

Being Fair to a Robot: “Shall I Treat It Like Him?”, a Five Years Old Wonders

Cinzia Di Dio (Catholic University of the Sacred Heart, Milan), Federico Manzi (Catholic University of the Sacred Heart, Milan – University of Neuchâtel), Shoji Itakura (Kyoto University), Takayuki Kanda (ATR Hiroshi Ishiguro Laboratories and Intelligent Robotics and Communication Laboratories, Kyoto University), Hiroshi Ishiguro (ATR Hiroshi Ishiguro Laboratories and Intelligent Robotics and Communication Laboratories, Kyoto – University of Osaka), Davide Massaro (Catholic University of the Sacred Heart, Milan) and Antonella Marchetti (Catholic University of the Sacred Heart, Milan)

Studies on adults investigating human-robot interactions suggest that individuals tend to behave similarly with humans and robots (e.g., Author, 2004; Author, 2008a; Author, 2008b; Nishio et al., 2012), often attributing to the robot mentalistic competences (Author, 2017; Author 2018), and namely, a Theory of Mind (Premack and Woodruff, 1978; Wimmer and Perner, 1983). Recently, the human-robot interaction was investigated in pre-schoolers playing the Ultimatum Game (UG) - as proposers and receivers - with another child or a robot (Author, 2018). Children recognized the robot as a distinct entity, as assessed through a States Attribution questionnaire evaluating emotions, intentions and desires, imagination, epistemic, and perceptive knowledge. Nevertheless, UG results revealed a comparable “self-centred” approach with both partners, with a slight inclination for children to be fairer towards the human. The analysis of the children’s justification for their behaviour, in both proposer and receiver roles, indicated a tendency to justify fair divisions either in terms of outcome-based (i.e., quantity), or in terms of equity-based reasoning. Conversely, unfair offers were mostly justified in terms of outcome. Mentalization-based justifications were few and mostly used when children made or accepted both fair and unfair divisions to/from the robot, and uniquely when making or accepting *fair* offers to/from the child. The lack of equity and mentalization-based justifications of unfair divisions with the human partner suggests an attempt of children to resolve a social cognitive conflict emerging from the discrepancy between a “socially-expected” equal behaviour and the actual “selfish” behaviour. This is supported by use of mentalistic justifications also of unfair divisions with the robot, towards which the socio-norms of fairness (Castelli et al., 2014) could be less salient, overall suggesting an inclination of children to regard the robot as “not perfectly like-me”.

References

- Castelli, I., Massaro, D., Bicchieri, C., Chavez, A., & Marchetti, A. (2014). Fairness norms and theory of mind in an ultimatum game: judgments, offers, and decisions in school-aged children. *PloS one*, 9(8), e105024.
- Author (2018).
- Author (2008a).
- Author (2008b).
- Author (2004).
- Author (2017).
- Author (2018).
- Nishio, S., Ogawa, K., Kanakogi, Y., Itakura, S., & Ishiguro, H. (2012). *Do robot appearance and speech affect people’s attitude? Evaluation through the Ultimatum Game*. Proceedings from 2012 ROMAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, Paris. doi:10.1109/ROMAN.2012.6343851
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences*, 1(4), 515-526. doi:10.1017/S0140525X00076512
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103-128. doi:10.1016/0010-0277(83)90004-5