COORDINATING HUMAN AND AGENT BEHAVIOR IN COLLECTIVE-RISK SCENARIOS

Elias Fernández Domingos[1,2,3], Juan Carlos Burguillo[3], Ann Nowé[2] and Tom Lenaerts[1,2]

[1] AI lab, Computer Science Department, Vrije Universiteit Brussel, Belgium
[2] MLG, Université Libre de Bruxelles, Belgium
[3] Department of Telematic Engineering. University of Vigo, Spain

MOTIVATION OF THE WORK

Many real-world scenarios are characterized by risk and uncertainty under time constraints. These constraints also create a dilemma that poses over the trade-off between immediate and future rewards. The importance of understanding and explaining human behavior under this scenarios range from being able to predict and anticipate the effects of sharing economies, to developing policies that can increase international awareness and cooperation towards the issue of the climate change. **These situations have been operationalized in Game Theory as the Collective-Risk Game (CRD) [1]**, where we place our study.

Ultimately, our goal is to apply the scientific and technological framework that we develop to the real-world scenarios already described and, in concrete, to P2P energy markets that form a micro-grid and 5G networks, where **humans have to interact with artificial agents** and **agents have to negotiate among themselves** taking into account human's preferences and utilities [2].

THESIS OBJECTIVES

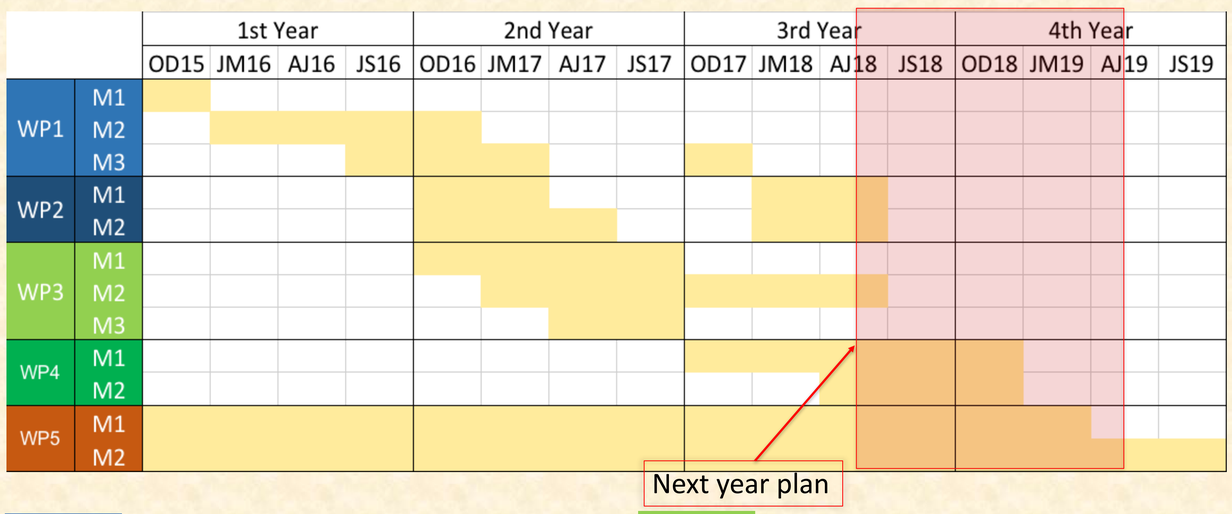
1 **Characterize the macro-behavior observed in the CRD** and provide policies to drive the system towards certain stable points. (WP2, WP4)

2 Create **models of individual human behavior** in the Collective-Risk Dilemma. We want to test our hypothesis **that anticipatory behavior** is a key element in human decision-making. (WP2, WP3)

3 **Validate the models** through behavioral experiments and data analysis. We will perform experiments with humans on the CRD and variants that we have designed, as well as hybrid experiments between humans and artificial agents. (WP1)

4 **Apply our models** in scenarios of risk an uncertainty that are similar to the CRD, e.g., energy smart-grids or 5G networks, where **agents need to be able to recognize and signal intentions** and effectively negotiate with other agents and humans. (WP5)

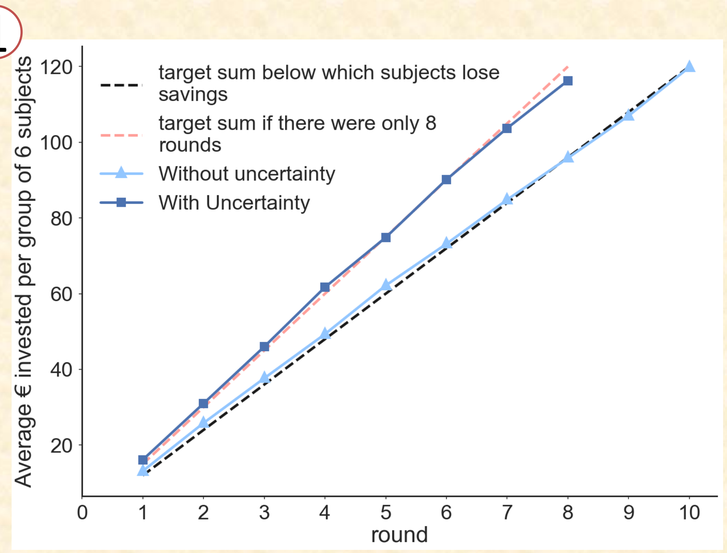
RESEARCH PLAN & NEXT YEAR PLANNING



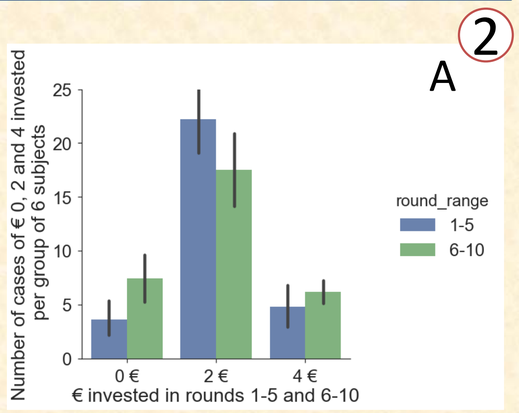
WP1-M1: Design of the experiments.
WP1-M2: Design and implement a specific framework required to perform the experiments.
WP1-M3: Perform the experiments and collect the data.
WP2-M1: Search for behavioral models in the data.
WP2-M2: Evaluate whether the inferred models lead to the observed macroscopic behavior.
WP3-M1: Experiment with previous approaches from the literature.
WP3-M2: Create new models to explain the data.
WP3-M3: Compare the behavior of different models.
WP4-M1: Multi-agent simulations.
WP4-M2: Search mechanisms and policies to influence behavior.
WP5-M1: Identify applications of our research to real-world problems.
WP5-M2: Write the PhD thesis.

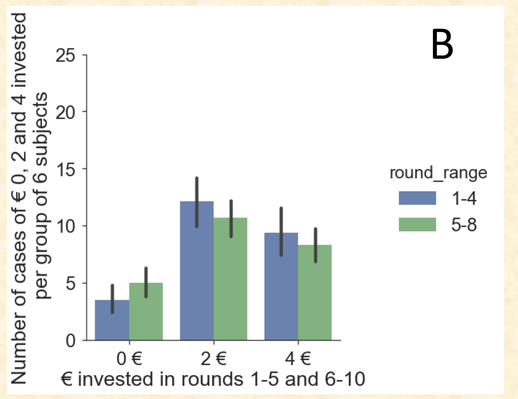
HOW DO HUMANS BEHAVE IN THE PRESENCE OF COLLECTIVE-RISK AND UNCERTAINTY?

1

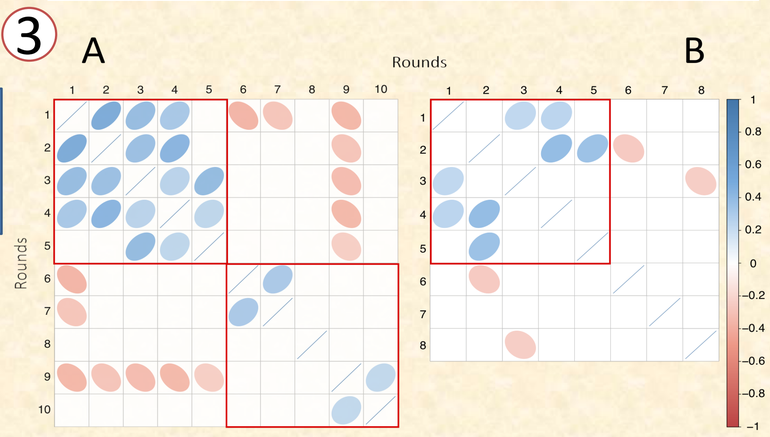


2





3

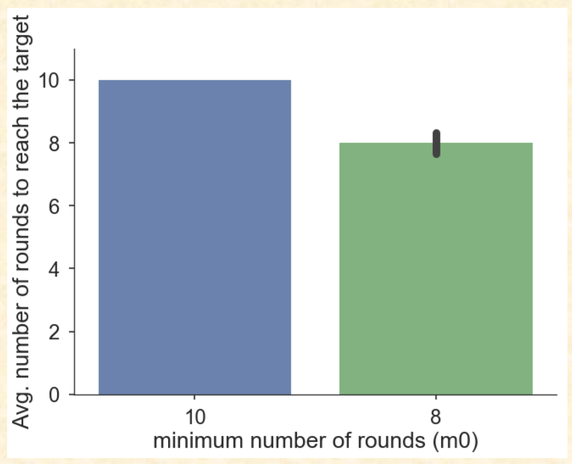


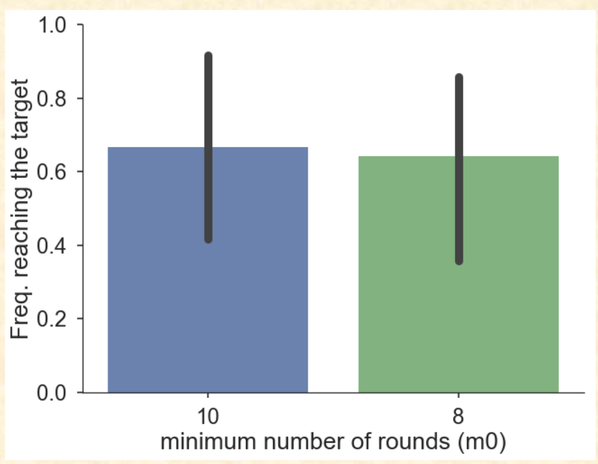
A **Without uncertainty**

B **With uncertainty**

We have performed behavioral experiments with 156 participants over 2 different treatments on the Collective-Risk game to study how the presence of risk and uncertainty affects collective and individual behavior. (1) We find that **under the presence of high risk, groups of 6 participants manage to reach on average the collective target.** Surprisingly, **when there was uncertainty about when the game would end**, **the groups substantially increased their donations**. (2) Participants also increased the "extreme" actions (0 and 4) by the end of the game. (3) This shift in behavior seems to be clear when we make an autocorrelation ($p < 0.05$) between the actions of the players at each time-step.

UNERTAINTY ABOUT THE END OF THE GAME MAKES PARTICIPANTS REACH THE TARGET EARLIER





Generally, it is considered that uncertainty diminishes cooperation, yet, in our experiments, not only we do not observe this, but also **participants reach the target earlier**. Therefore, we hypothesize that different types of uncertainty might have a different impact in multi-agent scenarios under collective risk and **the magnitude of uncertainty about the end of the game might be a key feature to drive cooperation**. We will perform more experiments to validate this hypothesis, but in the mean time we are expanding our analysis through agent-based simulations, where we also explore the dynamics of the models we presented in [2, 3]. **We are preparing a journal paper with these results**.

[1] Milinski, M.; Sommerfeld, R. D.; Krambeck, H.-J.; Reed, F. a.; and Marotzke, J. 2008. The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proceedings of the National Academy of Sciences of the United States of America* 105(7):2291–2294.
[2] **Fernández Domingos, E.; Burguillo, J.C.; Nowé, A.; and Lenaerts, T.** 2017. Coordinating Human and Agent Behavior in Collective-Risk Scenarios. In *Proceedings of the thirty-first AAAI conference*, AAAI.
[3] **Fernández Domingos, E.; Burguillo, J. C.; and Lenaerts, T.** 2017. Reactive versus anticipative decision making in a novel gift-giving game. In *Proceedings of the thirty-first AAAI conference*, AAAI.

VUB ULB UNIVERSIDADE DE VIGO fwo Research Foundation Flanders Opening new horizons

Check here for more information about me and the project