Collaboratively framed interactions increase the adoption of intentional stance towards robots *

A. Abubshait, J. Perez-Osorio, D. De Tommaso & A. Wykowska, *Member, IEEE*

Abstract— When humans interact with artificial agents, they adopt various stances towards them. On one side of the spectrum, people might adopt a mechanistic stance towards an agent and explain its behavior using its functional properties. On the other hand, people can adopt the intentional stance towards artificial agents and explain their behavior using mentalistic terms and explain the agents' behavior using internal states (e.g., thoughts and feelings). While studies continue to investigate under which conditions people adopt the intentional stance towards artificial robots, here, we report a study in which we investigated the effect of social framing during a color-classification task with a humanoid robot, iCub. One group of participants were asked to complete the task with iCub, in collaboration, while the other group completed an identical task with iCub and were told that they were completing the task for themselves. Participants completed a task assessing their level of adoption of the Intentional Stance (the InStance test) prior to - and after completing the task. Results illustrate that participants who “collaborated” with iCub were more likely to adopt the intentional stance towards it after the interaction. These results suggest that social framing can be a powerful method to influence the stance that people adopt towards a robot.

I. INTRODUCTION

When we socially interact with others, we recruit specific brain mechanisms that facilitate social interactions [1]. These mechanisms ensure social interactions' success by allowing us to predict others' thoughts, feelings, and intentions (i.e., Theory of mind; [2]). Although humans generally do not recruit similar mechanisms when interacting with artificial agents [1], studies investigating social interactions with artificial agents suggest that similar social cognitive mechanisms can be activated in specific contexts (e.g., [3], [4]). One context that can predict if we recruit these brain mechanisms is whether we adopt the intentional stance towards an artificial agent. Adopting the intentional stance occurs when humans behave as if other agents (or any complex system) would have thoughts, desires, and feelings [5]. As such, when we adopt the intentional stance towards an artificial agent we explain and predict its behavior using mentalistic descriptions, rather than using mechanistic descriptions (e.g., “iCub *wants* to play” or “iCub is *feeling* bored” as opposed to, “iCub has been *programmed* to play” or “iCub's *gears* are stuck”). Adopting the intentional or mechanistic stance provides an adaptive heuristic for understanding the world and its agents [6]. While studies have demonstrated that humans can adopt the intentional stance

towards artificial agents [7]–[9], it still unclear what contributes to adopting the intentional stance.

Studies that investigated the factors that influence how and when people explain an artificial agents' behavior as mentalistic suggest that these elements can be categorized as either factors related to the human (i.e., individual differences) or those that are related to the artificial agent (i.e., features of the artificial agent; [10]). Studies that investigated individual differences, for example, have shown that individual differences in resting-state beta rhythms in the brain [11], experiencing loneliness [12], and expectations regarding the capabilities of robots can predict whether people would explain a robot's behavior in mentalistic or mechanistic terms [13]. Studies that investigated robot factors have shown that robots that are embodied [14], look humanlike [15], behave unexpectedly ([10]), display variability in their behavior [16], or cheat [17], [18] are more likely to have their behaviors explained in mentalistic terms as opposed to mechanistic terms (For a review, see [19]). More recently, Lefkeli and colleagues [20], asked participants to play in cooperation with- or competition against a robot. They found that people tended to explain the robots' behavior in mentalistic terms (i.e., more mind perception) when they played against it and won as compared to losing. They also showed the opposite effect as people tended to explain the robots' behavior in mentalistic terms when they cooperated and lost compared to when they won [20]. While these studies suggest that people can adopt the intentional stance towards robots in specific contexts, the latter study suggests that simply the context in which an interaction occurs (i.e., telling participants that a robot is a teammate or an opponent) can influence the stance that we adopt towards artificial agents. Thus, the study presented here examined the effect of social interaction framing on the adoption of intentional stance.

According to framing theory [21], the way information is presented influences cognitive processes, decision-making and behavior in general. For example, would you choose an “80% sugar-free” yogurt or a “20% sugar” yogurt? Although both options are identical, the way the information is illustrated has the ability to bias our judgement towards the first alternative. This framing bias extends to multiple domains, including human-robot interaction. Some studies evaluated the impact of anthropomorphic framing of social robots during social interaction compared to a more mechanistic or functional framing. Anthropomorphic framing

* This work received support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant awarded to AW, titled “InStance: Intentional Stance for Social Attunement.” G.A. No: ERC-2016- StG-715058).

All authors are affiliated with the Istituto Italiano di Tecnologia, Genova, Italy, 16145. AW will serve as the corresponding author (A.W., phone: +39 020 8172 242, email: agnieszka.wykowska@iit.it)

can be induced by giving names [22], background stories [23] or both [23]. All three studies showed that anthropomorphic framing elicits empathetic responses and facilitates social interactions with robots. However, these studies were conducted online or using short stories that described a robot, thus the effects of framing in real interactions were unknown. To address this issue, Onnasch & Roesler [24] compared the effects of anthropomorphic versus functional framing during actual interactions with a robot (NAO). While they found no effect of anthropomorphic framing on the humanlike perception of the robot, they found that functional framing influenced participants to be more cooperative towards the robot, relative to the anthropomorphic condition. The fact that framing did not influence people’s tendency to explain a robot’s behavior as mentalistic/mechanistic, but influenced their interaction suggests that a more complex interaction between framing and real interactions with robots.

One explanation to why the framing effect did not influence perceptions of robots in past research could be due to the fact that there was no goal involved in the interaction. This is of high importance as literature on the framing effect identified goals as an essential determinant of human behavior as goals guide decision-making processes, modulate allocation of cognitive resources and influence how people perceive different social situations ([25], [26]; for a review see [27]).

Given this, the study presented here aimed to investigate the effect of framing on subjective perceptions of a robot during a goal-oriented task. Specifically, we asked if interacting with a robot in a task that is framed in a collaborative vs. individualistic manner during a color-categorization task influences the likelihood of adopting the intentional stance towards the robot. To investigate this question, all participants completed an identical color-categorization task with iCub [28] while we only manipulated the social framing of the task. One group of participants were instructed to complete the task in collaboration with iCub, while the other group completed the task with iCub, for themselves. On half of the trials, iCub would grasp a colored cube (i.e., blue or yellow) and handed it over to the participant. Participants would then take the cube and place it in one of two bins based on its color. During the other half of the task, participants saw iCub grasp the cube and then categorizing it itself. Since collaborative framing could highlight that iCub and the participant would experience the outcomes together, we hypothesized that participants would be more likely to adopt the intentional stance towards the robot, relative to the condition when the interaction is framed as individualistic. This hypothesis is driven by the fact that perception of robots can influence how people choose to collaborate with them [29].

II. METHODS

A. Participants

Forty-eight participants took part in the study, three of them were excluded due to technical problems with the robot. The remaining 45 participants (mean age = 24.64, range = 19-34, 13 males, all right-handed) were assigned to the Collaborative condition (n = 23) or the Individual condition (n = 22). All

participants reported no history of psychiatric or neurological conditions and received monetary compensation (10€ per

Figure 1. Task Space. (Participants were seated across from iCub with a touch screen in front of them that served as a table and a stimulus presentation platform).

hour) upon completion. Participants provided written consent prior to completing the study and were debriefed at the end of the experiment. The study was approved by the local Ethical Committee (Comitato Etico Regione Liguria), and strictly followed the Code of Ethics of the World Medical Association (Declaration of Helsinki).

B. Procedure

After providing consent, participants completed the first part of the InStance Test (IST PRE; [16]). Next, they were moved to a different room where the interaction with iCub would take place. Participants sat across from iCub and were provided instructions for the task. One group of participants was told that they were completing a color classification task with iCub where the points would go towards them as a team (i.e., Collaborative condition). The other group was instructed to complete the task with iCub where only they would receive the points (i.e., Individual condition). The task required participants to classify cubes (i.e., yellow or blue), by placing them in one of two bins. Next to each bin was a bar that had a ratio of yellow-to-blue colors. Participants were asked to drop the cube in the bin that was adjacent to the bar that had the dominant color (e.g., if the cube was yellow and the bar on the left had a larger yellow-to-blue ratio, then the cube should go to the left bin; see **Fig 1.** for the task set up). After completing the classification task with the robot, which took ~45 minutes, participants completed the second part of the InStance Test (IST POST; [16]) and the RoSAS questionnaire [30]. At the end of the experiment, participants were thanked and debriefed.

C. Task

Each trial of the classification task started with participants pressing a response button that was placed in front of them. Next, two differently colored bars would appear next to the bins. After the bars appeared on the screen, on half of the trials, iCub grasped the object, looked at it, handed it to the participant (i.e., 50% of the time). On the second half of the trials, iCub placed the cube in one of the two bins itself, where it dropped the cube in the correct bin (50% of the time) or

dropped the cube in the incorrect bin (i.e., the other 50% of the time). If iCub handed the cube to the participant, they released the button, grasped the cube, and dropped the cube into the corresponding bin (e.g., if the cube was yellow and the left bin was 80% yellow, then the correct bin is the left bin). After either the participant or iCub dropped the cube, correct or incorrect feedback was shown on the screen.

On trials that iCub completed the task, participants observed iCub and were asked to release the response button before iCub dropped the cube to ensure that participants were paying attention during the robot’s trials. After each block, participants saw their cumulative score, which was based on +2 points for each correct trial and -2 points for each incorrect trial. The classification task consisted of 8 trials per condition, where iCub placed the cube in the bin for 4 of those trials, with 2 correct trials (i.e., congruent) and 2 incorrect trials (i.e., incongruent), and handed the cube to the participants for the remaining 4 trials. The experiment consisted of 15 blocks (i.e., 120 trials overall).

D. Apparatus and stimuli

The experiment was presented on a 42” screen placed between the robot and the participant, which was used as a table and a stimulus presentation platform (see **Fig. 1**). Participant responses were collected using two physical switches placed under the boxes for detecting the drop time of the cube. Specifically, we used the buttons provided in the Logitech Gaming Kit [31]. The iCub humanoid robot was connected to a single workstation (SSD 250 Gb, 32 Gb RAM, Intel i7 7700CPU@3.60 GHz 4 cores, GPU GeForce GTX 1060 3Gb, running Ubuntu 20.04 LTS) through a peer-to-peer Ethernet connection. We controlled the behavior of the robot and the stimuli with Psychopy 2020.1.2 [32]. iCub was controlled using YARP [33] and Python wrappers. The grasping of objects was achieved by controlling the robot’s arms with the YARP IPositionControl, a position controller that generates joint trajectories following a minimum jerk profile [33].

E. Questionnaires

Two questionnaires were used to examine participants’ subjective perceptions of iCub responses. The first was the InStance, which measures participants’ propensity to adopt the intentional stance towards iCub [9]. The second questionnaire was the Robotic Social Attribute Scale (RoSAS), which measures participants’ perceptions of a robot’s social attributes on three dimensions; warmth, competence, and discomfort [30].

To measure whether participants changed their propensity to adopt the intentional stance after the interaction with the robot, we divided the InStance Test into two halves (i.e., Pre and Post). The division was based on prior work that examined how the items of the scale loaded using PCA and created two validated parts that measure how interaction with iCub can bias the intentional/mechanistic stance [16], [34]. IST-Pre consisted of 17 items and was administered prior to the interaction with iCub, while IST-Post consisted of 17 items and was administered after the interaction. The IST

consists of “scenarios” involving three pictures of iCub conducting an action (e.g., iCub playing a card game with a human) with two descriptions beneath the scenario. One description is written with mentalistic words, and the other with mechanistic terms. Participants are asked to use a slider and decide whether the mentalistic or the mechanistic description fits the scenario better. See Marchesi et al., 2019 for details about the presentation of the IST questionnaire. In the present experiment, scoring of the IST test was done by normalizing the scale so that 0 is always mechanistic and 100 is always intentional and calculating the average for each participant. The scale ranged from 0 (i.e., adopting the mechanistic stance entirely) to 100 (adopting the intentional stance entirely). The RoSAS scale was scored and the average was calculated for each dimension individually (i.e., warmth, competence, and discomfort). The scale ranged from 0 (i.e., low on the respective dimension) to 7 (i.e., high on the respective dimension).

F. Analysis

To examine whether the collaborative framing influenced participants’ propensity to adopt the intentional stance towards iCub, we ran a one-way ANCOVA examining differences in IST-POST ratings between *Framing* as a 2 level factor (Collaborative vs. Individual). The ANCOVA also had IST-PRE ratings as a covariate, to account for differences in IST ratings at baseline. To examine participants’ subjective ratings of the RoSAS scores, we ran a repeated measures mixed 2 x 3 ANOVA with *Framing* as a 2-level between factor (Individual vs. Collaborative) and *Dimension* as a 3-level within factor (Discomfort vs. Warmth vs. Competence). We additionally examined the relationships between POST-IST scores with each dimension of the RoSAS using three separate correlations (i.e., a correlation for dimension). Since this was an exploratory test, we averaged the ratings over groups in the different Framing conditions. We used this exploratory test to examine if adopting the intentional stance was related to robot perceptions. Multiple comparisons in the correlations were corrected using the false discovery rate procedure (FDR).

G. Results

Prior to running the ANOVA, we wanted to examine if the two groups scored differently at IST-PRE using a t-test. The t-test showed no significant differences between the collaborative and the individual groups ($t(42) = 1.71, p = .09$; $M_{\text{Collaborative}} = 39.43$ vs. $M_{\text{Individual}} = 29.29$). Results of the ANCOVA examining differences in IST-POST ratings showed a significant effect of *Framing* ($F(1, 41) = 13.58, p < .001$) after controlling for participants’ IST-PRE ratings ($F(1, 41) = 60.37, p < .001$). The difference showed that participants in the Collaborative condition had higher IST-POST ratings compared to those in the Individual condition ($M_{\text{Collaborative}} = 42.1$ vs. $M_{\text{Individual}} = 26.4$, Cohen’s $D = .71$).

The ANOVA examining differences in RoSAS scores showed no effect of *Framing* ($F(1, 39) = < .001, p = .99$, $ges < .001$), but a significant effect of *Dimension* ($F(1,73) = 123.8, p < .001$, $ges = .59$) with mean differences between the Competency ($M = 5.95$), Discomfort ($M = 1.66$) and Warmth

($M = 2.75$) dimensions. However, there was no *Framing X Dimension* interaction ($F(1,73) = 2.55, p = .08, \eta^2 = .08$).

Degrees of freedom were corrected using Greenhouse-Geisser correction.

Figure 2. Instance Test-POST rating. (The graph illustrates mean differences between those in the Collaborative condition vs. the Individual condition. The results showed that participants in the Collaborative condition were more likely to adopt the intentional stance towards iCub compared to those who completed the task alone. The error bars illustrate the standard error of the mean. The asterisks indicate a significant difference $p < .001$).

The correlation between Post-IST scores and the Competence Dimension of the RoSAS scale was significantly positive ($r(40) = .33, p = .04$) with higher Post-IST scores correlating with higher competence ratings. The Post-IST scores were also correlated with the Warmth dimension of the RoSAS scale ($r(40) = .43, p < .01$). The Post-IST scores, however, were not correlated with the Discomfort dimension ($r(40) = .06, p = .66$).

Figure 3. Correlation between IST-POST and RoSAS ratings. (The IST-POST and the RoSAS ratings were averaged over all participants in all conditions. Only the correlations between IST-POST and Competence, and IST-POST and Warmth were significant).

III. DISCUSSION

The goal of the current study was to examine if framing a task with a humanoid robot influences adoption of the intentional stance. Specifically, we asked whether a collaborative setting (where outcomes of a task affect both the human and the humanoid robot vs. a setting where only the participant was the beneficiary of the outcome) influences the adoption of intentional stance. To do so, we asked participants to complete a color categorization task with iCub, a humanoid robot. One group of participants was told that they were completing the task together with iCub and that both the participant and iCub were the recipients of the points that they received on correct trials. The second group of participants completed an identical task; however, they were told that only they (i.e., the participant) were the recipients of the points. We hypothesized that participants in the collaborative condition would be more likely to adopt the intentional stance towards the robot after interacting since the collaborative framing would highlight that the outcomes would be experienced by iCub and the participant. On the other hand, those who were in the individual condition would mostly attend to their own performance and disregard the robots' behavior, which would lead to adopting the intentional stance to a lesser extent (i.e., or adopting the design/mechanistic stance).

Results of the experiment show that, as hypothesized, participants in the collaborative framing condition were more likely to adopt the intentional stance towards iCub compared to those in the individual-framing condition, after the interaction. This effect held even after controlling for participants' baseline intentionality scores. We additionally examined whether framing influenced participants' scores on the different RoSAS dimensions (i.e., competence, warmth and discomfort). Although mean differences were evident between the dimensions, the effect of framing was not present on these dimensions. Exploratory analysis showed positive correlations between the IST test and the Warmth and Competence dimensions of the RoSAS scale, but not discomfort. This could suggest that adopting the intentional stance is related with person perception as these dimensions reflect positive social interactions with social beings [30]. However, since this was an exploratory finding, future work needs to investigate this further.

One explanation to why participants were more likely to adopt the intentional stance towards iCub when they were asked to complete the task in collaboration with iCub could be that such framing can influence how people perceive outcomes during decision making [35]. In other words, participants in the collaborative condition were more cognizant of the fact that their performance influenced the robots' points and that the robots' performance influenced their points, which could increase the likelihood of adopting the intentional stance. On the other hand, participants in the individual condition were only attentive to their own performance and not the iCub's as their performance and the robot's performance were mutually exclusive. The interpretation that framing can influence awareness to different outcomes is in line with prior work that

found that framing an interaction with a robot could influence the awareness of its properties [24], and that robot perceptions influence how people choose to interact with robots [1]. A second explanation could be that the collaborative context could have biased participants to perceive iCub as an ingroup member and as such would increase how anthropomorphic and intentional participants perceived it. This would be in line with the fact that ingroup robot members are perceived as being more anthropomorphic [36]. Yet, it is unclear to what extent does this change in adopting the intentional stance could influence subsequent interactions with robots. While we cannot make claims about how adopting the intentional stance influences behavior given the current data, prior work has shown that subjective perceptions towards robots are correlated with neural and behavioral indices of social behavior [3].

The use of framing in HRI has been investigated in the past, however, results were mixed in terms of finding an effect of framing on adopting the intentional stance (i.e., some studies showed the effect while others did not; e.g., [24], [37]). While we cannot draw direct comparisons between the studies, one explanation to why the current study did find this effect could be due to the fact that our manipulation included a goal-oriented task that was scored. In other words, because the scores were based on the framing manipulation, participants' motivation and/or attention towards a task increased, which could have exerted a stronger effect. This is consistent with prior work that suggests that goal-oriented behavior is essential for psychological research [27]. Our results extend on the prior studies of framing effects by showing that this effect can be observed in a goal-oriented task that can be helpful for researchers to use to examine factors that influence perceptions of robots in a real-interactive scenario.

The fact that social relations influence perceptions of autonomous systems (e.g., robots, algorithms, and artificial agents) is consistent with prior work that has shown that framing autonomous agents as teammates as opposed to tools positively influences subjective affective responses towards them [38]. Interestingly, these studies also provide insight into how the effect of framing can influence team/collaborative performance and not just subjective ratings [38]–[41], which future work should investigate further. Future research should also examine how framing effects in combination with other robot features (e.g., robot appearance and behavior) influences adoption of the intentional stance. Also, it is imperative to examine how well do these findings generalize to interactions with avatars/virtual agents and not just robots.

This study fits in with prior HRI literature that investigates under which conditions people adopt the intentional stance towards robots. Prior work has shown that factors such as a robots' physical appearance [1], behavior [16], exposure to reactive [42], or repetitive behavior [43] influence attribution of intentionality. Here, we illustrate that framing a task in a collaborative manner increases our tendency to adopt the intentional stance towards the robot. This implication can help improve the design of social-robot teams/collaborative settings as previous work has shown that adopting the intentional stance towards robots can influence their subsequent interaction with them [1]. Additionally, the findings of this study suggest that robot designers can make use of these

framing effects to influence adoption of the intentional stance towards robots when it is not feasible to manipulate other robot features (e.g., robot appearance), which can be a cost effective method to influencing how people perceive robots.

ACKNOWLEDGMENT

The authors would like to thank Matilde Piccini and Giulia Siri for their help with data collection.

REFERENCES

- [1] E. Wiese, G. Metta, and A. Wykowska, "Robots As Intentional Agents: Using Neuroscientific Methods to Make Robots Appear More Social," *Front. Psychol.*, vol. 8, p. 1663, 2017, doi: 10.3389/fpsyg.2017.01663.
- [2] D. Premack and G. Woodruff, "Does the chimpanzee have a theory of mind?," *Behav. Brain Sci.*, vol. 1, no. 4, pp. 515–526, 1978, doi: 10.1017/S0140525X00076512.
- [3] E. Wiese, G. A. Buzzell, A. Abubshait, and P. J. Beatty, "Seeing minds in others: Mind perception modulates low-level social-cognitive performance and relates to ventromedial prefrontal structures," *Cogn. Affect. Behav. Neurosci.*, vol. 18, no. 5, pp. 837–856, Oct. 2018, doi: 10.3758/s13415-018-0608-2.
- [4] K. Kompatsiari, F. Ciardo, V. Tikhanoff, G. Metta, and A. Wykowska, "On the role of eye contact in gaze cueing," *Sci. Rep.*, vol. 8, no. 1, p. 17842, Dec. 2018, doi: 10.1038/s41598-018-36136-2.
- [5] D. C. Dennett, *The Intentional Stance*, no. 1. MIT Press, 1987, pp. 891–892.
- [6] J. Perez-Orsorio and A. Wykowska, "Adopting the intentional stance toward natural and artificial agents," *Philos. Psychol.*, vol. 33, no. 3, pp. 369–395, Apr. 2020, doi: 10.1080/09515089.2019.1688778.
- [7] H. L. Gallagher, A. I. Jack, A. Roepstorff, and C. D. Frith, "Imaging the Intentional Stance in a Competitive Game," *NeuroImage*, vol. 16, no. 3, Part A, pp. 814–821, Jul. 2002, doi: 10.1006/nimg.2002.1117.
- [8] S. Thellman, A. Silvervarg, and T. Ziemke, "Folk-Psychological Interpretation of Human vs. Humanoid Robot Behavior: Exploring the Intentional Stance toward Robots," *Front. Psychol.*, vol. 8, p. 1962, 2017, doi: 10.3389/fpsyg.2017.01962.
- [9] S. Marchesi, D. Ghiglini, F. Ciardo, J. Perez-Orsorio, E. Baykara, and A. Wykowska, "Do We Adopt the Intentional Stance Toward Humanoid Robots?," *Front. Psychol.*, vol. 10, p. 450, 2019, doi: 10.3389/fpsyg.2019.00450.
- [10] A. Waytz, J. Cacioppo, and N. Epley, "Who Sees Human?: The Stability and Importance of Individual Differences in Anthropomorphism," *Perspect. Psychol. Sci.*, vol. 5, no. 3, pp. 219–232, May 2010, doi: 10.1177/1745691610369336.
- [11] F. Bossi, C. Willemse, J. Cavazza, S. Marchesi, V. Murino, and A. Wykowska, "The human brain reveals resting state activity patterns that are predictive of biases in attitudes toward robots," *Sci. Robot.*, vol. 5, no. 46, p. eabb6652, Sep. 2020, doi: 10.1126/scirobotics.abb6652.
- [12] H. I. Shin and J. Kim, "My computer is more thoughtful than you: Loneliness, anthropomorphism and dehumanization," *Curr. Psychol.*, vol. 39, no. 2, pp. 445–453, Apr. 2020, doi: 10.1007/s12144-018-9975-7.
- [13] J. Perez-Orsorio, S. Marchesi, D. Ghiglini, M. Ince, and A. Wykowska, "More Than You Expect: Priors Influence on the Adoption of Intentional Stance Toward Humanoid Robots," in *Social Robotics*, Cham, 2019, pp. 119–129.
- [14] S. Kiesler, A. Powers, S. R. Fussell, and C. Torrey, "Anthropomorphic Interactions with a Robot and Robot-like Agent," *Soc. Cogn.*, vol. 26, no. 2, pp. 169–181, Apr. 2008, doi: 10.1521/soco.2008.26.2.169.
- [15] A. Abubshait, A. Momen, and E. Wiese, "Pre-exposure to Ambiguous Faces Modulates Top-Down Control of Attentional Orienting to Counterpredictive Gaze Cues," *Front. Psychol.*, vol. 11, 2020, doi: 10.3389/fpsyg.2020.02234.
- [16] S. Marchesi, J. Perez-Orsorio, D. De Tommaso, and A. Wykowska, "Don't overthink: fast decision making combined with behavior variability perceived as more human-like," in *2020 29th IEEE*

International Conference on Robot and Human Interactive Communication (RO-MAN), Sep. 2020, pp. 54–59. doi: 10.1109/RO-MAN47096.2020.9223522.

- [17] E. Short, J. Hart, M. Vu, and B. Scassellati, “No fair!! An interaction with a cheating robot,” *2010 5th ACM/IEEE Int. Conf. Hum.-Robot Interact. HRI*, pp. 219–226, 2010, doi: 10.1109/HRI.2010.5453193.
- [18] D. Ullman, L. Leite, J. Phillips, J. Kim-Cohen, and B. Scassellati, “Smart Human, Smarter Robot: How Cheating Affects Perceptions of Social Agency,” *Proc. Annu. Meet. Cogn. Sci. Soc.*, vol. 36, no. 36, 2014, Accessed: Apr. 02, 2021. [Online]. Available: <https://escholarship.org/uc/item/2jh800n1>
- [19] A. Henschel, G. Laban, and E. S. Cross, “What Makes a Robot Social? A Review of Social Robots from Science Fiction to a Home or Hospital Near You,” *Curr. Robot. Rep.*, vol. 2, no. 1, pp. 9–19, Mar. 2021, doi: 10.1007/s43154-020-00035-0.
- [20] D. Lefkeli, Y. Ozbay, Z. Gürhan-Canli, and T. Eskenazi, “Competing with or Against Cosmo, the Robot: Influence of Interaction Context and Outcome on Mind Perception,” *Int. J. Soc. Robot.*, Jun. 2020, doi: 10.1007/s12369-020-00668-3.
- [21] A. Tversky and D. Kahneman, “The framing of decisions and the psychology of choice,” *Science*, vol. 211, no. 4481, p. 453, Jan. 1981, doi: 10.1126/science.7455683.
- [22] A. Keay and M. Graduand, “Emergent phenomena of robot competitions: Robot identity construction and naming,” in *Advanced Robotics and its Social Impacts*, Oct. 2011, pp. 12–15. doi: 10.1109/ARSO.2011.6301972.
- [23] K. Darling, P. Nandy, and C. Breazeal, “Empathic Concern and the Effect of Stories in Human-Robot Interaction,” in *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Kobe, Japan, 2015, pp. 770–775. doi: 10.1109/ROMAN.2015.7333675.
- [24] L. Onnasch and E. Roesler, “Anthropomorphizing Robots: The Effect of Framing in Human-Robot Collaboration,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 63, no. 1, pp. 1311–1315, Nov. 2019, doi: 10.1177/1071181319631209.
- [25] P. Gollwitzer and J. A. Bargh, *The Psychology of Action: Linking Cognition and Motivation to Behavior*. Guilford, 1996.
- [26] V. Liberman, S. M. Samuels, and L. Ross, “The Name of the Game: Predictive Power of Reputations versus Situational Labels in Determining Prisoner’s Dilemma Game Moves,” *Pers. Soc. Psychol. Bull.*, vol. 30, no. 9, pp. 1175–1185, Sep. 2004, doi: 10.1177/0146167204264004.
- [27] S. Lindenberg, “Social Rationality, Semi-Modularity and Goal-Framing: What Is It All About?,” *Anal. Krit.*, vol. 30, no. 2, pp. 669–687, 2008, doi: 10.1515/auk-2008-0217.
- [28] G. Metta *et al.*, “The iCub humanoid robot: An open-systems platform for research in cognitive development,” *Soc. Cogn. Babies Robots*, vol. 23, no. 8, pp. 1125–1134, Oct. 2010, doi: 10.1016/j.neunet.2010.08.010.
- [29] E. Wiese, P. P. Weis, Y. Bigman, K. Kapsaskis, and K. Gray, “It’s a Match: Task Assignment in Human–Robot Collaboration Depends on Mind Perception,” *Int. J. Soc. Robot.*, Mar. 2021, doi: 10.1007/s12369-021-00771-z.
- [30] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, “The Robotic Social Attributes Scale (RoSAS): Development and Validation,” in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, New York, NY, USA, 2017, pp. 254–262. doi: 10.1145/2909824.3020208.
- [31] BusinessWire, “Logitech G Adaptive Gaming Kit Creates New Possibilities for Gamers with Accessibility Needs,” Nov. 2019.
- [32] J. Peirce *et al.*, “PsychoPy2: Experiments in behavior made easy,” *Behav. Res. Methods*, vol. 51, no. 1, pp. 195–203, Feb. 2019, doi: 10.3758/s13428-018-01193-y.
- [33] G. Metta, P. Fitzpatrick, and L. Natale, “YARP: Yet Another Robot Platform,” *Int. J. Adv. Robot. Syst.*, vol. 3, no. 1, p. 8, Mar. 2006, doi: 10.5772/5761.
- [34] N. Spatola, S. Marchesi, and A. Wykowska, “The Instance Task: how to measure the mentalistic bias in human-robot interaction,” *PsyArXiv*, May 2021. doi: 10.31234/osf.io/b3wtq.
- [35] K. E. Sip, D. V. Smith, A. J. Porcelli, K. Kar, and M. R. Delgado, “Social closeness and feedback modulate susceptibility to the framing effect,” *Soc. Neurosci.*, vol. 10, no. 1, pp. 35–45, Feb. 2015, doi: 10.1080/17470919.2014.944316.
- [36] F. Eyssel and D. Kuchenbrandt, “Social categorization of social robots: Anthropomorphism as a function of robot group membership,” *Br. J. Soc. Psychol.*, vol. 51, no. 4, pp. 724–731, 2012, doi: 10.1111/j.2044-8309.2011.02082.x.
- [37] S. Wallkötter, R. Stower, A. Kappas, and G. Castellano, “A Robot by Any Other Frame: Framing and Behaviour Influence Mind Perception in Virtual but not Real-World Environments,” in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, New York, NY, USA, Mar. 2020, pp. 609–618. doi: 10.1145/3319502.3374800.
- [38] J. C. Walliser, E. J. de Visser, E. Wiese, and T. H. Shaw, “Team Structure and Team Building Improve Human–Machine Teaming With Autonomous Agents,” *J. Cogn. Eng. Decis. Mak.*, vol. 13, no. 4, pp. 258–278, Dec. 2019, doi: 10.1177/1555343419867563.
- [39] A. Abubshait, P. J. Beatty, C. G. McDonald, C. D. Hassall, O. E. Krigolson, and E. Wiese, “A win-win situation: Does familiarity with a social robot modulate feedback monitoring and learning?,” *Cogn. Affect. Behav. Neurosci.*, Apr. 2021, doi: 10.3758/s13415-021-00895-9.
- [40] C. D. Hassall, A. Silver, D. J. Turk, and O. E. Krigolson, “We are more selfish than we think: The endowment effect and reward processing within the human medial-frontal cortex,” *Q. J. Exp. Psychol.*, vol. 69, no. 9, pp. 1676–86, 2016, doi: 10.1080/17470218.2015.1091849.
- [41] N. M. Hobson and M. Inzlicht, “The mere presence of an outgroup member disrupts the brain’s feedback-monitoring system,” *Soc. Cogn. Affect. Neurosci.*, vol. 11, no. 11, pp. 1698–1706, Nov. 2016, doi: 10.1093/scan/nsw082.
- [42] K. Terada, T. Shamoto, A. Ito, and H. Mei, “Reactive movements of non-humanoid robots cause intention attribution in humans,” in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct. 2007, pp. 3715–3720. doi: 10.1109/IROS.2007.4399429.
- [43] A. Abubshait and A. Wykowska, “Repetitive Robot Behavior Impacts Perception of Intentionality and Gaze-Related Attentional Orienting,” *Front. Robot. AI*, vol. 7, 2020, doi: 10.3389/frobt.2020.565825.