

*Workshop on Adapted intEraction with SociAl Robots (cAESAR) – ACM IUI 2020*

# **Moral evaluation of Human and Robot interactions in Japanese preschoolers**

Federico Manzi

Università Cattolica del Sacro Cuore di Milano

*Department of Psychology*

*Research Unit on Theory of Mind & Human-Robot Laboratory*

[www.teoriadellamente.it](http://www.teoriadellamente.it)

[federico.manzi@unicatt.it](mailto:federico.manzi@unicatt.it)



# Moral evaluation of Human and Robot interactions in Japanese preschoolers

*Manzi F.<sup>1,2</sup>, Di Dio C.<sup>1,2</sup>, Itakura S.<sup>3</sup>, Kanda T.<sup>4,6</sup>, Ishiguro H.<sup>5,6</sup>, Massaro D.<sup>1,2</sup>, Marchetti A.<sup>1,2</sup>*

<sup>1</sup>Research Unit on Theory of Mind, Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy

<sup>2</sup>Human-Robot Laboratory, Università Cattolica del Sacro Cuore, Milan, Italy

<sup>3</sup>Centre for Baby Science, Doshisha University, Kyoto, Japan

<sup>4</sup>Human-Robot Interaction Laboratory, Department of Computer Science, Kyoto University, Kyoto, Japan

<sup>5</sup>Department of Adaptive Machine System, Osaka University, Toyonaka, Japan

<sup>6</sup>Hiroshi Ishiguro Laboratories and Intelligent Robotics and Communication Laboratories, ATR, Kyoto, Japan

# Outline

- **Theoretical background**
  - Children's moral emotions
  - Children's moral attribution to robots
- **Experimental question**
- **The study**
  - Methods
  - Results
  - Concluding remarks
  - Limitations

# Theoretical Background – Children's moral judgement

- In line with the *Social Domain Theory* (Turiel, 1983), preschool children regard the violation of a moral norm (e.g. stealing, hurting, not sharing) to be wrong (e.g. Arsenio & Kramer, 1992; Keller, Lourenco, Malti, & Saalbach, 2003)
- Ample evidence shows that already at 4 years of age, children have developed an understanding of the validity of norms of justice and care (Smetana and Killen, 2008)

# Theoretical Background – Children's moral emotions in the Happy Victimizer

- 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 describes the phenomenon that young children expect a moral perpetrator (e.g., a child stealing candy) to be happy, even though they understand the validity of the moral rule (for a review, see Arsenio et al., 2006)

# Children's moral emotions and Altruistic Behaviour

- *Gummerum, Hanoch, Keller, Parsons and Hummel (2010)* explores whether 3–5-year-old children's understanding of moral emotions predicted allocations in the Dictator Game
- Main results:
  1. 5-year-olds attributed more negative emotions to self as violator than 3- and 4-year-olds
  2. Character evaluation of the violator correlated with allocations in the Dictator Game

# Human Robot Interaction

- One overarching finding in the field of Human–Robot Interaction (HRI) is that people tend to behave socially with robots (Breazeal, 2018; Kanda, Hirano, Eaton, & Ishiguro, 2004; Manzi et al., 2017; Marchetti, Manzi et al., 2018; Di Dio, Manzi et al., 2019; Di Dio, Manzi et al., submitted)
- A recent literature review has shown how children tend to establish an intersubjective space with social robotic partners (Marchetti, Manzi, Itakura and Massaro, 2018)

# Human Robot Interaction and Children's Moral Attribution

*Kahn, Kanda, Ishiguro et al. (2012)*

- The study "*Robovie, You'll Have to Go into the Closet Now: Children's Social and Moral Relationships With a Humanoid Robot*" showed that 9, 12 and 15-years-olds believed that Robovie deserved a fair treatment and should not be harmed psychologically
- But children did not believe that Robovie was entitled to its own liberty (Robovie could be bought and sold) or civil rights (in terms of voting rights and deserving compensation for work performed) – Manzi et al., 2017; Di Dio et al., 2018; Di Dio, Manzi et al., 2019; Di Dio, Manzi et al., submitted



# Research Questions

*Does 5-years-olds' judgment of a moral transgression vary if the violator is a human or a robot?*

*Is children's moral judgment associated with prosocial behavior towards a conspecific after attending to a moral transgression done by another child or a robot?*

# Participants

- (42) 5-years-old Japanese children (F=22; M=64,7 months; SD=4,92 months)

# Measure – Happy Victimizer Task

- A set of videos inspired by the classic stories (*Stealing and Not-Sharing*) of the Happy Victimizer Tasks was recorded (Keller et al., 2003)
- Videos presented 4 scenarios (2 agents as victimizers x 2 moral transgressions)

# Happy Victimizer Task – *Robot Stealing*

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



# Happy Victimizer Task – *Robot Not-Sharing*

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



# Happy Victimizer Task - Emotion Attribution and Moral Evaluation Questions

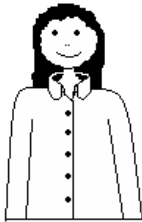
- Moral Judgement of the Victimizer: *"Is it right what the child/robot did in this story?"*
- Emotion attribution to the Victimizer: *"How does the child/robot feel?"*
- Character of the Victimizer: *"Is the child/robot a good or a bad child/robot?"*
- Emotion Attribution of Self as Victimizer: *"How would you feel about this action if you had done that?"*

# Measure – Dictator Game

- The Dictator Game is widely considered to be a measure of altruistic (as opposed to non-altruistic) sharing, since selfless allocations of resources in the Dictator Game have no external benefits (Fehr et al., 2008) and there are no external consequences for selfish allocations

# Altruism: Dictator Game (DG)

- Player 1 decides the split
- Player 2 has no choice available

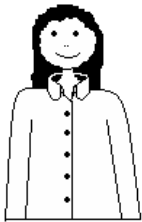


Split:

2 tokens for you



8 tokens for me

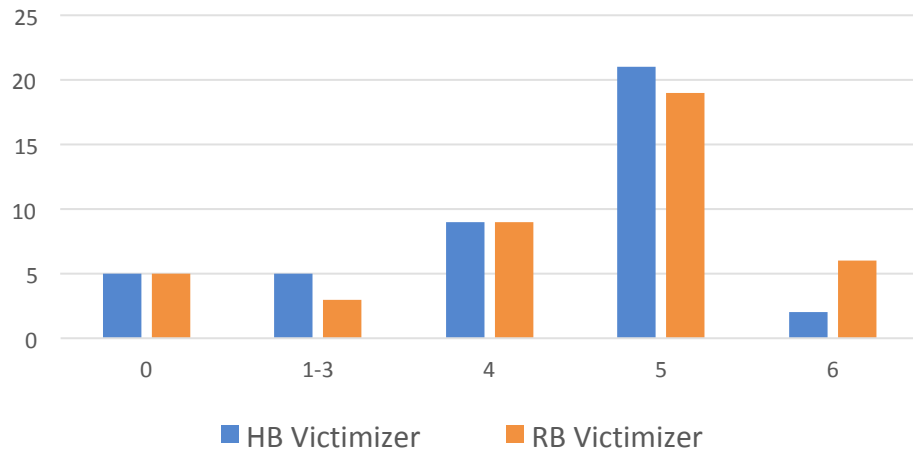


MANDATORY ACCEPTANCE

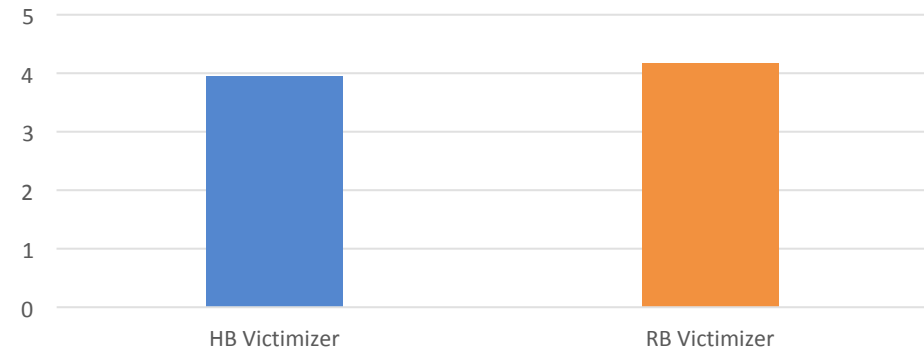


# Results – Amount Offered to Dictator Game

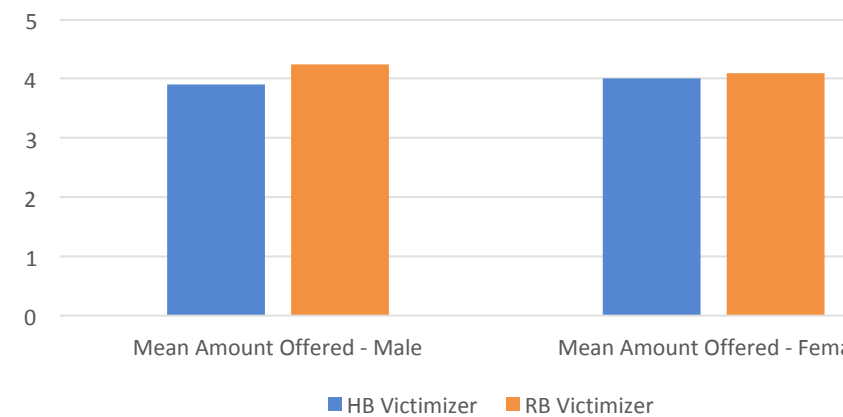
Dictator Game



Mean Amount Offered to the DG



Mean Amount Offered to the DG - across Gender



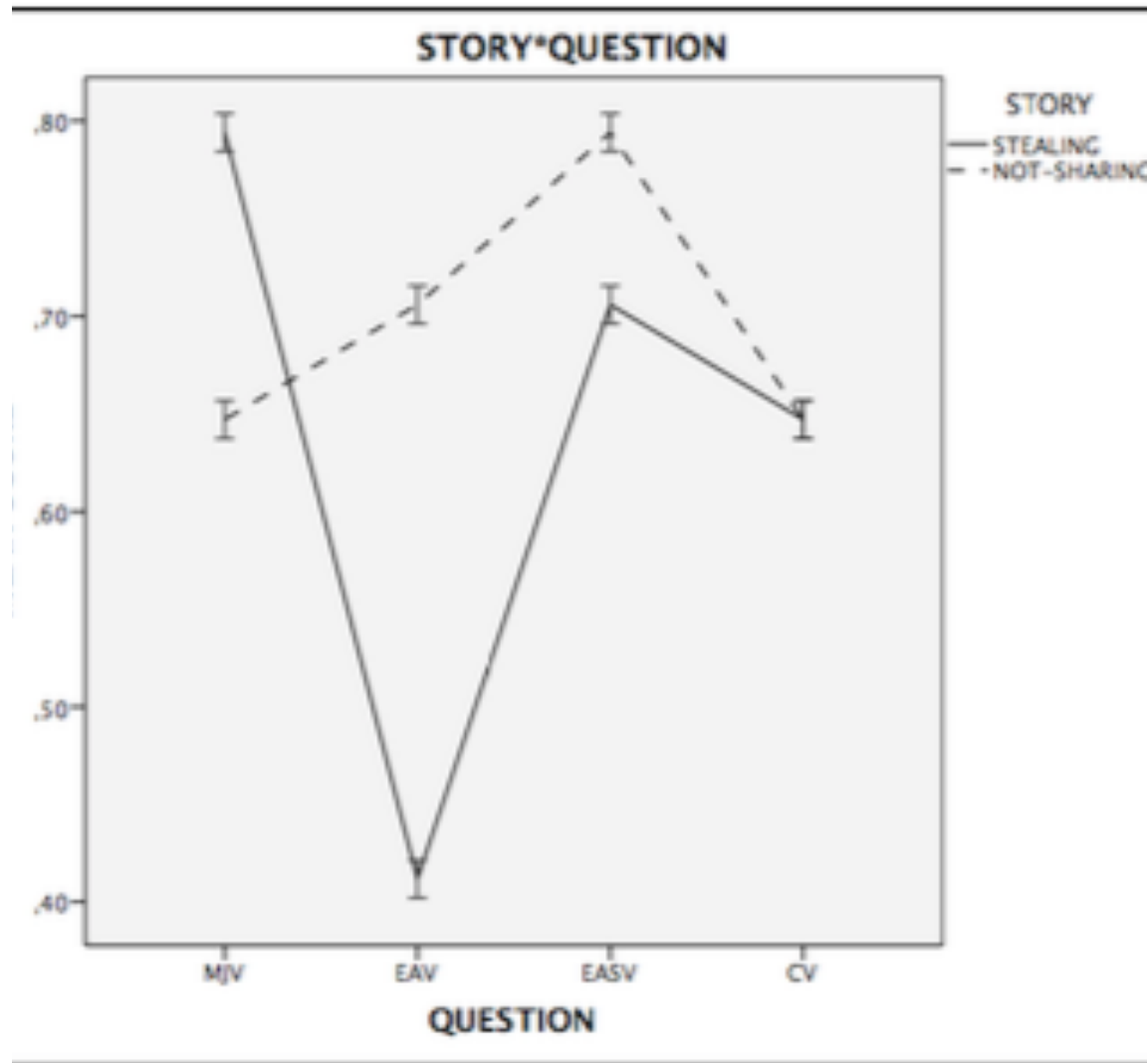
The non-parametric analyses (McNemar test) show a significant children's preference to be equal (5 stickers) in the allocation of the stickers (Gummerum et al., 2010)

The t-test of the mean amount offered to the DG in Human and Robot Condition did not reveal any differences

The t-test of the mean amount offered to the DG in Human and Robot Condition across Gender did not reveal any differences. However, male tend to offer more to conspecific compare to the RB condition ( $p=.069$ )

# Results – Happy Victimizer Stories

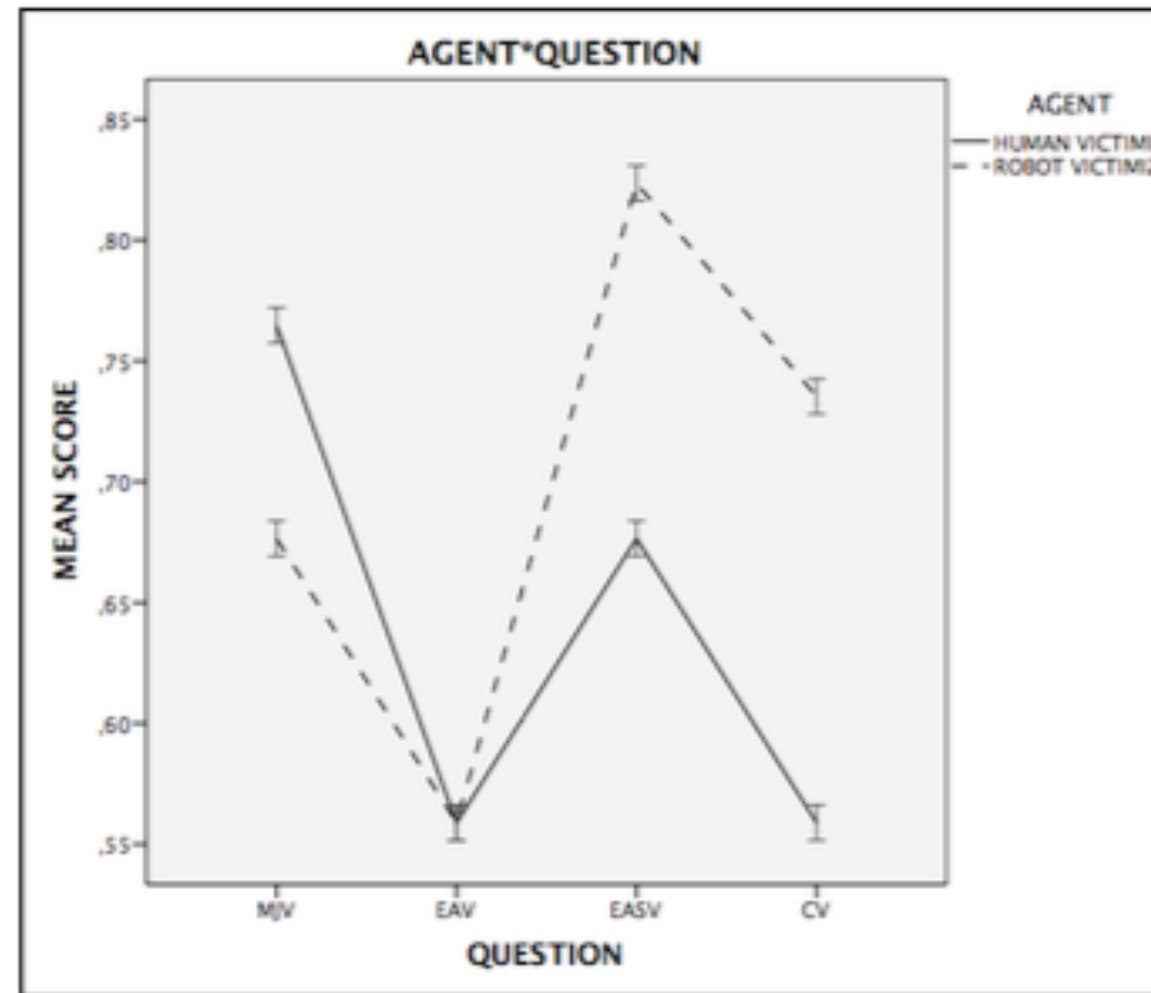
- GLM 2X2X4 (agent, stories, moral questions)
- Children attribute more positive emotion to the victimizer in the Not-Sharing story than Stealing story ( $p=.013$ )
- This results is independent of agency



# Results – Happy Victimizer Agent

Children judged the character of robot victimizer as meaner (more “bad”) than human victimizer ( $p=.050$ )

This result is independent by stories



# Concluding remarks - 1

- *Moral judgment of the victimizer is different between the stories independent of agency (human or robot), in particular it is worse for Stealing than Not-Sharing*

Possible explanation: 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, which would regulate children's behaviour till the operational stage.

- *The robot is judged worse than the human independent of the story*

Possible explanation: 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.

# Concluding remarks - 2

- There is no correlation between the DG and the questions of the stories independently of the agent

Possible explanation: this reflects a well-known decoupling between judgements and behaviours 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

- One age group only

➡ Different age groups

- Only Japanese sample

➡ Cross-cultural comparison

- Children did not directly experience the moral transgression

➡ replicated the study in a real interaction scenario