Utrecht University

Research Seminar in Social Neuroscience

Research Proposal Motivation

Laura Verhoeven (5605105)

Laurie Spapens (6388116)

Esmée Teinsma (6506461)

Does Socializing Between Humans and Robots have a Different Outcome in the Prisoner's Dilemma?

Introduction

In this research proposal, the cooperation between humans and robots in a prisoner’s dilemma game will be discussed. We will be elaborating on previous research that will be mentioned in this proposal by researching the effect of human-robot interaction prior to the prisoner’s dilemma on human-robot cooperation. Through different studies we have come to the following research question: Does socializing between humans and robots have a different outcome in the prisoner's dilemma?

The prisoner’s dilemma

In a study of Sandoval and colleagues (2015), they used a (repeated) prisoner’s dilemma in order to see whether people would reciprocate with a robot as well as with a human agent. The prisoner’s dilemma is an economic game of cooperating and defecting. It depends on whether the one ‘prisoner’ can predict whether the other would cooperate (i.e. remain silent) or defect (i.e. talk) for cooperation to be the best strategy (Sandoval et al., 2015). The agents used one of two strategies, either a Tit for Tat strategy or a random strategy. In a Tit for Tat strategy, the agent chooses a strategy relative to the strategy the participant chose in the game before.

The results of this study showed that the participants collaborated more with human agents than with robot agents. They were however equally reciprocal with both agents. This means that if the robot agent cooperated, the participant would (in the next game) be more likely to cooperate as well.

FRN

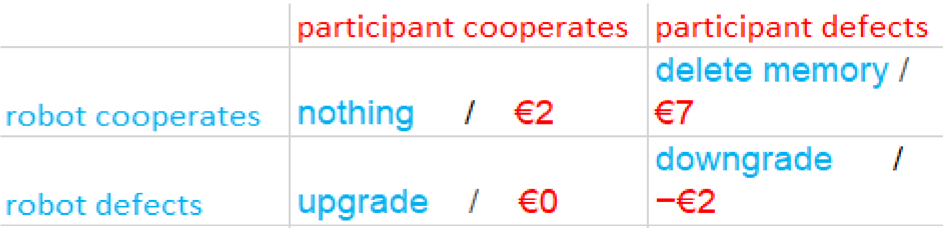
In a study performed by Cervantes Constantino et al. (2020) the neural processes for cooperation between individuals was measured during the iterated prisoner's dilemma. The feedback stage of the prisoner’s dilemma was measured with electroencephalography (EEG). The results showed that when the individuals received unreciprocated cooperation, the decision thereafter relates to changes to the feedback related negativity (FRN). The FRN occurs when external feedback is received, indicating that performance is worse than expected (Crowley, 2013). The FRN is a deflection elicited 200 to 350 ms after feedback.

Study design

In this study, 80 participants will be divided into 4 groups of 20 participants each. Each participant will be in only one of the four groups. These four groups differ on the time spent interacting with the robot. The first group interacts with the robot for a total of 10 hours prior to the study, the second group interacts for 5 hours, the third group interacts with the robot for 30 minutes right before the study, and the last group does not interact with the robot at all. This ten-hour socialisation time was inspired by similar research done by Tanaka et al. (2007). We added the five-hour and thirty-minute socialisation times to see if there would be a significant difference in cooperation between the three.

In each group, there will be 10 repetitions of the prisoner's dilemma task. The participants will not know how many rounds they will play against the robot. This means that their decisions will be conditioned by the possibility of interacting with the agent in an undetermined number of rounds (Sandoval et al., 2015): people tend to be more reciprocal and collaborative due to the reputation of the opponent in the previous round when they do not know the number of rounds.

The participants in this study will hear a script beforehand. This script will briefly explain the basics of the prisoner's dilemma game, and it will introduce the robot that the participant will be playing the game with. In this script, the participant will be told that if they lose they cannot keep the money they earned, and the robot will receive an upgrade (Abubshait et al., 2020). The participant can keep the money they earned if they win from the robot. In this case, the participant will be told that the robot's memory will be deleted. The matrix for this is represented in Figure 1. This memory deletion 'punishment' is used because research by Seo et al. (2015) has shown that such a prompt is useful in eliciting people's real concerns and empathy towards a robot.



	participant cooperates	participant defects
robot cooperates	nothing / €2	delete memory / €7
robot defects	upgrade / €0	downgrade / −€2

Figure 1. The cooperate/defect matrix.

Conclusion

In the article of Abubshait et al. (2020), they found that in the Iowa gambling task that Reward Positivity (RewP) amplitudes were enhanced for participants who interacted with the robot in comparison to subjects who did not. Based on this we would expect the FRN to show more deflection for the situation in which the human agents have socialized more with robots. The study of Cervantes Constantino et al. (2020) showed that the decision for the next round was based on the FRN. This would mean that when the FRN shows a greater deflection the chance of defecting for the human increases.

Because the ERP's will show greater responses when there is more interaction between humans and robots, we would expect that they would respond more like they did in the human-human prisoner's dilemma and will increase cooperation.

We would expect that in the case that the robot does not cooperate for the 10-h condition the FRN will show the greatest deflection. Because the group with no socializing has not interacted with the robot before we would expect the feedback related negativity to have a lower deflection for this group when there is unreciprocated cooperation.

We also expect that human agents will become more cooperative when they have interacted for more hours. The group that interacts with the robot with a total of 10 hours, will have the most cooperation with the robot during the prisoner's dilemma. The group with no interaction will show the least cooperation.

This outcome would show that socializing between humans and robots would lead to more cooperation between humans and robots for the prisoner's dilemma. This means that interaction time is connected with the cooperation between humans and robots. Although we believe that after 10 h of socializing the cooperation will not increase further between humans and robots, because of the 10-h barrier*. Only after 5 months of interaction between humans and robots the cooperation could go up again (Tanaka et al. 2007).

*The 10-h barrier was one of the concepts that emerged from the discussions at the National Science Foundation's Animated Interfaces and Virtual Humans Workshop in Del Mar, California in April 2004.

References

Abubshait, A., Beatty, P., McDonald, C., Hassall, C. D., Krigolson, O., & Wiese, E. (2020, May 14). A win-win situation: Does familiarity with a social robot modulate feedback monitoring and learning?. https://doi.org/10.31234/osf.io/6z75t

Cervantes Constantino, F., Garat, S., Nicolaisen-Sobesky, E., Paz, V., Martínez-Montes, E., Kessel, D., … Gradin, V. B. (2020). Neural processing of iterated prisoner's dilemma outcomes indicates next-round choice and speed to reciprocate cooperation. *Social Neuroscience*, 1–18. https://doi.org/10.1080/17470919.2020.1859410

Crowley M.J. (2013) Feedback-Related Negativity. In: Volkmar F.R. (eds) Encyclopedia of Autism Spectrum Disorders. Springer, New York, NY. https://doi-org.proxy.library.uu.nl/10.1007/978-1-4419-1698-3_731

Hsieh, T. Y., Chaudhury, B., & Cross, E. S. (2020, March). Human-Robot Cooperation in Prisoner Dilemma Games: People Behave More Reciprocally than Prosocially Toward Robots. In Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (pp. 257-259).

Sandoval, E. B., Brandstetter, J., Obaid, M., & Bartneck, C. (2016). Reciprocity in human-robot interaction: a quantitative approach through the prisoner's dilemma and the ultimatum game. International Journal of Social Robotics, 8(2), 303-317.

Seo, S. H., Geiskkovitch, D., Nakane, M., King, C., & Young, J. E. (2015, March). Poor thing! Would you feel sorry for a simulated robot? A comparison of empathy toward a physical and a simulated robot. In 2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 125-132). IEEE.

Tanaka, F., Cicourel, A., & Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. Proceedings of the National Academy of Sciences, 104(46), 17954-17958.