior research, individuals spontaneously incorporate another person's viewpoint when making these judgements. When letters are facing another person, people find it easier (measured in terms of recognition times, RTs) to judge whether the character is in its regular or mirrored presentation. Conversely when the letter is facing away from another person, individuals find the task more difficult [37–39]. This suggests that people are spontaneously assimilated their own egocentric viewpoint with a generative model simulating a likely estimation of the other person's altercentric frame of reference.

Here, we test whether people take the perspective of a physically present humanoid robot, and whether this depends on the robot's socially reactive behaviours in unconstrained human-robot interactions and/or its goal-directed engagement in a task alongside the participant, two factors that have been shown to be important for human-to-human perspective taking [10]. In a between-subjects design, participants are randomly assigned to one of four experimental conditions that arise from the factorial combination of the Task Reactivity and Social Reactivity factors (see Figure 1). Thus, across participants, the robot will be in one of four states: (i) Being completely inanimate in and outside the task, (ii) actively completing the task alongside the participant but being passive outside of it, (iii) being passive during the task but actively engaging with participant and experimenter outside the task, and (iv) engaging in the task with the participant and interacting with them outside of it.
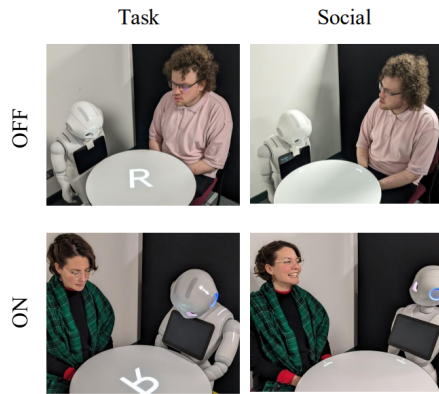
Task   Social

Figure 1: Examples of robot behaviours across the task and social reactivity dimensions.

Our predictions for the upcoming user-study are as follows: (i) The robot's mere-appearance will be sufficient to elicit overall VPT; (ii) if VPT reflects the ad-hoc generation of a model of the robot's mind for predicting its functional behaviour within the task, VPT should increase when participants are paired with the task-reactive robot; (iii) if VPT requires the attribution of agentive mental states to the robot, such that it is believed that the robot is capable of representing and interacting with a more unpredictable social environment, VPT should increase for the socially reactive robot modes of behaviour; (iv) if VPT requires evidence of both functional and agentive mental states, then both measures should interact and reveal VPT only when the robot is both socially and task-active.

2.2 Robot and Behaviour

From the robots available to us, we chose to use the Pepper robot [24] due to its humanoid appearance, and social interaction capabilities. Pepper's behaviour is governed via a simple Finite State Machine (FSM) programmed in Python 2.7 [34]. Transitions in Peppers behavioural state (described below) are instigated by the Inquisit experimental script.

VPT towards humans has been shown to be driven by joint task behaviour [10] with agentive partners [30]. We therefore vary, across participants, Pepper's behaviour along two independent dimensions that allow us to independently evaluate whether perspective taking is driven by (i) Pepper's engagement in the same task as the participant, (ii) by Peppers social behaviour outside of the task, and (iii) by an interaction of both factors.

2.2.1 Task Reactivity. The Task Reactivity factor codes Pepper's behaviour during the task. In its task-on state, Pepper exhibits goal-oriented behaviour by appearing to complete the task alongside the participant. Its posture is focused and its gaze is directed towards the task stimuli. It squeezes each hand as if to respond to the stimuli, with a variable delay imitating human-like performance. Like the participant's, its hands are concealed under the table, such that participants are unable to see its responses (though its mechanical hand actuation is audible). In its task-off state, Pepper is completely stationary during the task, its head is lowered towards its chest and it does not respond to the task stimuli.

2.2.2 Social Reactivity. The Social Reactivity factor codes whether Pepper interacts with the participant outside the task. In its social-on state, Pepper responds to environmental stimuli, turning its head to look at sources of sound (the participant speaking or the experimenter). It establishes eye contact with the participant and experimenter. The robot's posture is upright, and more relaxed than in the task state. There are slight periodic movements of Pepper's upper body. In its social-off state Pepper is completely stationary, its head is lowered towards its chest and it does not respond to any behaviour of experimenter or participant.

2.3 Apparatus, experimental conditions and procedure

2.3.1 Apparatus. The study takes place in the behavioural testing environment of the Action Prediction lab of the University of Aberdeen, within a square area (side length $\approx 2m$) fenced off from the larger lab by four room dividers. In the centre is a circular table of $60cm$ diameter in the centre of the testing space. The experiment is administered using Inquisit (Version 6) [21]. Stimuli are presented from an ceiling mounted projector (Sony, XGA VPL-EX7) onto the table. Participant responses are made on a portable miniature keyboard (cimetech Wireless Numeric Keypad) with UP and DOWN (coloured blue and red respectively) as response keys.

2.3.2 Procedure. Before the experiment begins, participants receive experimental instruction and complete 8 practice trials, in which rotated alphanumeric characters appear on a desktop monitor and participants respond with the UP (blue) key for normally presented characters, and the DOWN (red) key for mirror-inverted characters. The practice trials are near-identical to that of the main experiment, with the exception that feedback (CORRECT / INCORRECT) is given. Once the practice trials are completed, participants are taken to the testing environment, where they first encounter the robot. Participants are sat around the table, with their robot partner 90° to the left or right of them. For participants paired with a social-on robot, Pepper reacts to the participants presence, meet their gaze, and directs attention to any noises the participant or experimenter will make. The social-off robot will not respond to the participant at this time.

Participants then complete 8 blocks of 32 trials. At the start of each trial, a fixation cross is presented in the centre of the table for 1000 ms. Afterwards, after a pseudo-random delay of 1100 ms to 1800 ms that prevents participants from anticipating exactly when the letter will appear [37] an alphanumeric character is presented on the table. The character is presented in its normal ("R", "4", "G", "F") or mirror inverted ("Я", "Ϥ", "Ɔ", "Ꮧ") form, in one of eight possible orientations relative to the participant (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°). Characters are presented for 3000 ms. During this time, participants judge whether the presented character is normal or mirror inverted with a handheld miniature keyboard, on labelled keys. Recognition times are measured from character onset.

For participants paired with a task-on robot, Pepper will complete the task alongside them. The task-off robot will not respond to the stimuli and remains passive.

After each block, participants and Pepper are moved around the table such that both sit at each the four locations of the circular table over the course of the experiment, and Pepper has been on the participant's left and right an equal number of times. This ensures that all

potential sources of visual or spatial differences in the testing area cancel each other out in subsequent analysis. Between experimental blocks the participant is able to rest. When Pepper is in its social-on state it will engage with participant as described above.

2.4 Experimental Measures

Analysis follows our prior work on this task [37–39]. Dependent measures are derived from average recognition times (RTs) measured from character onset, for each of the eight character orientations (RTs, coded relative to upright to the participant), depending on the agent's location relative to the participant (left, right). Participants' RTs for each orientation and condition are parameterised onto two orthogonal and statistically independent summary measures for each participant, such that comparison across conditions is possible without alpha inflation due to multiple comparisons. These summary measures result from representing each participant's RTs for each character orientation and partner location (left/right of the participant) as a vector in a polar coordinate system where RT serves as the magnitude of the vector, and character orientation the polar angle from the origin (upright to the participant).

$$P_{self} = -RTcos(\theta_{self}) \tag{1}$$
$$P_{other} = -RTcos(\theta_{other}) \tag{2}$$

The Own-Perspective summary measure (Equation 1) quantifies how much each participant is anchored in their own perspective when judging letter forms. It quantifies the extent to which response times across character orientations follow the classic mental rotation pattern [29], becoming faster the more characters are orientated towards the participant (0°, 45°, 315°) and slower when oriented away (180°, 135°, 225°). It is calculated as the average of the recognition times for each character orientation, with the contribution of each weighted by the negative cosine of the angular disparity between letter orientation and participant. Positive values represent faster recognition times the more letters are oriented towards the participant.

The Other-Perspective summary measure (Equation 2) is computed analogously, but quantifies how much the participant takes the other agent's perspective. It quantifies the extent to which a participant's recognition times are faster the more characters are oriented towards the robot (e.g., for a robot on the left, 270°, 225°, 315°) than when oriented away from them (for an agent on the left, 45°, 90°, 135°). It is calculated as the average of the recognition times for each character orientation, with the contribution of each weighted by the negative cosine of the angular disparity between letter orientation and the other agent. Positive values show that a participant's recognition times are faster the more the characters are oriented towards the robot and slower the more they are oriented away from them, that is, the extent to which they are spontaneously judging the letter forms from the robot's perspective.

For each participant and experimental condition the mean is calculated independently on both summary measures. Note that because the other agent is always sitting at a 90° angle to the participant, the Other-Perspective and Self-Perspective measures are orthogonal to each other and statistically independent.
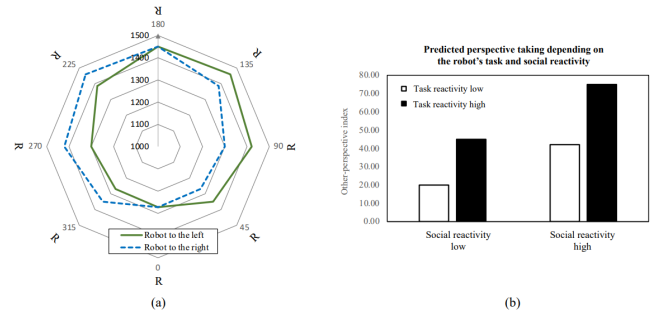

Figure 2: Expected pattern of results: (a) Polar plot showing RT increase the more characters are rotated away from the participant, as well as RT decrease and increase when characters are rotated towards or away from the robot; (b) Bar plot showing example data for Other-perspective summary measure when VPT is driven by independent contributions of task and social behaviours, so that VPT is strongest when both are combined.

3 CONCLUSION

Human-robot collaboration is forecast to play a crucial role in numerous industries [20], notably manufacturing [18] and construction [6]. Visual perspective taking is an essential mechanism for human collaboration, and is therefore of considerable importance for human-robot collaboration. Although progress has been made allowing a robot to take their human interaction partner's perspective, the design-space of features that support human-to-robot perspective taking is severely under-researched. Prior work has found predominately that a robot's human-like appearance promotes perspective taking, though we argue that the typical non-situated and disembodied mode of presentation on computer-based experimental designs can test little else.

Here, we present a novel task that resolves previous methodological problems and can reliably reveal whether human partner's take a robot's perspective in a face-to-face interaction, and which robot behaviour features (i.e., social and task reactivity) promote such perspective taking. Importantly, this task provides a direct measure of whether participants assimilate the robot's spatial reference with their own egocentric perceptual processing, so that it is available for action planning and prediction in shared task spaces.

Our contribution is an experimental design and accompanying analysis that can robustly explore the VPT feature-space with real-world interactions between individuals and a physically embodied robot. This grants investigation into attributes deeper than mere-appearance, such as dynamic goal-driven and autonomously driven socially reactive behaviours. We are optimistic that our contribution will lay the groundwork for investigating behavioural that improve human-robot collaboration, and pave the way towards the design of more understandable robots. We look forward to updating the community on the findings of our participant study.

ACKNOWLEDGMENTS

REFERENCES

[1] João Tiago Almeida, Iolanda Leite, and Elmira Yadollahi. 2023. Would You Help Me? Linking Robot's Perspective-Taking to Human Prosocial Behavior. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, New York, NY, United States, 388–397.

[2] Patric Bach, Wendy Fenton-Adams, and Steven P Tipper. 2014. Can't touch this: The first-person perspective provides privileged access to predictions of sensory action outcomes. *Journal of Experimental Psychology: Human Perception and Performance* 40, 2 (2014), 457.

[3] C Daniel Batson, Shannon Early, and Giovanni Salvarani. 1997. Perspective taking: Imagining how another feels versus imaging how you would feel. *Personality and social psychology bulletin* 23, 7 (1997), 751–758.

[4] Cynthia Brosque, Elena Galbally, Oussama Khatib, and Martin Fischer. 2020. Human-Robot Collaboration in Construction: Opportunities and Challenges. In *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. IEEE, New York, NY, United States, 1–8. https://doi.org/10.1109/HORA49412.2020.9152888

[5] Henryk Bukowski, Jari K Hietanen, and Dana Samson. 2015. From gaze cueing to perspective taking: Revisiting the claim that we automatically compute where or what other people are looking at. *Visual cognition* 23, 8 (2015), 1020–1042.

[6] Alan G Burden, Glenda Amayo Caldwell, and Matthias R Guertler. 2022. Towards human–robot collaboration in construction: current cobot trends and forecasts. *Construction Robotics* 6, 3-4 (2022), 209–220.

[7] Boyuan Chen, Yuhang Hu, Robert Kwiatkowski, Shuran Song, and Hod Lipson. 2021. Visual Perspective Taking for Opponent Behavior Modeling. *2021 IEEE International Conference on Robotics and Automation (ICRA 2021)* (5 2021), 13678–13685. http://arxiv.org/abs/2105.05145

[8] Sarah H Creem-Regehr, Kyle T Gagnon, Michael N Geuss, and Jeanine K Stefanucci. 2013. Relating spatial perspective taking to the perception of other's affordances: Providing a foundation for predicting the future behavior of others. *Frontiers in human neuroscience* 7 (2013), 596.

[9] Tobias Fischer and Yiannis Demiris. 2020. Computational Modeling of Embodied Visual Perspective Taking. *IEEE Transactions on Cognitive and Developmental Systems* 12, 4 (2020), 723–732. https://doi.org/10.1109/TCDS.2019.2949861

[10] Martin Freundlieb, Ágnes M Kovács, and Natalie Sebanz. 2016. When do humans spontaneously adopt another's visuospatial perspective? *Journal of experimental psychology: human perception and performance* 42, 3 (2016), 401.

[11] Tiziano Furlanetto, Cristina Becchio, Dana Samson, and Ian Apperly. 2016. Altercentric interference in level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Journal of Experimental Psychology: Human Perception and Performance* 42, 2 (2016), 158.

[12] Esther Herrmann, Josep Call, María Victoria Hernàndez-Lloreda, Brian Hare, and Michael Tomasello. 2007. Humans Have Evolved Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis. *Science* 317, 5843 (2007), 1360–1366. https://doi.org/10.1126/science.1146282 arXiv:https://www.science.org/doi/pdf/10.1126/science.1146282

[13] Cecilia M. Heyes and Chris D. Frith. 2014. The cultural evolution of mind reading. *Science* 344, 6190 (2014), 1243091. https://doi.org/10.1126/science.1243091 arXiv:https://www.science.org/doi/pdf/10.1126/science.1243091

[14] Anil K. Inkulu, Raju B. MVA, Ashok Dara, and K. SankaranarayanaSamy. 2022. Challenges and opportunities in human robot collaboration context of Industry 4.0 - a state of the art review. *The Industrial Robot* 49, 2 (2022), 226–239. https://www.proquest.com/scholarly-journals/challenges-opportunities-human-robot/docview/2626883272/se-2 Copyright - © Emerald Publishing Limited; Last updated - 2023-09-11.

[15] Maria Kozhevnikov, Michael A. Motes, Bjoern Rasch, and Olessia Blajenkova. 2006. Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Applied Cognitive Psychology* 20, 3 (2006), 397–417. https://doi.org/10.1002/acp.1192 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/acp.1192

[16] Séverin Lemaignan, Mathieu Warnier, E Akin Sisbot, Aurélie Clodic, and Rachid Alami. 2017. Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence* 247 (2017), 45–69.

[17] Mingjiang Liu, Chengli Xiao, and Chunlin Chen. 2022. Perspective-Corrected Spatial Referring Expression Generation for Human–Robot Interaction. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 52, 12 (2022), 7654–7666. https://doi.org/10.1109/TSMC.2022.3161588

[18] Eloise Matheson, Riccardo Minto, Emanuele GG Zampieri, Maurizio Faccio, and Giulio Rosati. 2019. Human–robot collaboration in manufacturing applications: A review. *Robotics* 8, 4 (2019), 100.

[19] Bradley D Mattan, Pia Rotshtein, and Kimberly A Quinn. 2016. Empathy and visual perspective-taking performance. *Cognitive neuroscience* 7, 1-4 (2016), 170–181.

[20] George Michalos, Panagiotis Karagiannis, Nikos Dimitropoulos, Dionisis Andronas, and Sotiris Makris. 2022. Human robot collaboration in industrial environments. *The 21st century industrial robot: when tools become collaborators* (2022), 17–39.

[21] millisecond. 2022. Inquisit6. https://www.millisecond.com/

[22] Manisha Natarajan, Esmaeil Seraj, Batuhan Altundas, Rohan Paleja, Sean Ye, Letian Chen, Reed Jensen, Kimberlee Chestnut Chang, and Matthew Gombolay. 2023. Human-robot teaming: grand challenges. *Current Robotics Reports* 4, 3 (2023), 81–100.

[23] Kohei Ogawa, Shuichi Nishio, Kensuke Koda, Koichi Taura, Takashi Minato, Carlos Toshinori Ishii, and Hiroshi Ishiguro. 2011. Telenoid: Tele-presence android for communication. In *ACM SIGGRAPH 2011 emerging technologies*. Association for Computing Machinery, New York, NY, United States, 1–1.

[24] Amit Kumar Pandey and Rodolphe Gelin. 2018. A Mass-Produced Sociable Humanoid Robot: Pepper: The First Machine of Its Kind. *IEEE Robotics & Automation Magazine* 25, 3 (2018), 40–48. https://doi.org/10.1109/MRA.2018.2833157

[25] Lorenzo Parenti, Uma Prashant Navare, Serena Marchesi, Cecilia Roselli, and Agnieszka Wykowska. 2023. Theta Synchronization as a Neural Marker of Flexible (re-) use of Socio-cognitive Mechanisms for a New Category of (artificial) Interaction Partners. *Cortex* 169 (2023), 249–258.

[26] Eileen Roesler, Dietrich Manzey, and Linda Onnasch. 2023. Embodiment matters in social hri research: Effectiveness of anthropomorphism on subjective and objective outcomes. *ACM Transactions on Human-Robot Interaction* 12, 1 (2023), 1–9.

[27] Dana Samson, Ian A Apperly, Jason J Braithwaite, Benjamin J Andrews, and Sarah E Bodley Scott. 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 (2010), 1255.

[28] Steven Samuel, Klara Hagspiel, Madeline J. Eacott, and Geoff G. Cole. 2021. Visual perspective-taking and image-like representations: We don't see it. *Cognition* 210 (5 2021). https://doi.org/10.1016/j.cognition.2021.104607

[29] Roger N Shepard and Jacqueline Metzler. 1971. Mental rotation of three-dimensional objects. *Science* 171, 3972 (1971), 701–703.

[30] Andrew Surtees, Ian Apperly, and Dana Samson. 2016. I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition* 150 (2016), 43–52.

[31] Andrew DR Surtees, Stephen A Butterfill, and Ian A Apperly. 2012. Direct and indirect measures of level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology* 30, 1 (2012), 75–86.

[32] Sam Thellman, Maartje de Graaf, and Tom Ziemke. 2022. Mental state attribution to robots: A systematic review of conceptions, methods, and findings. *ACM Transactions on Human-Robot Interaction (THRI)* 11, 4 (2022), 1–51.

[33] J Gregory Trafton, Nicholas L Cassimatis, Magdalena D Bugajska, Derek P Brock, Farilee E Mintz, and Alan C Schultz. 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 (2005), 460–470.

[34] Guido Van Rossum and Fred L Drake Jr. 1995. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

[35] Basil Wahn and Leda Berio. 2023. The influence of robot appearance on visual perspective taking: Testing the boundaries of the mere-appearance hypothesis. *Consciousness and Cognition* 116 (2023), 103588.

[36] Basil Wahn, Leda Berio, Matthias Weiß, and Albert Newen. 2023. Try to see it my way: Humans take the level-1 visual perspective of humanoid robot avatars. *International Journal of Social Robotics* (2023), 1–12.

[37] Eleanor Ward, Giorgio Ganis, and Patric Bach. 2019. Spontaneous Vicarious Perception of the Content of Another's Visual Perspective. *Current Biology* 29 (3 2019), 874–880.e4. Issue 5. https://doi.org/10.1016/j.cub.2019.01.046

[38] Eleanor Ward, Giorgio Ganis, Katrina L. McDonough, and Patric Bach. 2020. Perspective taking as virtual navigation? Perceptual simulation of what others see reflects their location in space but not their gaze. *Cognition* 199 (6 2020). https://doi.org/10.1016/j.cognition.2020.104241

[39] Eleanor Ward, Giorgio Ganis, Katrina L. McDonough, and Patric Bach. 2022. Is implicit Level-2 visual perspective-taking embodied? Spontaneous perceptual simulation of others' perspectives is not impaired by motor restriction. *Quarterly Journal of Experimental Psychology* 75 (7 2022), 1244–1258. Issue 7. https://doi.org/10.1177/17470218221077102

[40] Agnieszka Wykowska. 2020. Social robots to test flexibility of human social cognition. *International Journal of Social Robotics* 12, 6 (2020), 1203–1211.

[41] Tian Ye, Fumikazu Furumi, Daniel Catarino da Silva, and Antonia Hamilton. 2021. Taking the perspectives of many people: Humanization matters. *Psychonomic Bulletin & Review* 28 (2021), 888–897.

[42] Tian Ye, Takashi Minato, Kurima Sakai, Hidenobu Sumioka, Antonia Hamilton, and Hiroshi Ishiguro. 2023. Human-like interactions prompt people to take a robot's perspective. *Frontiers in Psychology* 14 (2023).

[43] Jing Zhai, Jiushu Xie, Jiahan Chen, Yujie Huang, Yuchao Ma, and Yanli Huang. 2021. The presence of other-race people disrupts spontaneous level-2 visual perspective taking. *Scandinavian Journal of Psychology* 62, 5 (2021), 655–664.

[44] Xuan Zhao, Corey Cusimano, and Bertram F. Malle. 2015. Do People Spontaneously Take a Robot's Visual Perspective?. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts* (Portland, Oregon, USA) *(HRI'15 Extended Abstracts)*. Association for Computing Machinery, New York, NY, USA, 133–134. https://doi.org/10.1145/2701973.2702044

[45] Xuan Zhao and Bertram F Malle. 2022. Spontaneous perspective taking toward robots: The unique impact of humanlike appearance. *Cognition* 224 (2022), 105076.