The impact of game outcomes and Agent-based Feedback on Prosociality in Social Dilemmas

Jennifer Renoux
Center for Applied Autonomous Sensor Systems, Örebro University, Sweden

Neziha Akalin
Center for Applied Autonomous Sensor Systems, Örebro University, Sweden

Joana Campos
INESC-ID and Instituto Superior Técnico, University of Lisbon, Portugal

Filipa Correia
Instituto Superior Técnico, University of Lisbon, and ITI/LARSYS, Portugal

Lucas Morillo-Mendez
Center for Applied Autonomous Sensor Systems, Örebro University, Sweden

Fernando P. Santos
Informatics Institute, University of Amsterdam, The Netherland

Ana Paiva
INESC-ID and Instituto Superior Técnico, University of Lisbon, Portugal

ABSTRACT

We implemented the Pest Control Game, simulating a collective risk dilemma. We evaluated how observing game outcomes and receiving feedback from an artificial agent impact prosociality. We found that events, even those affecting other players, positively impacted the participants' prosociality. We did not find an impact of the agent's feedback.

This work has been conducted as part of the HumaneAI-Net micro-project calles "Social Dilemma with Information Asymmetry (SDIA)"

Figure 1: Screenshot of a tutorial round of the Pest Control Game. During the tutorial, the participants learn the rules of the game and how to interact with it.

1 INTRODUCTION

From vaccination to climate change, averting catastrophic events often requires individuals to cooperate for the collective good. This can prove challenging as altruistic cooperation entails a social dilemma: it involves a costly contribution, which can tempt individuals to free-ride and rely on the efforts of others [4]. The dynamics of human behaviour when facing these dilemmas have been extensively studied both experimentally [2, 9] and through mathematical models and computer simulations [1, 8]. The dilemma associated with public goods cooperation changes when risks are involved [8] or when there are asymmetries between individuals [5]. Social robots and virtual agents have the potential to serve as prosocial behaviour elicitors [6]. In this project, we investigated human prosocial behaviour when playing a game called *The Pest Control* [7] in a setting where an artificial agent provides feedback and players can directly observe the state of other individuals and the environment.

2 EXPERIMENTAL PROTOCOL

In the Pest Control Game (Figure 1), a set of farmers is placed in a 2-dimensional hexagonal grid. The pest starts in a cell and can propagate to adjacent cells. When the pest reaches a cell occupied by a farmer, that player loses the game and all accumulated resources. The game is played over multiple rounds. At each round, farmers receive an endowment and decide how much to invest to prevent the pest from spreading. The pest is controlled with a probability proportional to the total investment by all farmers. By the end of the game, surviving players are rewarded based on what they accumulated. Here lies the dilemma: investing in pest control averts collective losses, although players maximise their payoffs if they do not invest and remain safe through the contributions of others. We implemented an artificial agent, named *Pat the Bot*, which gives information to the participants through the game. The information

HumanE AI Conference, 2022, Stockholm, Sweden

https://doi.org/

given by Pat the Bot was of two categories, following the work in [3]. The *Problem Awareness* (PA) category contains utterances to help the player understand the social dilemma, while the *Player Strategies* (PS) category contains utterances that relates to players actions. We tested two hypotheses:

H1: The feedback of the artificial non-playing agent positively affects the participants' prosocial actions.

H2: Personal characteristics such as personality traits and gender influence participants' prosocial actions in the game.

We conducted a user study with 177 participants, separated in four conditions:

(1) **Control Condition (34 participants):** The participant did not receive any information from *Pat the Bot* and were not even made aware of its existence.
(2) **PA Condition (54 participants):** The participant only received information of the PA type, and received two utterances for each PA concept in a random order.
(3) **PS Condition (34 participants):** The participant only received information of the PS type, and received two utterances for each PS concept in a random order.
(4) **PA+PS Condition (55 participants):** The participant received information from both PA and PS types, and received one utterance for each PA and PS concepts in a random order.

The participant first played a short tutorial game followed by the study game. Finally, they answered a questionnaire assessing various personal characteristics. Upon successful completion of the study, the participants received a fixed monetary compensation for participating, plus a bonus proportional to the number of coins they had at the end of the study game.

3 RESULTS AND CONCLUSION

We did not find support for H1. However, when analysing the difference between the contributions in consecutive rounds (Figure 2) , we found that threatening game events have a significant impact on the prosocial behaviours of individuals. This result suggests that feedback by an artificial agent could be more effective when provided based on the situation, which was not the case in our study. Regarding H2, we found that female participants has significantly less coins in their wallet at the end of the game than male participants, indicating that they contributed more throughout the game. We did not find any other significant correlation between other personal characteristics and final wallet.

We conclude that communication types that have been found to work when used by players in a public good game ([3]) did not impact significantly the players when used by a non-playing character (NPC). Several reasons may be invoked to explain this difference. First, the fact that the agent is an NPC renders impossible to add an "agreement" stage [3] which strengthens the impact of the communication strategies. The feedback from an NPC might require another combination of strategies to be efficient. Second, the fact that our agent was external and neutral, might have driven players to ignore its feedback and consider it as a "disturbance".

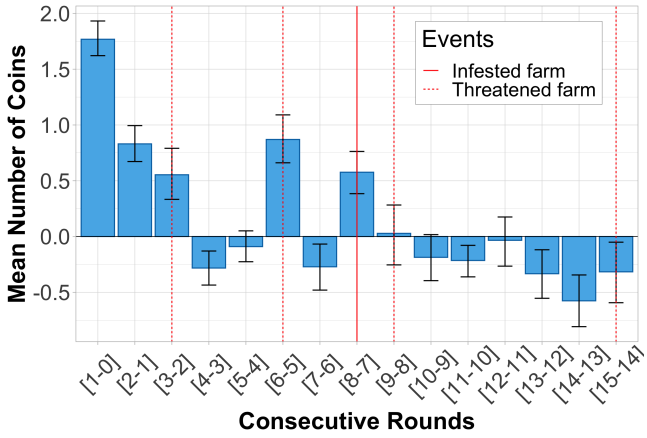

Figure 2: Mean contribution difference between consecutive rounds. Error bars represent 95% bootstrapped CIs. Mean difference at [1-0] is equal to mean contribution at round one.

Given the impact of the game outcomes on player's contribution, we hypothesise from this study that focusing the agent's feedback on actual game events might reinforce the its impact. Future work will evaluate this hypothesis. In addition, we consider making the agent team with the participant.

REFERENCES

[1] Wolfram Barfuss, Jonathan F Donges, Vítor V Vasconcelos, Jürgen Kurths, and Simon A Levin. 2020. Caring for the future can turn tragedy into comedy for long-term collective action under risk of collapse. *Proceedings of the National Academy of Sciences* 117, 23 (2020), 12915–12922.

[2] Elias Fernández Domingos, Jelena Grujić, Juan C Burguillo, Georg Kirchsteiger, Francisco C Santos, and Tom Lenaerts. 2020. Timing uncertainty in collective risk dilemmas encourages group reciprocation and polarization. *Iscience* 23, 12 (2020), 101752.

[3] Ann-Kathrin Koessler, Juan Felipe Ortiz-Riomalo, Mathias Janke, and Stefanie Engel. 2021. Structuring Communication Effectively—The Causal Effects of Communication Elements on Cooperation in Social Dilemmas. *Environmental and Resource Economics* 79, 4 (2021), 683–712.

[4] Simon A Levin. 2014. Public goods in relation to competition, cooperation, and spite. *Proceedings of the National Academy of Sciences* 111, Supplement 3 (2014), 10838–10845.

[5] Ramona Merhej, Fernando P Santos, Francisco S Melo, and Francisco C Santos. 2021. Cooperation between independent reinforcement learners under wealth inequality and collective risks. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*. 898–906.

[6] Raquel Oliveira, Patrícia Arriaga, Fernando P Santos, Samuel Mascarenhas, and Ana Paiva. 2021. Towards prosocial design: A scoping review of the use of robots and virtual agents to trigger prosocial behaviour. *Computers in Human Behavior* 114 (2021), 106547.

[7] T. Reeves, H. Ohtsuki, and S. Fukui. 2017. Asymmetric public goods game cooperation through pest control. *Journal of theoretical biology* 435 (2017), 238–247.

[8] Francisco C Santos and Jorge M Pacheco. 2011. Risk of collective failure provides an escape from the tragedy of the commons. *Proceedings of the National Academy of Sciences* 108, 26 (2011), 10421–10425.

[9] Alessandro Tavoni, Astrid Dannenberg, Giorgos Kallis, and Andreas Löschel. 2011. Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proceedings of the National Academy of Sciences* 108, 29 (2011), 11825–11829.