Unfair game: how age and robot deception shape the attribution of mental states in virtual reality

Ludovica Misino[1] [0009-0006-7698-8803], Oronzo Parlangeli[1] [0000-0002-2040-3507], Luca Lusuardi[1] [0000-0003-4217-5768], Alessandro Innocenti[1][0000-0003-4550-4643] and Stefano Guidi[1][0000-0001-8304-8680]

[1] Università di Siena, Siena SI 53100, Italy

ludovica.misino@unisi.it
stefano.guidi@unisi.it

Abstract.
The following study investigates the dynamics of human-robot interaction (HRI) by examining how a robot's behavior (fair vs. unfair) and its perceived age (adult vs. child) influence the attribution of mental states and moral judgment in a competitive game. Using an immersive virtual reality environment, a between-subjects design was employed in which participants interacted with a robot under four different conditions, manipulating the robot's behavior and perceived age. Participants' psychological and social responses were assessed through questionnaires. Results indicate that perceived age affects responsibility and intentionality attributions, with younger robots judged less morally accountable. Additionally, unfair behavior reduces trust and likability, decreasing the interaction enjoyment. These findings emphasize the importance of designing social robots that behave in ways that are morally aligned with human expectations to foster trust and cooperation.

Keywords: human robot interaction, virtual reality, mental state attribution, perceived morality.

1 Introduction

The field of Human-Robot Interaction (HRI) has expanded rapidly, integrating robots into various aspects of human life, from healthcare and education to customer service and entertainment [8]. Beyond technical development, understanding how humans attribute agency, intentionality, and morality to robots is crucial, as these perceptions influence trust and cooperation [3, 8]. Such attributions are shaped not only by a robot's behavior but also by its physical and social characteristics, including apparent age [14, 18, 22, 27, 35, 36].

Anthropomorphism plays a central role in shaping these perceptions, as humans instinctively ascribe mental states and social attributes to robots based on their appearance and behavior [11, 13, 14, 18, 27, 32, 41]. In particular, the perceived age of a robot—whether it appears child-like or adult-like—may influence how people attribute responsibility and moral agency to its actions [22, 35, 41]. While prior research has extensively examined factors such as human-likeness, trust, and ethical decision-making in HRI [6, 27, 38, 46], the role of perceived age in shaping moral attributions remains largely unexplored. Given that younger individuals are often judged as less accountable for moral transgressions in human interactions [22], we hypothesize that this bias may extend to social robots.

To test these hypotheses, we employed immersive virtual reality (VR) to simulate realistic interactions, allowing participants to engage with virtual robots in competitive game scenarios [9, 17, 18, 26, 40]. Our findings indicate that perceived age significantly modulates moral attributions [22], with child-like robots judged as less responsible for their actions. Additionally, physical design characteristics influence perceptions of agency and trust, highlighting the importance of robot morphology in human judgments. This research contributes to a deeper understanding of the cognitive and ethical dimensions of HRI, offering practical insights for designing socially acceptable and trustworthy robotic agents. The results have direct implications for developing robots in education, healthcare, and assistive technologies, where perceptions of agency and moral responsibility can affect user trust and long-term adoption [1].

2 Related work

Social robots are designed to interact naturally with humans, adapting to social norms and expectations [3, 8]. Beyond their functional roles, social robots evoke cognitive, emotional, and moral responses, influencing how humans trust, cooperate, and attribute intentionality to them. One of the central challenges in HRI is understanding how humans ascribe mental states and agency to artificial agents [14, 18, 22, 28, 35, 36]. This process is deeply influenced by anthropomorphism, the tendency to attribute human-like qualities to non-human entities [14,18]. Anthropomorphism facilitates interaction by making robotic behavior more predictable, but it also leads to complex social and moral judgments, particularly when robots engage in unfair or ethically ambiguous behaviors [11, 13, 28, 32, 41].

2.1 Embodiment and Anthropomorphism in HRI

A crucial element in HRI is embodiment, which refers to the physical presence and perceived sensorimotor capabilities of a robotic system [8]. According to Glenberg [20], cognitive processes are inherently shaped by bodily morphology, sensory-motor systems, and emotional responses. Thus, how humans analyze and respond to a robot is closely linked to its physical presence and interactive capabilities. Embodiment is also tightly connected to anthropomorphization. When robots enter human social spaces, humans naturally project their interpretations onto robot actions [11, 13, 14, 18, 28, 32, 41, 42]. Research suggests that the more a robot appears human-like, the stronger the tendency to apply social norms and expectations to it [6]. However, high anthropomorphism can also create unrealistic expectations regarding a robot's competence, emotions, and morality. The Uncanny Valley Hypothesis [31] suggests that as robots become more human-like, they initially elicit positive emotional responses, but beyond a certain threshold, they provoke discomfort or repulsion.

As Foner [18] pointed out, in human-computer interaction, excessive anthropomorphism may lead to disillusionment when the system fails to meet human-like expectations. Beyond behavior, a critical factor shaping human perceptions of robots is their physical design, particularly their perceived age [22, 35, 41].

2.2 Theory of Mind and Perceived Moral Agency in HRI

A fundamental cognitive framework for interpreting robot behavior is Theory of Mind (ToM), which refers to the human ability to infer the mental states of others, including beliefs, intentions, and desires [4, 37]. Research in cognitive neuroscience has shown that the same neural circuits involved in human social cognition—such as the temporoparietal junction (TPJ) and medial prefrontal cortex—are also activated during interactions with robots [24, 12, 25]; this suggests that, under specific conditions, humans process robots in ways similar to human agents.

While ToM focuses on cognitive attributions of intentionality and agency, a related concept in HRI is Perceived Moral Agency (PMA), which refers to the extent to which humans attribute moral responsibility to artificial agents based on their behavior [44]. Research suggests that robots demonstrating ethical behavior foster greater trust and social acceptance, whereas those engaging in deceptive or morally ambiguous actions may trigger negative perceptions while also increasing engagement [2, 10, 44]. A key study by Short et al. (2010) [39] investigated how people react to deceptive robot behavior in a competitive game (rock, paper and scissors). Their findings revealed that action-based deception (cheating through movement) elicited stronger attributions of intentionality and moral responsibility compared to verbal deception. Interestingly, despite perceiving the deceptive robot as untrustworthy, participants engaged more with it, suggesting a paradox where morally ambiguous behavior enhances social interaction. However, Short et al. (2010) [**39**] did not consider whether the robot's perceived age modulates these moral judgments—a gap this study aims to address.

2.3 Physical Design and Perceived Age of Robots

More recent studies [22, 32, 35] demonstrate that subtle morphological variations — such as head-to-body ratio — can significantly influence perceptions of age and mind attributions. Guidi et al. (2021) [22] explored the impact of robot proportions on perceived age, showing that increasing the head-to-body ratio makes robots appear younger and cuter. The features that give robots a human-like appearance is analysed in the ABOT (Anthropomorphic roBOT) dataset, which is a systematically curated collection of images representing real-world humanoid robots [36]. Research using this dataset [35] has shown that head-to-body ratio, limb proportions, and facial expressivity significantly shape how humans interpret a robot's age; for instance, robots with larger heads relative to their bodies are often perceived as younger and more child-like, whereas those with more proportionate or elongated features are associated with an adult-like appearance. Further studies have shown that higher perceived age of robots is associated with higher perceived agency and lower perceived experience [32]. However, the implications of perceived age for moral attributions and trust remain largely unexplored.

2.4 Importance of Virtual Reality in HRI Research

Virtual Reality (VR) has emerged as a powerful tool in HRI research, offering immersive environments where human-robot interactions can be studied under controlled yet ecologically valid conditions [9, 17, 19, 26, 40]. VR allows researchers to manipulate robot attributes, such as appearance and behavior, in a systematic manner, providing insights into how these factors influence moral attributions and trust. However, despite its advantages, VR presents challenges. The lack of physical embodiment may alter user responses compared to real-world interactions. Additionally, immersion levels, avatar realism, and motion fidelity influence user perceptions of agency and trustworthiness in virtual robots [15]. Nevertheless, VR remains an invaluable tool for exploring human responses to robotic agents, allowing for precise control of experimental variables while avoiding limitations associated with physical robot interaction.

2.5 The study

While previous research has examined robot deception [2, 10, 39, 44] and physical design's effect on perceived age [22, 35], the intersection of these two domains remains underexplored. Specifically, how does a robot's perceived age influence moral attributions, agency judgments, and trust when engaging in unfair behavior? By integrating insights from moral psychology, ToM research, and robot design, this work offers new perspectives on HRI and provides practical implications for developing robots that align with human social and ethical expectations.

2.6 Research Questions and Hypotheses

This study is structured around two main axes: the impact of robot behavior (fair vs. cheating) and the role of perceived age (adult-like vs. child-like appearance). We formulated the following research questions and hypothesis:

RQ1.1: To what extent fairness (as opposed to unfairness) in the behavior of a robot influences attributions intentionality, mental states, and humanlikeness to the robot? **H1.1**: Participants who interact with cheating robots will attribute more intentionality and mental states to the robot, and will consider it more similar to a human than participants who interact with a robot that acts fairly.

RQ1.2: To what extent are the attributions of moral abilities and rights to a robot influenced by the fairness of its behavior? **H1.2**: Levels of perceived morality of the robot will be lower when the robot is cheating than when it's acting fairly.

RQ1.3: To what extent the fairness of a robot's behavior influences the evaluation of the overall interaction experience with the robot and the desire for future interactions? **H1.3**: The overall interaction experience will be evaluated less positively when the robot behaves unfairly.

These first three questions aim to assess the impact of robot behavior on users' cognitive and emotional responses. A second set of research questions concerns the role of the robot's perceived age:

RQ2.1: To what extent the perceived age of a robot influences the attributions of intentionality, mental states and humanlikeness? **H2.1**: A child-like robot will receive significantly lower attributions of intentionality, mental states and humanlikeness than an adult-like robot.

RQ2.2: To what extent the perceived age of a robot influences the attributions of moral abilities and rights to the robot? **H2.1**: A child-like robot will be judged as less morally responsible for its action than an adult-like robot.

RQ2.3: To what extent does the perceived age of a robot moderate the assessment of its behaviour? **H2.3**: Participants who interact with a child-like robot that cheats will perceive the interaction less negatively than those who interact with the cheating adult - like robot.

A third set of questions concerns the effect of the individual tendency to anthropomorphisms in the evaluation of the robot.

RQ3.1: To what extent the individual tendency to anthropomorphism influences the attributions of mental states, moral capabilities and rights to a robot, and the overall evaluation of the users' interaction with it? **H3.1**: Participants with a higher tendency to anthropomorphism will attribute more mental states, moral capabilities and rights to the robots, and will evaluate more positively the interaction with it.

RQ3.2: To what extent does the individual tendency to anthropomorphism moderate the effect of robot behavior and age on the attributions of mental states, moral capabilities and rights to a robot, and the overall evaluation of the users' interaction with it? This question was exploratory, and therefore we do not have specific predictions to test.

3 Methods

3.1 Experimental Design

The experiment was structured following a 2x2 between subjects factorial design with two independent variables: the *behavior* of a robot in a series of rounds of a competitive game (fair or occasionally cheating), and the perceived age of the robot (child or adult). The game was rock, paper and scissor, and was adapted from [39]. The Rock-Paper-Scissors game was chosen due to its strong precedent in HRI research, particularly in the study by Short et al. (2010), allowing for direct comparison with prior findings. Additionally, its simplicity and visual clarity made it well-suited for immersive VR, where hand-tracking and gesture-based interaction could enhance the realism of the robot's behavior.

The robots used in this study were adapted from the first and fourth conditions of Guidi et al. [22], a study investigating the influence of body proportions on perceived robot age and agency. The original study demonstrated that changes in head-to-body ratio and limb proportions significantly affect the perceived cuteness and maturity of humanoid robots. For this experiment, two robot models were employed:

1. Adult-like robot: based on the control condition (group 1) of Guidi et al. [22], originally 170 cm tall, but modified to 190 cm for this study. The head-to-body ratio and limb proportions remained unchanged to maintain an adult-like appearance.
2. Child-like robot: based on condition 4 of Guidi et al. [22], originally 120 cm tall. The model was adjusted to 100 cm, with an increased head-to-body ratio (+30%) and reduced limb size (-20%) with respect to the adult-like robot, enhancing its child-like features.

These physical modifications were validated through a pre-test that involved 20 participants, ensuring that they recognized the age differences between the two robots. The design and proportions of the robots used in the present study are illustrated in Figure 1, showing the distinct characteristics that define the adult and child conditions.

In the cheating condition, the robot initially plays honestly. However, in certain rounds, after one or more losing turns, if its current move turns out to be a losing one it performs a rapid arm movement to change it into the winning move, and finally declares the false victory.

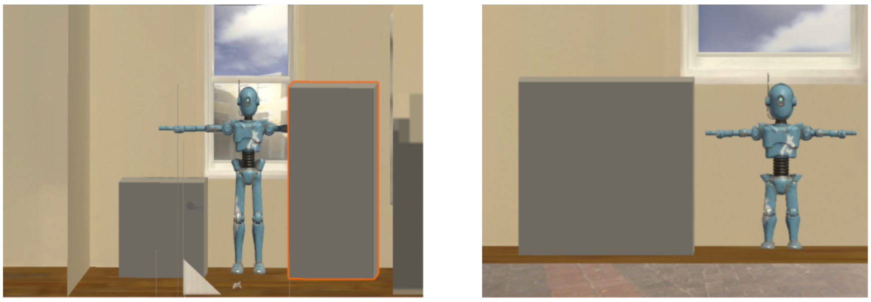

Fig. 1. Screenshot of the 3D models of the robots for the adult (left) and child (right) conditions.

3.2 Participants

The study involved a total of 82 participants, primarily recruited from the student body of the Anonymous University. The recruitment process was managed through the university's online platform, which allowed for participant registration and scheduling of experimental sessions. The decision to recruit primarily university students was based on logistical constraints related to the use of immersive virtual reality headsets, which required in-person participation. University students were readily accessible and represented a population accustomed to using digital technologies, which helped ensure a smooth and reliable VR experience. Moreover, the use of student samples is a well-established practice in experimental psychology, particularly in exploratory studies involving emerging technologies such as virtual reality. Participants received €8 as a compensation for their participation in the study. The average age of participants was

23.2 years (SD = 2.86). Despite attempts to maintain gender balance, the final sample consisted of 62.2% female and 36.6% male participants, the rest chose not to answer. Regarding educational background, most participants (47.6%) held a high school diploma, likely due to the predominance of undergraduate students in the sample. Additionally, 36.6% had obtained a bachelor's degree, while 12.2% held a master's degree.

3.3 Materials: Virtual Reality Setup and Questionnaires

The experiment was conducted using Meta Quest 2 and Meta Quest 3 VR headsets, which allowed a high level of immersion and experimental control. The VR Lab at Anonymous Lab (Anonymous University) developed the virtual environment to minimize external interference and ensure a consistent experience for all participants. After the interaction in VR, participants completed an online questionnaire divided into five sections which is reported in the "Post-interaction questionnaire" found in the supplementary material. The first measured the attribution of mental states to the robot using 18 items derived from Gray et al. [21] answered on 7-point Likert scales to investigate how individuals attribute mental states, emotional experience, perceived agency and sociability to robotic agents. 11 items measured perceived Experience (the ability to have feelings) and 7 items measured perceived Agency (having intentions, free will, and being able to pursue goals) [21]. The second section included the Perceived Moral Agency Scale, a tool developed to assess how people perceive moral agency in interactive systems such as chatbots or social robots [5], which comprises 10 items rated on 7-point agreement scales, which measure two constructs: perceived Morality (6 items) and Dependency (4 items). The third section investigates the moral status and rights of robots with questions drawn from a study by Lima G. et al. [29], it employs 11 items answered on 7-point agreement scales. The fourth section includes 5 items on a semantic differential scale designed to assess the robot's impression of likeability [7] and 4 items asking about trust, desirability of interaction, acceptability of collaboration, and realism on 7-point Likert scales [33]. In the last section of the questionnaire there are questions about the socio-demographic information of the participants accompanied by 5 items from the standardized IDAQ - Individual differences in anthropomorphism questionnaire to measure the tendency to anthropomorphize technological systems. [43].

3.4 Experimental Procedure

Prior to the main experiment, a pilot study was conducted with 8 participants to test the smoothness of the procedure and solve any technical problems.

The experiment received ethical approval (Opinion No.60/2024) from the University Ethics Committee.

The experimental procedure was divided into three main phases. Initially, participants received a general explanation of the study and an informed consent form was signed by them, guaranteeing anonymity and the possibility of withdrawal. In the second phase, the actual virtual reality interaction took place in which participants played several rounds of Rock-Paper-Scissors against a virtual robot. The number of rounds was fixed to 20 in the fair robot condition, and varied in the cheating condition, in which sometimes the robot would show a first move and, if losing, perform a rapid arm movement to modify it. The “action cheat” follows a pseudo-random logic, inspired by the study of Short et al. [39], but adapted to avoid predictable patterns. More specifically, cheating occurs between rounds 5 and 42; a variable delay regulates cheating: first cheating: 5th round, second: after 3rd round, third: after 4th round, fourth: after 2 rounds. fifth: after 5 rounds, if necessary. If the robot wins honestly in the rounds scheduled for cheating, cheating is postponed to the next round lost. A more detailed description is available 1 in the supplementary material *The cheating algorithm*.

The robot in the cheating condition is programmed to cheat 5 times before ending the game and starts cheating only after actually losing a series of rounds so that it appears to be an intentional and not a random choice; the 'cheating took place at predetermined times (e.g. 5th, 8th, 15th rounds). The final phase involved computer administration of post experience questionnaires in which participants rated the robot and overall interaction.

4 Results

4.1 Group Equivalence and Design Checks

Following the random assignment of participants to the experimental conditions, 20 participants were assigned to each of the adult robot conditions, and 21 to each of the child robot conditions. We compared participants’ age, gender distribution, familiarity with technologies and tendency to anthropomorphism across experimental groups. For none of the variables the distribution or the mean scores varied across groups.

4.2 Scales Consistency

We assessed the internal consistency of the robot perception scales, computing Cronbach's alpha for each scale or subscale. For all the variables except likeability (α = .71), alpha was higher than .84, showing good consistency. We therefore computed the scores for each variable averaging the scores of the corresponding items.

4.3 Effects of Robot Behavior and Age

Statistical analysis was conducted using a series of two-way factorial ANOVAs, exploring the main and interaction effects of the independent variables: Robot Behavior (control vs. cheating); Robot Age (adult-like vs. child-like). The dependent variables considered (in different models) include all the scales about the robot perception: Experience; Agency; Perceived Moral Agency subscales (Morality and Dependency); Robot Moral Rights; Godspeed Likeability (pleasantness and trust in the robot); Trust and Interaction Enjoyment; Humanlikeness (perceived humanity of the robot).

In the following paragraphs we report the results for the significant effects. The results for the tests of all the main effects and interactions for all the dependent variables are reported in Supplementary Table 1 in the supplementary material.

The results showed significant main effects of robot behaviour on perceived Experience (F = 4.50, ηp2 = .055, p = .037), desire for future interactions with the robot (F = 7.36, ηp2 = .086, p = .008) and willingness to accept help from the robot (F = 7.78, ηp2 = .091, p = .007). The main effect of behaviour was also marginally significant on perceived likeability (F = 2.83, ηp2= .034, p = .096) and trust (F = 3.35, ηp2 = .043, p = .064). For all the variables, the mean ratings were lower for the cheating (M_{Exp} = 1.73, M_{Like} = 3.08, M_{Trust} = 2.94, M_{Inter} = 3.54, M_{Help} = 3.95) than for the fair robot (M_{Exp} = 2.36, M_{Like} = 3.36, M_{Trust} = 3.61, M_{Inter} = 4.61, M_{Help} = 5.22). The plots of the marginal means as a function of robot behaviour are presented in figure 2 below.

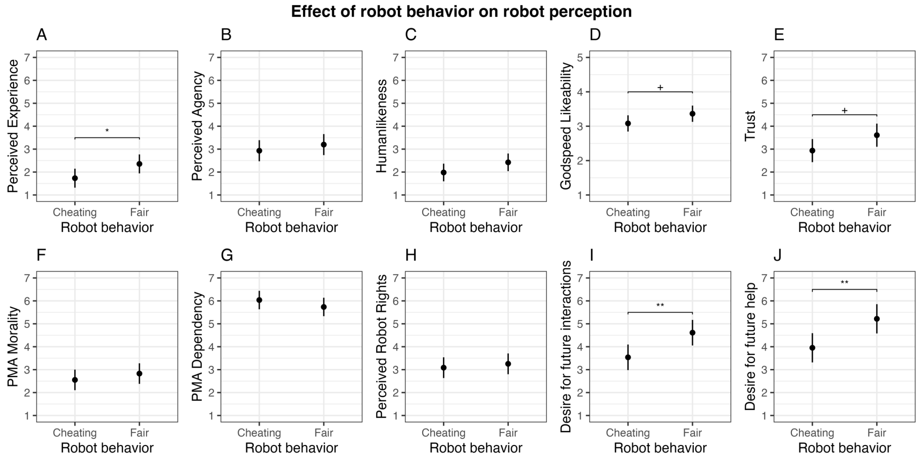

Fig. 2. Plots of the marginal means of the dependent variables as a function of Robot Behavior. Error bars are 95% confidence intervals for the means. A. Experience, B. Agency, C. Humanlikeness. D. Likeability. E. Trust, F. PMA Morality, G. PMA Dependency, H. Robot Moral Rights, I. Desire for future interactions, J. Desire for future help.

Concerning robot age, the results (Figure 3) showed a significant main effect of age on perceived morality (F = 4.29, ηp2 = .052, p = .042), resemblance to a human being (F = 4.75, ηp2 = .057, p = .032), and, marginally, perceived experience (F = 3.09, ηp2 = .038, p = .083). The child robot was considered having less perceived morality (M = 2.36), less resembling a human being (M = 1.9) and having less experience (M = 1.79) than the adult robot ($M_{Morality}$ = 3.02, M_{Human} = 2.5, M_{Exp} = 2.3). No other significant main effect or interaction was found on any of the dependent variables considered.

4.4 Effects of the Tendency to Anthropomorphism

To explore the effect of the tendency to anthropomorphism on judgments about the robot, we first computed the correlations between participants' IDAQ scores and the ratings for all the robot perception variables. The IDAQ was significantly and positively correlated with the levels of experience (r = 0.51) and agency (r = 0.42) attributed to the robot, with the perceived morality (r = 0.42) and with the level of moral rights (r = 0.48). It was instead negatively correlated with the (PMA) perceived dependency (r = -0.57).

We then repeated the ANOVAs including participants' standardized IDAQ score as a covariate. The table with the tests of all the effects is reported in the supplementary material (Supplementary Table 2). The results of the ANCOVAs showed that IDAQ was significantly associated with all the dependent variables but likeability. When controlling for the tendency for anthropomorphisms, significant main effects of robot behavior were found only for willingness to interact with the robot (F = 6.30, ηp2 = .076, p = .014) and willingness to accept help from it (F = 6.68, ηp2 = .080, p = .012), and marginally also for perceived experience (F = 3.37, ηp2 = .042, p = .070). The main effect of age, instead, was only marginally significant for human-likeness ratings (F = 4.29, ηp2 = .052, p = .081). No other significant main effect or interaction was found on any other dependent variable in the ANCOVAs.

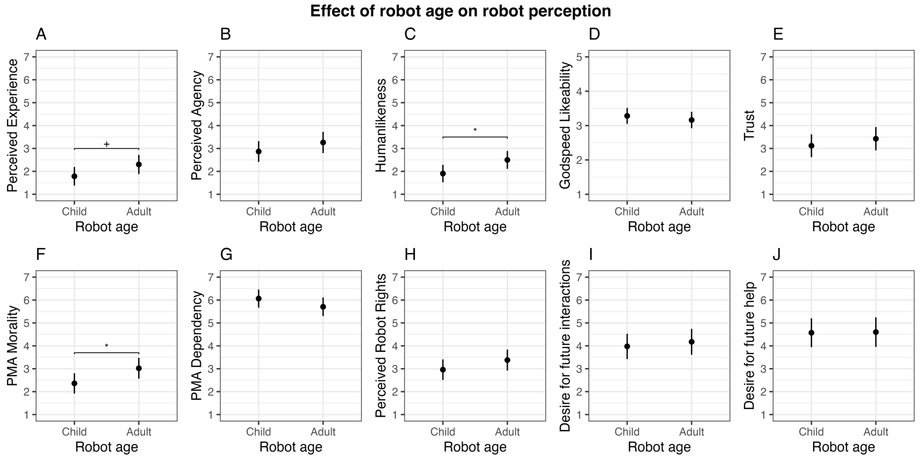

Fig. 3. Plots of the marginal means of the dependent variables as a function of Robot Age. Error bars are 95% confidence intervals for the means. A. Experience, B. Agency, C. Humanlikeness. D. Likeability. E. Trust, F. PMA Morality, G. PMA Dependency, H. Robot Moral Rights, I. Desire for future interactions, J. Desire for future help.

Lastly, to explore the possible moderating role of the tendency to anthropomorphisms on the effects of the robot behavior and age, we first classified participants as having low or high tendency to anthropomorphisms based on the comparison of their IDAQ scores to the median score for that variable on the sample (M = 1.8). We then conducted a further series of ANOVAs (one for each dependent variable), including the IDAQ level (low vs high) as a predictor, and allowing it to interact with the other design factors. The table with the tests of all the effects is reported in the supplementary material (Supplementary Tables 3A and 3B).

The results showed a significant robot behaviour by IDAQ level interaction on the willingness to accept help (F = 5.43, ηp2 = .068, p = .023), and significant robot behaviour by robot age by IDAQ interactions on experience (F = 4.15, ηp2 = .053, p = .045) and moral rights (F = 5.44, ηp2 = .068, p = .022). Concerning the willingness to accept help (Figure 4.A), pairwise comparisons of the marginal mean (averaged across robot age) showed that only for participants with low IDAQ the ratings were significantly lower for the cheating robot (M = 3.21, SE = 0.38) than for the fair robot (M = 5.20, SE = 0.44, t(74) = -3.43, p = .001), while no differences were found for high-IDAQ participants (M_{cheat} = 5.33, SE = 0.53, M_{fair} = 5.24, SE = 0.43, t(74) = 0.14, p = .887).

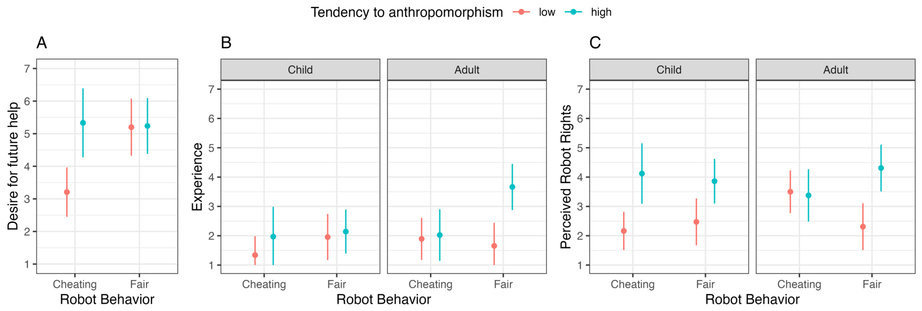

Fig. 4. Plots of the marginal means of (A) Desire for future help as a function of robot behavior and tendency for anthropomorphism, (B) Experience and (C) attribution of moral rights as a function of robot behavior, robot age and tendency for anthropomorphism. Error bars are 95% confidence intervals for the means.

Concerning the three-way interactions, the test of the simple effects of robot behavior on experience showed (Figure 4.B) that only for participants with high IDAQ the adult cheating robot (M = 2.02, SE = 0.44), but not the child one, was rated as having significantly less experience than the fair robot (M = 3.66, SE = 0.39, t(74) = -2.77, p = .007). The tests of the simple effects of robot behavior on the attributions of moral rights to the robot (Figure 4.C), instead, showed that only for low-IDAQ participants, and for an adult robot, were attributed more rights to the cheating robot (M_{cheat} = 3.5, SE = 0.37) than to the fair robot (M_{fair} = 2.31, SE = 0.40, t(74) = 2.19, p = .031). This difference (for the adult robot) was not significant for high-IDAQ participants, and it was in the opposite direction (M_{cheat} = 3.38, SE = 0.45, M_{fair} = 4.31, SE = 0.40, t(74) = -1.55, p = .124) while for a child robot it was not significant neither for high-IDAQ (p = .685) nor for low-IDAQ participants (p = .552).

5 Discussion

The present study was conducted to test a series of research questions and associated hypotheses about the role of robot behavior and robot perceived age on the attribution of mental states, moral capabilities, moral rights to the robot, and on several HRI dimensions.

We predicted that a robot behaving unfairly (e.g. cheating in a game) could be attributed more mental states intentionality and humanlikeness (H1.1), less morality and rights (H1.2), and would be more negatively evaluated in the interaction and in general (H1.3) than a fair robot. The first two predictions were not confirmed by the results, which actually found that a cheating robot was seen as having less capability to feel and have emotions (Experience mind dimension) than a fair robot, but not different agency, humanlikeness and morality.

Interestingly, the hypothesis that deceptive behavior would increase attributions of intentionality (H1.1) was not supported by our data. While previous studies [39] reported stronger intentionality attributions in response to robotic deception, these divergences may stem from methodological differences. In our study, mental state attribution was assessed using a standardized self-report instrument [21], whereas Short et al. employed a modified version of the Interactive Experiences Questionnaire by Lombard & Ditton [30], analyzing open-ended responses coded by human raters. Thus, the two studies likely capture different facets of the construct of intentionality, and the discrepancy in findings reflects these conceptual and methodological distinctions.

It was also not attributed less rights than the fair robot. H1.3 was instead fully confirmed, as the cheating robot was liked and trusted less than the fair robot, and participants expressed less desire to interact with it or receive help from it. This is consistent with previous research that [39] showed how cheating robots reduce human trust and are perceived as less agreeable.

The second set of research questions concerned the effects of the robot age. We predicted that a child-like robot would be attributed less mental states, intentionality and humanlikeness (H2.1), and less morality and rights (H2.2) than a fair robot. The first prediction was partially confirmed by the results, which found that a child-robot was seen as less like a human and tended to be attributed less experience (but not less agency) than an adult robot.

This is partially in contrast with previous findings [23] that did not find differences in the perceived levels of agency and experience of child-like and adult robots. However, in that study the interaction with the robot was extremely limited, and thus it is possible that the effect of age requires an extended and more complex interaction to manifest itself. This finding, interestingly, contrasts with the results of [34] which found a negative relationship between the age and perceived experience of 80 robots from the ABOT database. This inconsistency could be due to the different type and range of stimuli used in the experiments.

Evidence for H2.2 was mixed, as a child robot was indeed attributed less morality, but not less moral agency or rights than an adult robot, although the pattern of means was in line with the hypothesis for all the variables. We then predicted (H2.3) that robot age could moderate the effect of robot behavior, but this hypothesis was not confirmed by the results, which did not find any significant interaction between robot age and robot behavior.

The third set of questions regarded the role of the tendency for anthropomorphism on robot perception and interaction evaluation. We predicted that participants higher in this individual trait could express higher ratings for all the variables (and lower for dependency) (H3.1), and this prediction was clearly confirmed by the significant association of IDAQ with almost all the dependent variables, consistently with previous research [23, 34, 45]. The size of these effects, moreover, tended to be moderate and strong.

Lastly, we were interested in the possible moderating role of the tendency for anthropomorphism in the effects of behavior and age on the robot perception. The results of the analyses uncovered interesting significant interactions that are worth considering. First of all, the perceived experience level of the robot decreased with cheating only for high IDAQ participants, and only for the adult robot. The pattern of means in Figure 4.B seems to indicate that perceived experience was in general low, and only for an adult robot behaving fairly, it was significantly increased by the tendency to anthropomorphize. The second result is that the attributions of rights to the robot were significantly higher in participants high on this trait than in participants low on it, except for an adult cheating robot. It is not clear, however, why a low tendency for anthropomorphism would bring participants to attribute an adult cheating robot more rights than to the same adult robot behaving fairly. Lastly, the results showed that willingness to accept help from the robot was reduced by the robot cheating (regardless of age) only for participants with low tendency to anthropomorphize. It is possible that anthropomorphizing the robot would make participants more tolerant of the robot's cheating behavior, but it is not clear why no significant effect of cheating was found in participants having greater tendency for anthropomorphism. We must however notice that we did not have predictions related to the moderating role of the tendency for anthropomorphism, and therefore these results are only exploratory and should be further investigated and empirically tested.

6 Conclusions

The study is part of a strand of research examining the attribution of mental states to robots and the influence of their behavior on social trust and acceptance. The results of this study highlight that robot behavior is a determining factor in human perception. Confirming findings from previous research on transparency and honesty in human-robot interactions, we showed that a robot's cheating has a negative impact on the trust and likability of the interaction.

Recent work by Dula, Rosero & Phillips (2023) [16] on *dark patterns* in social robots warns that a child-like appearance can be exploited to conceal manipulative behavior and engender undue trust. We therefore argue that designers should avoid using youthfulness as a strategy to mask deception, and instead adopt ethical guidelines—such as behavioral transparency, clear indication of scripted behaviors, and user briefing—to prevent misuse in HRI applications.

We extend previous findings showing that a cheating robot is also seen as having less experience, and that is less wanted for future interactions or assistance. Moreover, the fact that we replicated in an immersive VR environment findings obtained with real robots indicates that VR can be a valuable tool for the study of human-robot interaction, as it allows highly controlled experiments without the limitations of interaction with physical robots. However, it is necessary to investigate how also the new findings obtained in VR in our study are transferable to interactions with real robots.

One of the most interesting findings in our study concerns the effect of the robot's perceived age. The data show that robots with childlike features are perceived as less morally capable than adult robots, suggesting a parallel with social judgment mechanisms applied to humans. This result extends previous findings about the perception of agency in artificial agents. Age, however, does not seem to significantly mitigate the negative judgment toward cheating, suggesting that dishonesty is perceived as an inherent characteristic of the agent, regardless of its apparent maturity. But the fact that child-like robots tended to be seen as having less experience and less moral capability, suggest that child robots could also be seen as less responsible and accountable for violations and misconduct than adult robots.

The results of this study have important implications for the design of social robots, especially in contexts where trust and transparency are crucial (e.g., health care, education). A robot designed to interact with humans must meet human moral expectations. Our research shows that robot behavior significantly affects trust, indicating that interactions must be clear and honest to avoid negative effects on agent perception. In addition, the design of a robot influences how it is perceived and judged. A robot with childlike features may be seen as less responsible for its actions, which may be relevant in educational or therapeutic applications. However, if the goal is to promote interaction based on social rules and shared morality, a design that is too childlike may reduce the perception of its accountability.

Limitations

This study has some limitations. First, the sample is small and not very diverse because the participants were mainly college students, with relatively high familiarity with technology. This may have influenced the perception of the robot, reducing the generalizability of the results. Additionally, the gender imbalance (62% female) and the relatively homogeneous demographic characteristics of the sample may have introduced further biases in perception and judgment. These aspects should be taken into account when interpreting the findings. Future research should aim to include more diverse and balanced samples in terms of gender, age, cultural background, and technological familiarity in order to increase the external validity and generalizability of the results.

A second factor concerns the length of the interaction as some participants interacted with the robot for a longer number of rounds than originally planned, which may have affected their level of attention and involvement. Another limitation lies in the fact that we only used two robots, and this limits the generalizability of our findings. A last factor concerns the type of task used, a game typically played among children. It remains to be seen, in future studies, whether the effect of deception or misconduct on the perception of robots, could be also found in other tasks, more collaborative and having higher stakes than the competitive game used in our experiment; further studies should try to replicate this one with different tasks. Future research could also test these models in clinical or educational settings to see if the perceived age of the robot also affects interaction in more complex environments.

Acknowledgments. We thank Eva Venturini, Francesco Lomagistro and Lorenzo Pandolfi for supporting us conducting the study. This study was funded by University of Siena funds.

We thank the two anonymous reviewers.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abel, M., Buccino, G., Binkofski, F.: Perception of robotic actions and the influence of gender. *Front. Psychol.* 15, 1295279, 1–5 (2024).
2. Arkin, R.C.: Ethics of Robotic Deception. *IEEE Technol. Soc. Mag.* September, 18–19 (2018).
3. Asimov, I.: The Bicentennial Man and Other Stories. Doubleday, Garden City, NY (1976).
4. Atherton, G., Cross, L.: Seeing more than human: Autism and anthropomorphic theory of mind. *Front. Psychol.* 9, 1–18 (2018).
5. Banks, J.: A perceived moral agency scale: Development and validation of a metric for humans and social machines. *Comput. Human Behav.* 90, 363–371 (2019).
6. Bartneck, C., Hu, J.: Exploring the abuse of robots. *Interact. Stud.* 9(3), 415–433 (2008).
7. Bartneck, C., Kulić, D., Croft, E., Zoghbi, Z.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot.* 1, 71–81 (2009).
8. Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M., Šabanović, S.: *Human-Robot Interaction: an introduction*. Cambridge University Press, Cambridge (2024).
9. Biocca, F., Harms, C., Burgoon, J.K.: Toward a more robust theory and measure of social presence: review and suggested criteria. *Presence* 12(5), 456–480 (2003).
10. Briggs, G., Scheutz, M.: The case for robot disobedience. *Sci. Am.* 316(1), 44–47 (2017).
11. Broadbent, E.: Interactions with robots: The truths we reveal about ourselves. *Annu. Rev. Psychol.* 68, 627–652 (2017).
12. Carrington, S.J., Bailey, A.J.: Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Hum. Brain Mapp.* 30(8), 2313–2335 (2009).
13. Dennett, D.: *Kinds of Minds*. Basic Books, New York (1996).
14. Duffy, B.R.: Anthropomorphism and the social robot. *Robot. Auton. Syst.* 42, 177–183 (2003).
15. Duffy, J.: Trust in Second Life. *South. Econ. J.* 78(1), 53–62 (2011).
16. Dula, E., Rosero, A., Phillips, E.: Identifying dark patterns in social robot behavior. In: 2023 Systems and Information Engineering Design Symposium (SIEDS), pp. 7–12. IEEE, Charlottesville, VA, USA (2023)

17. Fiore, S.M., Harrison, G.W., Hughes, C.E., Rütstrom, E.: Virtual experiments and environmental policy. *J. Environ. Econ. Manage.* 57(1), 65–86 (2009).
18. Foner, L.: In: Duffy, B.R.: Anthropomorphism and the social robot. *Robot. Auton. Syst.* 42, 177–183 (2003).
19. Gigerenzer, G., Todd, P.M.: *Simple Heuristics That Make Us Smart*. Oxford University Press, New York (1999).
20. Glenberg, A.M.: Embodiment as a unifying perspective for psychology. *WIREs Cogn. Sci.* 1, 586–587 (2010).
21. Gray, K., Jenkins, A.C., Heberlein, A.S., Wegner, D.M.: Distortions of mind perception in psychopathology. Proc. Natl. Acad. Sci. USA 108(2), 477–479 (2011).
22. Guidi, S., Bracci, M., Currò, F., Innocenti, A., Lusuardi, L., Marchigiani, E., Palmitesta, P., Sirizzotti, M.: Not all sizes matter. The perception of robots' age and mental abilities based on their physical dimensions. In: European Conference on Cognitive Ergonomics 2024 (ECCE '24), Article 29, pp. 1–6. Association for Computing Machinery, New York, NY, USA (2024).
23. Guidi, S., Bracci, M., Currò, F., Innocenti, A., Lusuardi, L., Marchigiani, E., Palmitesta, P., Sirizzotti, M.: You look so young, you look so cute. The relationship between physical appearance, age and mental abilities in social robots. *Behaviour and Information Technology*, pp. 1–10 (2025).
24. Hortensius, R., Cross, E. S.: From automata to animate beings: The scope and limits of attributing socialness to artificial agents. *Annals of the New York Academy of Sciences* 1426(1), 93–110 (2018).
25. Hortensius, R., Kent, M., Darda, K. M., Jastrzab, L., Koldewyn, K., Ramsey, R., Cross, E. S.: Exploring the relationship between anthropomorphism and theory-of-mind in brain and behaviour. *Human Brain Mapping* 42, 4224–4241 (2021).
26. Innocenti, A.: Virtual reality experiments in economics. *Journal of Behavioral and Experimental Economics* 69, 71–77 (2017).
27. Keijsers, M., Bartneck, C.: Mindless Robots get Bullied. In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 125–126. ACM, New York (2018).
28. Kiesler, S., Goetz, J.: Mental models and cooperation with robotic assistants. In: Duffy, B. R. (ed.) *Anthropomorphism and the social robot*, pp. 177–183. Elsevier Science, Amsterdam (2003).
29. Lima, G., Kim, C., Ryu, S., Jeon, C., Cha, M.: Collecting the Public Perception of AI and Robot Rights. arXiv preprint arXiv:2008.01339 (2020).
30. Lombard, M., Ditton, T.B., Crane, D., Davis, B., Gil-Egui, G., Horvath, K., Rossman, J.: Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument. In: Presence 2000: The Third International Workshop on Presence, Delft, The Netherlands, pp. xx–xx (2000)
31. Mori, M.: The Buddha in the Robot. Charles E. Tuttle Co., Tokyo (1982).
32. Nass, C., Moon, Y.: Machines and mindlessness: social responses to computers. *Journal of Social Issues* 56(1), 81–103 (2000).
33. Parlangeli, O., Palmitesta, P., Masi, L., Tittarelli, M., Guidi, S.: It's a Long Way to Neutrality. An Evaluation of Gendered Artificial Faces. In: *International Conference on Human-Computer Interaction*, pp. 366–378. Springer, Cham (2023).
34. Perugia, G., Boor, L., van der Bij, L., Rikmenspoel, O., Foppen, R., Guidi, S.: Models of (Often) Ambivalent Robot Stereotypes: Content, Structure, and Predictors of Robots' Age and Gender Stereotypes. In: Proceedings of the 2023 ACM/IEEE International Conference

on Human-Robot Interaction (HRI '23), pp. 428–436. Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3568162.3576981

35. Perugia, G., Guidi, S., Bicchi, M., Parlangeli, O.: The shape of our bias: perceived age and gender in the humanoid robots of the ABOT database. In: *Proceedings of the 17th ACM/IEEE International Conference on Human-Robot Interaction*, pp. 273–282. ACM, New York (2022).
36. Phillips, E., Zhao, X., Ullman, D., Malle, B. F.: 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*, pp. 105–113. ACM, New York (2018).
37. Premack, D., Woodruff, G.: Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* 1(4), 515–526 (1978).
38. Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., Eimler, S. C.: An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics* 1(1), 17–34 (2013).
39. Short, E., Hart, J., Vu, M., Scassellati, B.: No fair!! An interaction with a cheating robot. In: *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*, pp. 219–226. ACM, New York (2010).
40. Smith, V. L.: Constructivist and ecological rationality in economics. *American Economic Review* 93(3), 465–508 (2003).
41. Thellman, S., De Graaf, M., Ziemke, T.: Mental State Attribution to Robots: A Systematic Review of Conceptions, Methods, and Findings. *ACM Transactions on Human-Robot Interaction* 11(4), Article 41, 51 pages (2022).
42. Watt, S.: A brief naive psychology manifesto. *Informatica* 19, 495–500 (1995).
43. Waytz, A., Cacioppo, J., Epley, N.: Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science* 5(3), 219–232 (2010).
44. Wester, J., Pohl, H., Hosio, S., Van Berkel, N.: "This chatbot would never..." Perceived moral agency of mental health chatbots. *Proceedings of the ACM on Human-Computer Interaction* 8(CSCW1), Article 133, 25 pages (2024).
45. Wullenkord, R., Lacroix, D., & Eyssel, F. Anthropomorphism and Human–Robot Interaction. In W. Barfield, Y.-H. Weng, & U. Pagallo (Eds.), The Cambridge Handbook of the Law, Policy, and Regulation for Human–Robot Interaction (1st ed., pp. 17–56). Cambridge University Press. https://doi.org/10.1017/9781009386708.005 (2024).
46. Zlotowski, J., Sumioka, H., Bartneck, C., Nishio, S., Ishiguro, H.: Understanding anthropomorphism: Anthropomorphism is not a reverse process of dehumanization. In: Keijsers, M., Bartneck, C. (eds.) *Mindless Robots get Bullied*, pp. 1–6. IEEE (2017).