When appearance meets personality: robot age and individual differences in the evaluation of cheating robots

Ludovica Misino
Università di Siena
Siena, Italy
ludovica.misino@unisi.it

Stefano Guidi
Università di Siena
Siena, Italy
stefano.guidi@unisi.it

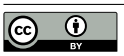
Luca Lusuardi
Università di Siena
Siena, Italy
luca.lusuardi@unisi.it

Alessandro Innocenti
Università di Siena
Siena, Italy
innocenti@unisi.it

Oronzo Parlangeli
Università di Siena
Siena, Italy
oronzo.parlangeli@unisi.it

Abstract

As robots increasingly engage in autonomous and socially meaningful actions, understanding how humans evaluate morally ambiguous robot behaviour is a central challenge in Human–Robot Interaction (HRI). This study investigates how robot age cues and individual differences influence the evaluation of deceptive behaviour displayed by robots in immersive virtual reality. Drawing on theories of mind perception and moral agency, we examine whether Dark Triad personality traits and lie acceptability shape tolerance toward robot cheating and interact with robot age (adult-like vs. child-like) in predicting perceived agency, experience, moral responsibility, and trust. Participants interacted with a humanoid robot that engaged in systematic cheating during a competitive rock–paper–scissors task in virtual reality. Results ($N = 56$) show no significant main effect of robot age on moral or cognitive evaluations, suggesting that explicit cheating behaviour may override appearance-based cues. However, significant interactions emerged between robot age and individual differences. Exploratory analyses suggested that narcissism was positively associated with perceived agency and experience, particularly in the adult-like robot condition. Similarly, lie acceptability predicted higher perceived experience and lower dependency exclusively in the adult-like condition. These findings suggest that the evaluation of robot misconduct depends on the interaction between design features and observer dispositions rather than on robot characteristics alone. Adult-like robots appear to amplify individual differences, whereas child-like robots may attenuate them. The study contributes to HRI research by highlighting the importance of considering both user traits and design cues in morally sensitive interactions, with implications for the development of socially acceptable artificial agents taking into consideration ethical concerns regarding the strategic use of dark patterns in the field of social robotics. Ultimately, these findings advocate for HRI design standards that prioritize behavioral transparency and ethical accountability.



This work is licensed under a Creative Commons Attribution 4.0 International License.
ECCE 2026, Siena, Italy
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2515-9/26/09
https://doi.org/10.1145/3822301.3822326

CCS Concepts

• **General and reference → General conference proceedings**; • **Social and professional topics → User characteristics**; • **Human-centered computing → Human computer interaction (HCI)**.

Keywords

dark patterns, dark triad, human–robot interaction (HRI), lie acceptability, mind perception, perceived moral agency, robot age, virtual reality

ACM Reference Format:
Ludovica Misino, Stefano Guidi, Luca Lusuardi, Alessandro Innocenti, and Oronzo Parlangeli. 2026. When appearance meets personality: robot age and individual differences in the evaluation of cheating robots. In *European Conference on Cognitive Ergonomics (ECCE 2026), September 15–18, 2026, Siena, Italy*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3822301.3822326

1 Introduction

Human–Robot Interaction (HRI) research has increasingly examined how humans evaluate robots not only in terms of functionality but also as social and moral agents. When robots display autonomous, socially meaningful behaviours, people readily attribute to them intentions, mental states, and moral responsibility, similarly to human–human interaction [8, 27]. A central challenge concerns the evaluation of morally ambiguous behaviors such as deception. Although deceptive behaviours may improve interaction outcomes or sustain user engagement [1, 29], they can also reduce trust and moral acceptance [19, 29]. Identifying the factors that shape tolerance toward robot deception is therefore essential for designing socially acceptable artificial agents. Research has shown that robot characteristics, including anthropomorphism and perceived age, influence attributions of intentionality and moral responsibility [10]. In immersive virtual reality, age cues modulate evaluations of robot cheating, guilt, and intentionality during competitive interactions [19, 23, 24]. However, the role of individual differences in the evaluation of robot misconduct remains largely unexplored.

Personality research indicates that socially aversive traits are associated with manipulativeness, reduced empathy, and greater acceptance of unethical behaviour [12, 26], particularly when deception serves self-interest rather than altruistic goals [31]. Specifically, Machiavellianism reflects a pragmatic view of morality, narcissism is characterised by grandiosity and entitlement, and psychopathy

by pathological lying and lack of remorse. Although these findings derive from human–human interaction, people appear to rely on similar social-cognitive mechanisms when interpreting artificial agents [27]. Accordingly, we examine whether dark personality traits and lie acceptability, a general attitude toward deception for goal attainment [22], predict tolerance toward robot cheating in immersive virtual reality. We further investigate whether these individual differences interact with robot age cues [10] in shaping Perceived Moral Agency (PMA), namely the extent to which moral responsibility is attributed to an artificial agent [2].

2 Related Work

2.1 Perceived Age and Mind Perception in HRI

The perception of robots as social actors is shaped by morphological cues that activate mechanisms underlying human social cognition [3, 21]. Even minimal anthropomorphic features promote attributions of intentions, emotions, and personality traits, with body proportions playing a key role in first impressions. In particular, a larger head-to-body ratio increases perceived youthfulness, vulnerability, and cuteness, eliciting nurturing responses consistent with the baby schema theory [7, 10]. These perceptual effects are closely linked to mind attribution. According to the two-dimensional model of mind perception, artificial agents are evaluated in terms of Agency, namely the capacity for intentional action and self-control, and Experience, that is the capacity to feel emotions and sensations [8, 23, 24]. These dimensions also shape moral judgments by influencing attributions of responsibility and blame. Perceived age systematically modulates mind perception in HRI. Adult-like robots are typically associated with greater agency, competence, and autonomy, whereas child-like robots are perceived as warmer, more emotionally capable, but less intentional [4, 14]. Consequently, child-like robots are considered less responsible for norm violations and judged less blameworthy than adult-like agents [19]. Overall, perceived age affects not only appearance but also mind attribution and moral evaluation.

2.2 Robot Deception and Perceived Moral Agency (PMA)

Robot deception represents a major ethical challenge because, although it may improve interaction, it can also undermine trust [30]. Studies using the Rock–Paper–Scissors paradigm show that action-based cheating elicits stronger attributions of intentionality and moral responsibility than verbal deception [29]. Within this framework, Perceived Moral Agency (PMA) refers to the extent to which moral responsibility is attributed to an artificial agent [2]. Importantly, people attribute deceptive intent to robots much as they do to humans, regardless of whether deception succeeds [15]. Although deception generally reduces trust, child-like robots receive more lenient moral evaluations than adult-like robots, reflecting an age-related bias [19].

2.3 Individual Differences: The Dark Triad and Lie Acceptability

While robot design has been extensively studied, less attention has been devoted to observer characteristics. The Dark Triad, Machiavellianism, Narcissism, and Psychopathy, is associated with manipulativeness, reduced empathy, and a pragmatic view of morality [5, 26]. Individuals high in these traits also show greater moral disengagement and evaluate harmful or deceptive behaviours more instrumentally [13, 20]. Narcissism, in particular, is positively associated with Lie Acceptability, namely the general acceptance of lying as a strategy for achieving personal goals [22]. Consistent with this profile, individuals high in Dark Triad traits are more tolerant of self-serving deception, especially when it is goal-directed or strategically advantageous [12, 31]. In HRI, these traits may reduce sensitivity to social cues such as vulnerability, leading to more utilitarian evaluations of cheating robots regardless of age-related characteristics.

3 The Present Study

Building on prior HRI and personality research, this study investigates whether dark personality traits and lie acceptability influence the evaluation of cheating robots in immersive virtual reality. We further examine whether robot age moderates robots' perceptions, their cognitive and moral evaluation. Given that child-like agents are typically perceived as less intentional and less morally responsible [8], and prior HRI research has shown that such agents elicit more lenient moral evaluations in cases of misconduct [19], robot age is expected to shape the cognitive and moral evaluation of cheating robots. Moreover, Dark Triad traits have been associated with reduced sensitivity to the Experience dimension and target vulnerability [9], whereas high lie acceptability promotes a more instrumental evaluation of deception regardless of the agent's characteristics [22, 31]. Based on this theoretical framework, the study addresses the following research questions:

RQ1. How does robot age influence the evaluation of cheating robots?

RQ2. How do individual differences (Dark Triad, lie acceptability) influence these evaluations?

RQ3. Do these individual differences moderate the effect of robot age?

3.1 Hypotheses

To address these questions, we formulated the following hypotheses:

H1. Adult-like robots will be attributed higher levels of perceived agency, decision-making capacity, and moral responsibility, whereas child-like robots will elicit more lenient moral evaluations.

H2. Higher levels of Dark Triad traits will attenuate the effect of robot age on moral and cognitive evaluations, such that individuals high in these traits will show smaller differences between child-like and adult-like robots.

H3. Greater tolerance of lying will attenuate the effect of robot age on robot evaluation, such that participants high in lie acceptability will show smaller differences between child-like and adult-like robots in perceived agency, experience, moral attributes, and trust.



Figure 1: Adult-like (left) and child-like (right) robot conditions used in the VR experiment.

4 Methods

4.1 Experimental Design and Participants

The study employed a two-condition between-subjects design manipulating robot perceived age (adult-like vs. child-like). Participants interacted in immersive virtual reality with a cheating robot opponent during a competitive Rock–Paper–Scissors game. Cheating behaviour followed the algorithm developed in a previous study [19], ensuring a minimum number of cheating episodes while preserving variability across game rounds. Dark Triad traits and lie acceptability were measured as dispositional predictors. Participants were recruited among university students on a voluntary basis. Data collection is currently ongoing, with a target sample of 50–80 participants. The study was approved by the Ethical Committee of the University of Siena (approval no. 4/2026).

4.2 Virtual Reality Setup

The experiment was conducted using standalone Meta Quest 2, Meta Quest 3, and Meta Quest 3S head-mounted displays, providing fully immersive VR interaction through integrated hand tracking. The virtual environment, developed in Unity, reproduced a controlled Rock-Paper-Scissors interaction with a humanoid robot. Robot appearance was manipulated to convey either adult-like or child-like characteristics through body proportions and visual features, following previous studies [10, 19]. The two experimental conditions are illustrated in Figure 1.

4.3 Measures

4.3.1 Pre-Interaction Measures. Dark personality traits were measured using the Dirty Dozen inventory in the Italian version, a 12-item scale measuring Machiavellianism, narcissism, and psychopathy [5]. Moreover, a key element of this study is the inclusion of the Revised Lie Acceptability Scale (RLAS) [22], used to quantify the user's general attitude towards the acceptability of lying as a functional strategy for achieving objectives.

4.3.2 Post-Interaction Measures. At the end of the virtual reality interaction, participants completed a structured online questionnaire designed to explore their multidimensional perception of the robotic agent. The first section examined the attribution of mental states through 18 items derived from the model by Gray et al. [9]. To specifically assess moral responsibility, the Perceived Moral Agency (PMA) Scale [2] was administered, comprising 10 items that measure the extent to which the robot is perceived as an autonomous moral agent (Morality) or the extent to which its actions are considered a mere product of programming (Dependency). The moral status of the robot was explored using 11 items relating to robot rights [17], analysing users' opinions on issues such as the right to privacy, ownership of property, and protection from cruel treatment. Lastly, participants were asked to rate the degree they would trust the robot, on a 7-point scale from 1 (not at all) to 7 (very much) [25].

4.4 Procedure

Participants first completed a pre-interaction online questionnaire and were then immersed in the virtual environment where they played a competitive Rock–Paper–Scissors game against the robot. The VR interaction lasted approximately 10 minutes (about 30 minutes including questionnaires and briefing). During the game, the robot engaged in cheating behaviour without participants being informed beforehand. After that, participants completed the post-interaction questionnaire assessing moral evaluation, mind perception, trust, and lie acceptability and were then debriefed.

5 Results

5.1 Sample Characteristics

Fifty-six participants were recruited for the experiment from the University of Siena. The mean age of participants was 23.5 (SD = 2.14), 60.7% (N = 34) were women, 37.5% (N = 21) men and 1 participant preferred not to report their gender. We compared the scores of participants for the dark traits and lie acceptability across experimental groups defined by robot type, and no significant differences were found in these characteristics, or in age and gender distribution.

5.2 Robot Perception and Individual Characteristics

We analysed the effect of robot age and individual characteristics on the perception of the robot using a MANCOVA followed by a series of ANCOVAs, one for each robot perception variable measured in the study (Agency, Experience, PMA Morality, PMA Dependency, Robot Moral Status and Right, Trust in the robot). In each model participants' standardized scores on the Dark Triad personality traits (Machiavellianism, Psychopathy and Narcissism) and on the Revised Lie Acceptance Scale (RLAS) were included as predictors in interaction with the robot type (2 levels: Adult-like and Child-like). The results of the analyses are reported in Table 1.

The results of the MANCOVA showed a marginally significant interaction between robot age and Narcissism (Pillai's trace$(1, 6)$ = 2.22, p = .060). In the univariate models, the main effect of robot type was not significant for any of the dependent variables considered, and neither were the main effects of Machiavellism and Psychopathy. Marginally significant associations were found between Narcissism and perceived Agency (B = 0.44) and between RLAS and PMA Dependency (B = −0.37). The results showed also a significant interaction between robot type and Narcissism for perceived Agency (η_g = .13), Experience (η_g = .10) and Trust (η_g = .08), a significant interaction between robot type and RLAS for Experience (η_g = .08), and, marginally, also for Agency (η_g = .07) and PMA dependency (η_g = .06). No other significant effect was found.

Table 1: Results of the tests of the main effects and interaction of the ANCOVAs of the robot perception variables collected in the study. RA = Robot Age; M = Machiavellism; P = Psychopathy; N = Narcissism; RLAS = Revised Lie Acceptance Scale.

	Experience		Agency		PMA Morality		PMA Dependency		RMSR		Trust	
Effect	**F**	**p**	**F**	**p**	**F**	**p**	**F**	**p**	**F**	**p**	**F**	**p**
Robot Age	0.01	.925	0.29	.592	0.42	.522	0.18	.677	0.32	.577	0.04	.845
Machiavellism	0.01	.915	0.36	.553	0.00	.993	0.00	.999	0.01	.906	0.02	.886
Psychopathy	0.70	.406	0.77	.386	0.41	.525	1.06	.308	2.66	.110	0.35	.558
Narcissism	1.29	.262	2.99^{+}	.090	0.51	.477	0.01	.934	0.21	.651	0.05	.831
RLAS	0.50	.485	0.50	.482	2.17	.148	3.28^{+}	.077	0.40	.532	0.54	.467
Robot Age × M	1.33	.256	1.66	.204	0.33	.567	0.13	.724	0.29	.590	0.61	.437
Robot Age × P	1.03	.315	0.33	.571	0.09	.766	1.03	.317	0.20	.659	1.16	.287
Robot Age × N	4.92^{*}	.032	6.89^{*}	.012	2.03	.161	0.37	.544	0.07	.789	4.11^{*}	.048
Robot Age × RLAS	4.14^{*}	.048	3.35^{+}	.073	1.20	.279	2.90^{+}	.095	0.58	.452	0.62	.436

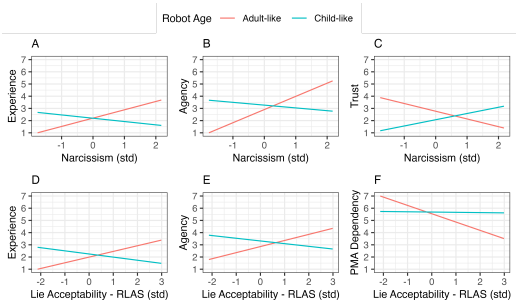


Figure 2: Plot of the marginal effects of (top row) Narcissism and (bottom row) Lie acceptability on the attributions of Experience (A, D), Agency (B, E), Trust (C) and PMA Dependency (F) to adult-like (red lines) and child-like robots (green lines). The lines represent the estimated regression lines for the effects (slopes) averaged across the other factors.

We examined the significant interactions by estimating the slopes for the effect of the covariate separately for adult-like and child-like robots, and using interaction contrast to check whether the slopes varied significantly across robot types. The results showed that Narcissism was significantly and positively associated with both perceived Agency ($B = 1.1$, $p = .01$) and perceived Experience ($B = 0.84$, $p = .043$), but only for the adult-like robot, and not for the child-like one ($p > .36$). The slopes for the effect of Narcissism on trust in the robot were not significant neither for the adult-like robot ($B = -0.63$, $p = .171$) nor for the child-like robot ($B = 0.51$, $p = .133$), but they changed significantly across robot types ($p = .048$).

Concerning the effect of participants' lie acceptability on robot perception, first of all it was never significant on the child-like robot. For the adult-like robot, instead, the results showed a marginally significant positive association between RLAS and perceived experience ($B = 0.53$, $p = .088$), and a significant negative association between RLAS and PMA Dependency ($B = -0.72$, $p = .03$).

6 Discussion

The present study examined how robot age cues and individual differences in socially aversive traits and lie acceptability shape the evaluation of cheating robots in immersive virtual reality. Given the exploratory nature of this work-in-progress contribution, the findings should be interpreted as preliminary. Contrary to H1, robot age did not exert a significant main effect on any dependent variable. This contrasts with previous findings showing that child-like robots are typically perceived as less intentional and less morally responsible [8, 19]. One possible explanation is that the salience of explicit cheating behaviour reduced sensitivity to appearance-based cues leading participants to rely more on behaviour than morphology when making moral evaluations [18, 29]. Despite the absence of a main effect, the findings provide preliminary evidence consistent with H2, showing that Dark Triad traits, particularly narcissism, moderate the relationship between robot age and evaluation. Higher narcissism was associated with greater perceived agency and experience, particularly for the adult-like robot. This is consistent with research suggesting that individuals high in narcissism adopt more agentic and instrumental interpretations of others [12, 26]. Rather than attenuating the effect of robot age as initially expected, these findings suggest that individual traits such as narcissism increase the perceived differences between adult-like and child-like robots. The results also provide preliminary evidence consistent with H3. Lie acceptability showed no main effects but interacted with robot age in predicting perceived experience and PMA dependency. Higher lie acceptability was associated with greater perceived experience and lower dependency only for the adult-like robot, consistent with the view of deception as an intentional and goal-directed strategy. Overall, these findings suggest that moral evaluations of cheating robots emerge from the interaction between robot design and observer characteristics rather than robot age alone. However, given the exploratory nature of this work-in-progress study, the multiple outcome measures, interaction analysis and current sample size require cautious interpretation. The reported effects should therefore be considered preliminary and confirmed through future studies with larger and more diverse samples.

7 Conclusion

The findings provide preliminary implications for the design of social robots operating in morally ambiguous situations. First, the results suggest that robot appearance interacts with user characteristics rather than exerting a direct effect on moral evaluation. Adult-like robots appear to amplify the influence of socially aversive

traits, whereas child-like designs may function as a psychological buffer, standardizing user responses and mitigating the aggressive or overly agentic interpretations that characterize Dark Triad profiles [26]. Second, the absence of a main effect of robot age suggests that explicit deceptive behaviour may outweigh appearance-based cues. Consequently, designers should not rely on child-like or empathic appearances to mitigate the moral consequences of unethical robot behaviour. Instead, transparency and reliability should remain central design principles [11, 28]. Furthermore, these results align with emerging concerns regarding dark patterns in social robotics, which refer to deceptive design strategies programmed to manipulate users through unidirectional emotional bonds [6]. Specifically, the strategic use of "cute" or child-like morphologies can itself constitute a dark pattern when it is intentionally employed to elicit over-attachment or to encourage 'data myopia', where users overlook privacy risks or deceptive functions due to the emotional response prompted by the interface [16]. Future research should validate these results with larger and more diverse samples and further investigate how user characteristics and robot design jointly shape moral evaluations.

7.1 Limitations

The findings should be interpreted considering several limitations. First, the current sample size may have reduced statistical power, particularly for detecting weaker main and interaction effects. Second, participants were recruited exclusively from a single university, limiting the generalizability of the findings. Third, the study relied on a competitive VR rock–paper–scissors paradigm which, while offering strong experimental control, may not fully capture the complexity of deception and moral judgment in more real-world settings. Finally, robot age was manipulated primarily through visual appearance. Future research should employ larger and more diverse samples and examine whether similar effects emerge when age is conveyed through additional cues, such as voice, language style, or interaction behavior.

Acknowledgments

We thank Eva Venturini, Francesco Lomagistro, Lorenzo Pandolfi, Benedetta Angeli and Giada Anna Benedetto for supporting us in conducting the study. This study was funded by University of Siena funds. We thank the anonymous reviewers.

8 Design Implications and Ethical Considerations

This research contributes to the ethical design of Human-Robot Interaction (HRI) by uncovering how individual personality traits, such as dark personality characteristics, and individual attitudes toward deception affect the perception of adult-like and child-like robots. While these findings can help designers develop more transparent and trustworthy robotic systems, they also highlight a potential risk of future misuse, where artificial agents could be tailored to exploit human psychological vulnerabilities or manipulate specific user profiles. To mitigate these risks, future developments should adhere to strict guidelines regarding transparency and emotional alignment in HRI.

References

[1] R. C. Arkin. 2018. Ethics of robotic deception. *IEEE Technology and Society Magazine* 37, 3 (2018), 18–19.
[2] J. Banks. 2019. A perceived moral agency scale: Development and validation of a metric for humans and social machines. *Computers in Human Behavior* 90 (2019), 363–371.
[3] C. Bartneck, T. Belpaeme, F. Eyssel, T. Kanda, M. Keijsers, and S. Šabanović. 2024. *Human-Robot Interaction: An Introduction.* Cambridge University Press.
[4] C. Breazeal. 2003. Toward sociable robots. *Robotics and Autonomous Systems* 42, 3–4 (2003), 167–175. doi:10.1016/S0921-8890(02)00373-1
[5] C. Chiorri, C. Garofalo, and P. Velotti. 2019. Does the dark triad manifest similarly in men and women? Measurement invariance of the dirty dozen across sex. *Current Psychology* 38 (2019), 659–675.
[6] E. Dula, A. Rosero, and E. Phillips. 2023. Identifying dark patterns in social robot behavior. In *Proceedings of the Systems and Information Engineering Design Symposium (SIEDS).* IEEE, 7–12.
[7] M. L. Glocker, Daniel D. Langleben, K. Ruparel, J. W. Loughead, J. N. Valdez, M. D. Griffin, N. Sachser, and R. C. Gur. 2009. Baby schema in infant faces induces cuteness perception and motivation for caretaking in adults. *Ethology* 115, 3 (2009), 257–263. doi:10.1111/j.1439-0310.2008.01603.x
[8] H. M. Gray, K. Gray, and D. M. Wegner. 2007. Dimensions of mind perception. *Science* 315, 5812 (2007), 619.
[9] K. Gray, A. C. Jenkins, A. S. Heberlein, and D. M. Wegner. 2011. Distortions of mind perception in psychopathology. *Proceedings of the National Academy of Sciences USA* 108, 2 (2011), 477–479.
[10] S. Guidi, M. Bracci, F. Currò, A. Innocenti, L. Lusuardi, E. Marchigiani, P. Palmitesta, and M. Sirizzotti. 2024. Not all sizes matter. The perception of robots' age and mental abilities based on their physical dimensions. In *Proceedings of the European Conference on Cognitive Ergonomics (ECCE).* ACM, 1–6.
[11] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, and R. Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors* 53, 5 (2011), 517–527.
[12] P. K. Jonason, M. Lyons, L. M. Baughman, and P. A. Vernon. 2014. What a tangled web we weave: The dark triad traits and deception. *Personality and Individual Differences* 70 (2014), 117–119. doi:10.1016/j.paid.2014.06.038
[13] P. K. Jonason, M. Lyons, E. Bethell, and R. Ross. 2013. Different routes to limited empathy in the sexes: Examining the links between the Dark Triad and empathy. *Personality and Individual Differences* 62 (2013), 127–131.
[14] P. H. Kahn, Jr., R. E. Gary, and S. Shen. 2012. Children's social relationships with current and near-future robots. *Child Development Perspectives* 6, 4 (2012), 373–379. doi:10.1111/cdep.12011
[15] M. Kneer. 2021. Can a robot lie? Exploring the folk concept of lying as applied to artificial agents. *Cognitive Science* 45, 10 (2021), e13032.
[16] C. Lacey and C. Caudwell. 2019. Cuteness as a 'dark pattern' in home robots. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI).* 374–381.
[17] G. Lima, C. Kim, S. Ryu, C. Jeon, and M. Cha. 2020. Collecting the public perception of AI and robot rights. *arXiv preprint arXiv:2008.01339* (2020).
[18] A. Litoiu, D. Ullman, J. Kim, and B. Scassellati. 2015. Evidence that robots trigger a cheating detector in humans. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI).* 165–172. doi:10.1145/2696454.2696456
[19] L. Misino, O. Parlangeli, L. Lusuardi, A. Innocenti, and S. Guidi. 2025. Unfair game: How age and robot deception shape the attribution of mental states in virtual reality. In *Proceedings of the International Conference on Social Robotics (ICSR).* Springer, 313–328.
[20] C. Moore, J. R. Detert, L. K. Treviño, V. L. Baker, and D. M. Mayer. 2012. Why employees do bad things: Moral disengagement and unethical organizational behavior. *Personnel Psychology* 65, 1 (2012), 1–48.
[21] C. Nass and B. Reeves. 1996. *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge University Press.
[22] C. M. Oliveira and T. R. Levine. 2008. Lie acceptability: A construct and measure. *Communication Research Reports* 25, 4 (2008), 282–288.
[23] O. Parlangeli, M. C. Caratozzolo, and S. Guidi. 2014. Multitasking and mentalizing machines: How the workload can have influence on the system comprehension. In *Engineering Psychology and Cognitive Ergonomics.* Springer, 50–58. doi:10.1007/978-3-319-07515-0_6
[24] O. Parlangeli, S. Guidi, E. Marchigiani, P. Palmitesta, A. Andreadis, and S. Roncato. 2020. How guilty is a robot who kills other robots?. In *Proceedings of the International Conference on Information, Intelligence, Systems and Applications (IISA).* IEEE, 1–5.
[25] O. Parlangeli, P. Palmitesta, L. Masi, M. Tittarelli, and S. Guidi. 2023. It's a long way to neutrality. An evaluation of gendered artificial faces. In *Proceedings of the International Conference on Human-Computer Interaction (HCII).* Springer, 366–378.
[26] D. L. Paulhus and K. M. Williams. 2002. The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality* 36, 6

(2002), 556–563. doi:10.1016/S0092-6566(02)00505-6

[27] D. Premack and G. Woodruff. 1978. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* 1, 4 (1978), 515–526.

[28] N. Shamsudhin and F. Jotterand. 2022. Social robots and dark patterns: Where does persuasion end and deception begin? In *Artificial Intelligence in Brain and Mental Health*. Springer, 89–110.

[29] E. Short, J. Hart, M. Vu, and B. Scassellati. 2010. No fair!! An interaction with a cheating robot. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM, 219–226.

[30] A. R. Wagner and R. C. Arkin. 2011. Acting deceptively: Providing robots with the capacity for deception. *International Journal of Social Robotics* 3 (2011), 5–26. doi:10.1007/s12369-010-0073-8

[31] P. G. Yarbrough and C. L. Hart. 2025. Lies in the dark: An analysis of dark personality traits and dishonesty. *Personality and Individual Differences* 240 (2025), 113161.