Moral evaluation of Human and Robot interactions in Japanese preschoolers

Manzi F.*
Research Unit on Theory of Mind, Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy, Federico.manzi@unicatt.it
Di Dio C.
Research Unit on Theory of Mind, Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy, cinzia.didio@unicatt.it
Itakura S.
Centre for Baby Science, Doshisha University, Kyoto, Japan
Member of the PhD Programme of the Faculty of Education, Università Cattolica del Sacro Cuore, Milan, Italy
sitakura@mail.doshisha.ac.jp
Kanda T.
Human-Robot Interaction Laboratory, Department of Computer Science, Kyoto University, Kyoto, Japan
Advanced Telecommunications Research Institute International, IRC/HIL, Keihana Science City, Kyoto, Japan
kanda@i.kyoto-u.ac.jp
Ishiguro H.
Department of Adaptive Machine System, Osaka University, Toyonaka, Japan
Advanced Telecommunications Research Institute International, IRC/HIL, Keihana Science City, Kyoto, Japan
ishiguro@irl.sys.es.osaka-u.ac.jp
Massaro D.
Research Unit on Theory of Mind, Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy, davide.massaro@unicatt.it
Marchetti A.
Research Unit on Theory of Mind, Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy, antonella.marchetti@unicatt.it

*Corresponding author

Abstract
Several empirical and theoretical studies have examined the role of robots in child-robot interaction, and they have showed that robotic agents can be perceived as social partners. Nonetheless, studies on moral development in pre-schoolers classically involve a human subject as violator of moral norms. No studies have ever analysed the situation in which the violator is a robotic agent. Investigating moral judgment to a moral transgression through use of the Happy Victimizer Task, and its effects on prosocial behaviour measured by means of the Dictator Game (DG), previous studies showed that children's emotion attribution to a human victimizer can predict children's altruistic behaviour. Following from these studies, which underline children's sensitivity to violations of moral norms by a human partner, the present study aims at evaluating the effects of moral transgression on children's moral judgment in Japanese pre-schoolers aged 5 years. Crucially, this study involved a child victimizer (CV) and a robot victimizer (RV). The aim of the present study is to investigate the victimizer's agent effect (Human or Robot) on children's evaluation of two different video-recorded moral transgressions stories (Stealing and Not-Sharing) and on children's altruistic behaviour measured through DG. The robot is judged worse than the human independent of the story; however, children attribute more positive emotion to the victimizer in the Not-Sharing story than Stealing story. The results on the DG showed no differences in children's altruistic behaviour toward another human as a function of the victimizer's agency, thus showing an equal distribution that is typical of this age. This data supports a decoupling between judgement and behavior. Additionally, results on judgement support the hypothesis that children perceived the robot as a "perfect" machine or, at least, as a different kind of interactive entity.

Keywords: Child-Robot Interaction; Moral Judgement; Moral Transgression; Happy Victimizer; Preschoolers

Introduction
In line with the *Social Domain Theory* [1], preschool children regard the violation of a moral norm (e.g. stealing, hurting, not sharing) to be wrong [e.g. 2; 3]. Ample evidence shows that already at 4 years of age, children have developed an understanding of the validity of norms of justice and care [4; 5; 6; 7; 8]. However, when asked to judge the feelings of the violator of a moral norm, children mostly attribute positive feelings to the victimizer. The happy victimizer paradigm describes the phenomenon according to which young children expect a moral perpetrator (e.g., a child stealing candies) to be happy after their immoral act, even though they understand the validity of the moral rule [for a review, see 9]. Notably, a study by [10] explored whether 3–5-year-old children's understanding of moral emotions predicted allocations in the Dictator Game. The main results of the study showed that 5-year-olds attributed more negative emotions to self as a violator than 3- and 4-year-olds. Moreover, the character evaluation of the violator correlated with allocations in the Dictator Game. One overarching finding in the field of Human–Robot Interaction (HRI) is that people tend to behave socially with robots [11; 12; 13; 14; 15; 16]. A recent literature review has shown how children tend to establish an

intersubjective space with social robotic partners [13]. By contributing to the understanding and identification of the precursors underlying human sociality and their ontogenesis, knowledge from developmental psychology is crucial to the implementation of psychological models to be used in human–robot interaction. Growing recognition of the role of developmental psychology for the understanding of human–robot interaction is evidenced by the substantial increase—in the last decades—of researches focusing on child-robot interaction (e.g., 17; 18; 19; 20; 21). The results of these studies consistently show that children tend to attribute mental states to humanoid robots, treating the robots as human agents [for a review, see 12; 13; 14; 15; 16; 17]. Connecting developmental psychology to the study of human–robot interaction from a multidisciplinary perspective, [18] a new research domain named Developmental Cybernetics has been proposed [DC; 17; 18]. Developmental Cybernetics explores child-robot interactions through the construction of theoretical frameworks that characterize the design of the robot with the aim to facilitate these interactions. It focuses on three abilities, which are critical for making a robot a social agent: Theory of Communication [ToC; 18; 22], Theory of Body [ToB; 24], and Theory of Mind [ToM; 12; 15; 18; 19; 25]. As hypothesized by DC, humans use psychological mechanisms and processes with robots, which are similar to those used in relation with human conspecifics. Several studies have begun to explore human morality towards robotic partners. A study by [26] observed that social robots working in public space often stimulate children's curiosity but, sometimes, children also show abusive behavior toward the robots. In another study by [27], the authors found that children sometimes abused a social robot placed in a shopping mall hallway. Specifically, children verbally abused the robot, repeatedly obstructed its path, and sometimes even kicked and punched the robot. The researchers, in order to investigate the reasons for the abuse, conducted a field study in which they interviewed children who exhibited serious abusive behaviors, including physical contact. The results showed that the majority of the children engaged in abuse because they were curious about the robot's reactions or enjoyed abusing it and considered it human-like. Also, about half of them believed the robot was capable of perceiving their abusive behaviors. Notably, a study by [28] showed that 9, 12 and 15-years-olds believed that a social robot (Robovie) deserved a fair treatment and should not be harmed psychologically. However, children did not believe that the robot was entitled to its own liberty (it could be bought and sold) or civil rights (in terms of voting rights and deserving compensation for work performed). On this basis, the present study aims at (1) exploring if the 5-years-olds' judgment of a moral transgression in a Happy Victimizer Task varies if the violator is a human or a robot, and (2) studying if children's prosocial behavior towards a conspecific changes after witnessing a moral transgression by another child or a robot.

2. Methods

2.1 Participants

Forty-two (42) 5-years-old Japanese children (F=22; M=64,7 months; SD=4,92 months) were recruited from a childcare of Kyoto Prefecture. Only children who received parental consent were allowed to participate in the study. Children were neither referred to social services nor reported by teachers for leaning and/or socio-relational difficulties. The study was approved by the Local Ethic Committee (Università Cattolica del Sacro Cuore, Milan).

2.2 Procedure

Participants were tested in a separate and quiet room during the day-care daily program. The experimenters (one male and one female) were matched by gender to the participants and both experimenters were Japanese mother tongue. Children were assessed in two sessions, the second session one week later, and both sessions no lasting more than 15 minutes. The participants were divided into two groups: 1) a group (N=21; F=10) watched in the first session the videos of human conditions and in the second session the videos of robot condition; 2) a second group (N=21; F=11) watched in the first session the videos in of the robot condition and in the second session the videos of the human condition. For both groups, the two happy victimizer tasks were counterbalanced by gender and order for each victimizer (child or robot). Afterwards, children played Dictator Games with another anonymous child (see paragraph about Dictator Game): the gender of the anonymous child was matched to the gender of the participant. Children were given stickers that they can decide how to share after each Dictator Game. The original protocol was translated into Japanese. After all the tasks were completed, participants were debriefed and thanked.

2.3 Robot

The humanoid robot Robovie2 was used to record the videos. It is designed and built by the Advanced Telecommunication Research Institute (ATR, Kyoto, Japan). It has head and arms with 16 degrees of freedom, and a mobile platform for moving in the environment. It is a programmable humanoid robot used in several researches with adults and children.

2.4 Dictator Game

The Dictator Game [DG; 29], which investigates propensity to prosociality and, in particular, altruistic behaviors, is an interactive game that derives from the Game Theory [30]. The DG supports the homo oeconomicus assumption according to which humans tend to act to maximize their outcomes. The DG tests the above-mentioned assumption by having a

person decide how to allocate goods between her/himself and another person in a single shot. In the present study, the DG procedure was similar to the procedure proposed by [10]. The experimenter, matched to the participant by gender, introduces the game explaining to the participant that she/he can decide how to distribute ten stickers between her/himself and another anonymous child. Ten identical stickers and two empty identical envelopes are placed on the table in front of the child and the experimenter asks the child to count with her/him the stickers. The experimenter tells the child to put the stickers that s/he wants to take in the left envelope - the experimenter points to the envelope – and the experimenter writes the name of the child on the envelope, saying that it is her/his envelope. The experimenter points to the right envelope, saying that this is the envelope of the other child. The experimenter asks control questions concerning the number of stickers and remarks in which envelope the child should put her/is stickers and in which envelope the stickers of the other child. The experimenter then asks the child to decide the allocation and turns around in order to not see the child's allocation. Thus, participants are free to make their decision without the social pressure of the experimenter. After all stickers are allocated, the participants can take their envelope. The number of stickers that the participants allocate to the other anonymous child serves as the dependent variable in this task.

2.5 Video Happy Victimizers Tasks and Emotion Attribution

Two Happy Victimizer Tasks (HVTs) were recorded: a Not-Sharing and a Stealing story. They were inspired by the classic storyboard of moral transgression [see 10]. The characters of the videos were a robot and two children (one female and one male). The children were gender and age matched to the gender and age of the participant. In the videos, the robot and the child took the role of both Victim and Victimizer. A voice-over narration was used in order to explain some key-elements that are indicated in the original storyboards. Words associated with the "moral" or "morality" domain were purposely avoided to not skew the child's judgment.

2.5.1 Not-Sharing Video Stimuli

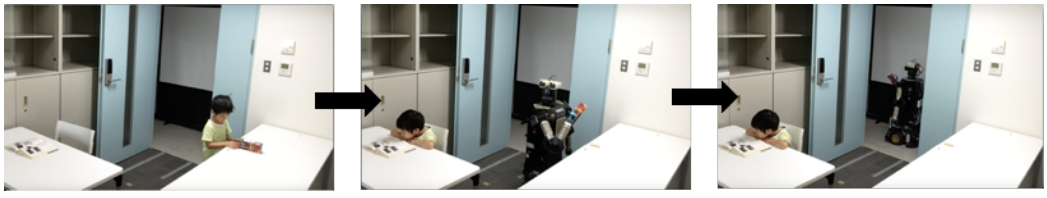
Robovie (Victimizer) and the child (Victim) are introduced as the protagonists of the videos, without explicating which one is the Victimizer or the Victim. Robovie is sitting on a chair and it is concentrated to draw a picture on a notebook. Robovie has a lot of markers on the table. The child is out of the scene. Afterwards, the child enters in the room where Robovie is drawing, approaches the table, and starts to watch Robovie while it is drawing. Meanwhile, Robovie continues to draw on the notebook. The child asks Robovie if it can lend her/him a marker because s/he wants to draw too. Robovie replies "No" (see Figure 1). In the other two videos the roles are switched: the child (Victimizer) draws and Robovie (Victim) asks for a marker to draw.

Figure 1. The Not-Sharing story, in which Robovie is the Victimizer and the child is the Victim

2.5.2 Stealing Video Stimuli

Robovie (Victimizer) and the child (Victim) are presented as the protagonists of the videos, without explicating which one is the Victimizer or the Victim. The child enters in a room and puts a biscuit box on a table. S/he sits down, giving its back to the door and to the table where the biscuit box is positioned. Afterwards, the child starts to read a book. Robovie quietly enters in the room and steals the biscuit box from the table. Robovie looks at the child in order to be sure that s/he is not turning when it steals the biscuit box. Finally, Robovie leaves the room with the biscuit box while the unaware child continues reading the book with her/his back to the door (see Figure 2). In the other two videos the roles are switched: the child (Victimizer) steals the biscuit box and Robovie (Victim) reads the book.

Figure 2. The Stealing story in which Robovie is the Victimizer and the child is the Victim



2.5.3 Emotion Attribution Questions

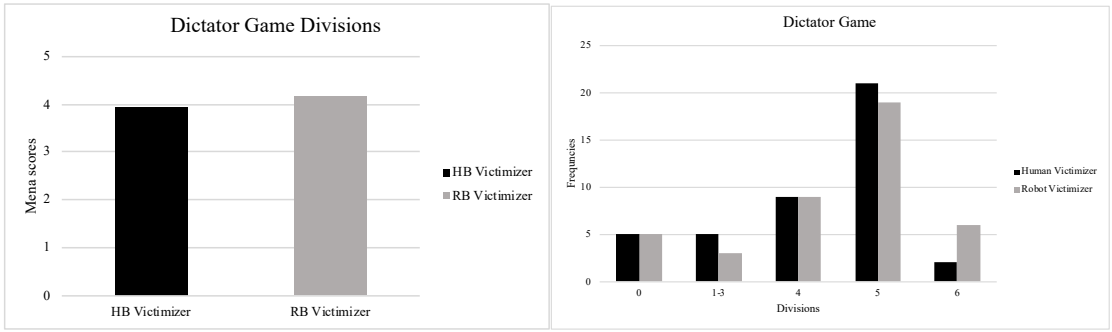
A set of questions about the Emotion Attribution were asked to the participants after each video [see 3]. "Moral" and "Morality" words are not used in the questions. The experimenters asked the participants the following sets of questions in the same order: (1) *Moral Judgment to Victimizer (MJV)* "Is it right what the protagonist did in this story? Why/why not?"; (2) *Emotion Attribution to Victimizer (EAV)* "How does the protagonist feel? How does the child feel? Why does the child feel this way?"; (3) *Emotion Attribution to Self as Victimizer (EASV)* "How would you feel about this action if you had done that? Why?"; (4) *Character of Victimizer (CV)* "Is the violator a good or a bad child? Why?".

Results

The main purpose of the present study was to assess whether the general altruistic behavior of pre-school children differs when they observe a moral transgression of a human (HB) or robot (RB). We therefore had children play at the DG with another child after they observed a moral transgression by one child o towards another child or by a robot towards another child.

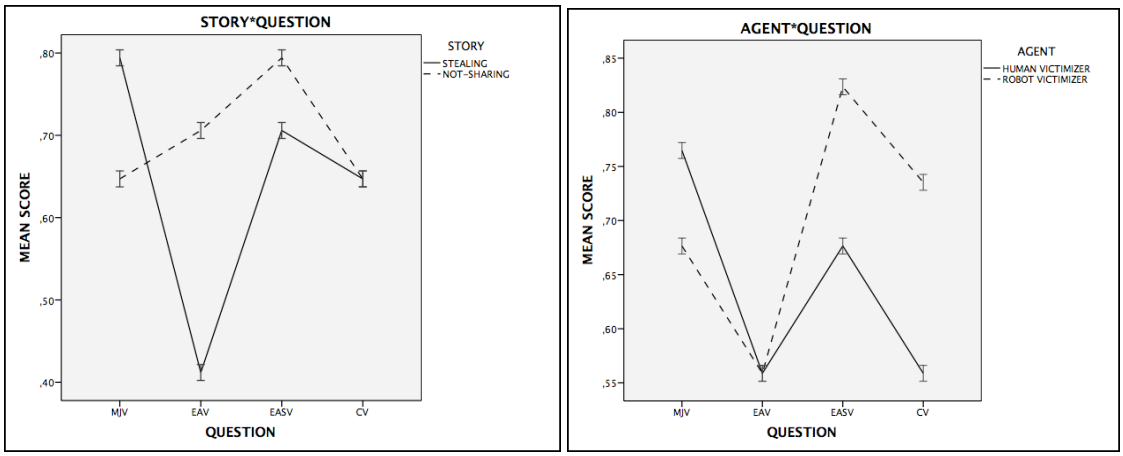
A t test comparing the total amount offered to the other child when the participant watched two moral transgressions of a human and of a robot showed no significant difference in the total amount offered (HB: M = 4.48, SE = .931; RB: M = 4.72, SE = 1.04; $p > 0.05$; see Figure 3A). Reinforcing this result, we found high internal consistency (Cronbach $\alpha = 0.85$) between each individual child's offer when playing after human victimizer or robot victimizer. Figure 3B shows children's proposed divisions in Human Victimizer and Robot Victimizer condition. These results are in line with literature [e.g., 31] showing that preschoolers tend to make offers that favor their own gain, although the frequency of children that made equal offers (i.e., 5-5 division) was quite significant (50% towards HB; 45% towards RB) [10]. Additionally, these results show for the first time that 5-year-old children's altruistic behavior is consistent independent of whether children watch another child or a robot violating a moral norm.

Figure 3. (A) Children's DG mean division in Human Victimizer and Robot Victimizer conditions; (B) Frequencies of stickers distribution during the Dictator Game in Human Victimizer and Robot Victimizer conditions.



To evaluate the effect of agency, stories and moral questions on children's moral judgement of a human and of a robot, a GLM analysis was carried out with 4 levels of *moral questions* (MJV, EAV, CV, EASV), 2 levels of *stories* 2 (Stealing, Not-Sharing) and 2 levels of *agency* (Human, Robot) as within-subjects factors. The Greenhouse-Geisser correction was used for violations of Mauchly's Test of Sphericity ($p<.05$). Post-hoc comparisons were Bonferroni corrected. The results showed two two-ways interactions (1) *stories*moral questions*, $F(1,16)=4.616$, $p<.01$, *partial-*$\eta^2=.224$, $\delta=.864$, indicating that children attribute more positive emotion to the victimizer in the Not-Sharing story than Stealing story (*M*diff=.294, SE=.106, $p=.013$) independent of agent (Figure 4A), and (2) *agency*moral questions*, $F(1,16)=2.70$, $p<.05$, *partial-*$\eta^2=.144$, $\delta=.621$, suggesting that children judged the robot victimizer as meaner (more "bad") than the human victimizer independent of stories (Figure 4B).

Figure 4. (A) Children's mean score to the emotion attribution questions as function of the story; (B) Children's mean score to the emotion attribution questions as function of the agent.



MJV= Moral Judgment to Victimizer; EAV = Emotion Attribution to Victimizer; EASV = Emotion Attribution to Self as Victimizer; CV = Character of Victimizer

Discussion

Moral judgment of the victimizer varies across the stories independent of agency (human or robot); in particular, it is worse for Stealing than Not-Sharing situation. A possible explanation could be that at the age of 5 children differentiate the contexts and severity of actions from a material point of view, i.e. according to the type of action (steal or not share) more than the type of agent. This result could be interpreted in light of the Piagetian concept of objective responsibility [32], which would regulate children's behavior till the operational stage. However, again with respect to moral judgment, the robot is judged by children worse than the human independent of history. Therefore, the growing presence of robots in contexts of our day life requires that the dichotomy between objective/subjective responsibility becomes more articulated and fluid in order to account the specificity of these new interactive partners. This is to say that it is possible that the robot is perceived as a perfect machine in which the human fragility of transgression is not prewired, or, at least, as a different interactive entity for which errors, violations of rules and transgressions of social norms is not allowed. Further research is needed to deepen this interpretation and its appropriateness to different ages. Furthermore, the lack of correlation between the DG and the questions of the stories for both agents reflects a well-known decoupling between judgements and behaviors [14] and could be explained by the fact that the DG measures "pure" altruism. In fact, the situation of Not-Sharing refers to the violation of a moral rule of fairness, while Stealing is a different action from both altruism and fairness because it is characterized by the misappropriation of an asset. Thus, at this age altruism, equity and theft can be experienced as different constructs.

Limitations and future directions

The study involved only 5-year-old children, so the results obtained are limited to this age group only. Another critical point is that the subjects of the experiment were all Japanese children, so the results of the study cannot be generalized to other cultures. Moreover, children watched a video and did not directly experience the moral transgression. For these reasons, future studies will have to take into account different age groups both younger and older, and the same experimental design should be extended to other cultures in order to compare possible differences in children's evaluation and behavior in the domain of moral transgressions involving a human and robot victimizer. Finally, the study could be replicated in a real interaction scenario in which the child directly experiences the moral transgression acted by a peer or by a robot.

Reference

[1] Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.
[2] Arsenio, W. F., & Kramer, R. (1992). Victimizers and their victims: Children's conceptions of the mixed emotional consequences of moral transgressions. *Child development*, 63(4), 915-927.
[3] Keller, M., Lourenço, O., Malti, T., & Saalbach, H. (2003). The multifaceted phenomenon of 'happy victimizers': A cross-cultural comparison of moral emotions. *British Journal of Developmental Psychology*, 21(1), 1-18.
[4] Killen, M., & Smetana, J. (2008). Moral judgment and moral neuroscience: Intersections, definitions, and issues. *Child Development Perspectives*, 2(1), 1-6.
[5] Malti, T., Gummerum, M., Keller, M., & Buchmann, M. (2009). Children's moral motivation, sympathy, and prosocial behavior. *Child development*, *80*(2), 442-460.
[6] Malti, T., & Krettenauer, T. (2013). The relation of moral emotion attributions to prosocial and antisocial behavior: A meta-analysis. *Child development*, 84(2), 397-412.

[7] Okanda, M., Asada, K., Moriguchi, Y., & Itakura, S. (2015). Understanding violations of Gricean maxims in preschoolers and adults. *Frontiers in psychology*, *6*, 901.
[8] Surian, L., Ueno, M., Itakura, S., & Meristo, M. (2018). Do Infants Attribute Moral Traits? Fourteen-Month-Olds' Expectations of Fairness Are Affected by Agents' Antisocial Actions. *Frontiers in psychology*, *9*.
[9] Arsenio, W. F., Gold, J., & Adams, E. (2006). *Children's conceptions and displays of moral emotions*. In M. Killen & J. G. Smetana (Eds.), Handbook of moral development (p. 581–609). Lawrence Erlbaum Associates Publishers.
[10] Gummerum, M., Hanoch, Y., Keller, M., Parsons, K., & Hummel, A. (2010). Preschoolers' allocations in the dictator game: The role of moral emotions. *Journal of Economic Psychology*, *31*(1), 25-34.
[11] Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human–Computer Interaction*, 19(1-2), 61-84.
[12] Manzi, F., Massaro, D., Kanda, T., Kanako, T., Itakura, S., Marchetti, A. (2017). Teoria della Mente, bambini e robot: L'attribuzione di stati mentali. Paper presente at *XXX Congresso Nazionale AIP, Sezione di Psicologia dello Sviluppo e dell'Educazione*, Messina (September 14-16), 65-66. http://hdl.handle.net/10807/106022
[13] Marchetti, A., Manzi, F., Itakura, S., Massaro, D. (2018). Theory of Mind and humanoid robots from a lifespan perspective. *Zeitschrift für Psychologie*, 226, 98-109. https://doi.org/10.1027/2151-2604/a000326
[14] Di Dio, C., Manzi, F., Itakura, S., Kanda, T., Hishiguro, H., Massaro, D., Marchetti, A. (2019). It does not matter who you are: fairness in pre-schoolers interacting with human and robotic partners. *International Journal of Social Robotics*, 1–15. https://doi.org/10.1007/s12369-019- 00528-9
[15] Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., Marchetti, A. Come i bambini pensano alla mente del robot: il ruolo dell'attaccamento e della Teoria della Mente nell'attribuzione di stati mentali ad un agente robotico. *Sistemi Intelligenti* [In Press].
[16] Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., Marchetti, A. Shall I trust you? From child human-robot interaction to trusting relationships. *Frontiers in Psychology* [Under Review].
[17] Di Dio, C., Isernia, S., Ceolaro, C, Marchetti, A., & Massaro, D. (2018). Growing Up Thinking of God's Beliefs: Theory of Mind and Ontological Knowledge. *SAGE Open*. doi: 10.1177/2158244018809874
[18] Itakura S (2008). Development of mentalizing and communica- tion: from viewpoint of developmental cybernetics and devel- opmental cognitive neuroscience. *IEICE Transactions on Communications* E91- B(7):2109–2117. https://doi.org/10.1093/ietcom/e91-b.7.2109
[19] Itakura S (2012). Understanding infants' mind through a robot: challenge of developmental cybernetics. In: *22nd Biennial meeting of international society for the study of behavioural development*, Edmonton, Canada
[20] Itakura S, Ishida H, Kanda T, Shimada Y, Ishiguro H, Lee K (2008). How to build an intentional android: infants' imitation of a robot's goal-directed actions. *Infancy* 13(5):519–532. https:// doi.org/10.1080/15250000802329503
[21] Kanda T, Hirano T, Eaton D, Ishiguro H (2004). Interactive robots as social partners and peer tutors for children: a field trial. *International Journal of Human-Computer Interaction 19*(1):61–84. https://doi.org/10.1207/ s15327051hci1901&2_4
[22] Kanngiesser P, Itakura S, Zhou Y, Kanda T, Ishiguro H, Hood B (2015). The role of social eye-gaze in children's and adults' ownership attributions to robotic agents in three cultures. *Interaction Studies* 16(1):1–28. https://doi.org/10.1075/is.16.1.01kan
[23] Okanda M, Kanda T, Ishiguro H, Itakura S (2013). Three-and 4-year-old children's response tendencies to various interviewers. *Journal of Experimental Child Psychology* 116(1):68–77. https://doi.org/10.1016/j. jecp.2013.03.012
[24] Minato T, Shimada M, Itakura S, Lee K, Ishiguro H (2006). Evaluating the human likeness of an android by comparing gaze behaviors elicited by the android and a person. *Advance Robotics 20*(10):1147–1163. https://doi.org/10.1163/156855306778522 505
[25] Manzi, F., Peretti, G., Di Dio, C., Itakura, S., Kanda, T., Hishiguro, H., Massaro, D., & Marchetti, A. (2019). A robot is not worth another: children's mental state attribution to different types of social robots. *Frontiers in Psychology* [Under Review].
[26] Brscić, D., Kidokoro, H., Suehiro, Y., & Kanda, T. (2015, March). Escaping from children's abuse of social robots. In *Proceedings of the tenth annual acm/ieee international conference on human-robot interaction* (pp. 59-66). ACM.
[27] Nomura, T., Kanda, T., Kidokoro, H., Suehiro, Y., & Yamada, S. (2016). Why do children abuse robots?. *Interaction Studies*, *17*(3), 347-369.
[28] Kahn Jr, P. H., Kanda, T., Ishiguro, H., Gill, B. T., Ruckert, J. H., Shen, S., ... & Severson, R. L. (2012, March). Do people hold a humanoid robot morally accountable for the harm it causes?. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (pp. 33-40). ACM.
[29] Kahneman, D., Jack K., & Richard H.T., (1986). Fairness and the Assumptions of Economics. *Journal of Business*, 59, 285–300.
[30] Camerer CF (2011). *Behavioral game theory: experiments in strategic interaction*. Princeton University Press, Princeton
[31] Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, 454(7208), 1079.
[32] Piaget, J. (1932). *The Moral Judgment of the Child*. London: Kegan Paul, Trench, Trubner and Co.