Prospect Theory EcopG

July 30, 2025

Contents

1 Introduction

Climate change is a pressing global challenge that demands a deeper understanding of the drivers of cooperation. Most existing models rely on the assumptions of agents behaving as rational individuals guided by expected utility theory. However, to accurately reflect real-world decision making more, it is important to incorporate empirically validated frameworks such as Prospect Theory (hereafter, PT). Doing so would enable us to have a refined understanding of the drivers of cooperation under risk of ecological collapse. This would also contribute to the advancement of MARL frameworks that account for deviations from rational decision making.

The four features of prospect theory along with its potential applications to collective action is outlined under Research Objectives

1.1 Problem Statement

Developing behaviorally grounded models of collective action under climate risk to explain the gap between optimal and actual mitigation efforts.

1.2 Research Objectives

This project aims to extend multi-agent learning dynamics by incorporating PT-based preferences into agent decision-making. The resulting framework will enable the analysis of how systematic behavioral biases influence cooperation in the face of climate-related risks. Specifically, we can investigate the following features of prospect theory under climate risk:

(i) **Loss aversion**: Individuals respond asymmetrically to gains and losses. This asymmetry, between the perceived short-term gains from defection and the long-term losses from gradual or catastrophic climate change, can be examined to understand how it influences the emergence of cooperation.

(ii) **Reference dependence**: In climate policy discourse, the choice of reference point strongly shapes how contributions are perceived [5]. When the reference is set at the level of contribution needed to meet climate goals, costs are seen as necessary investments to avert future risks. In contrast, if the status quo is the reference point, any contribution is perceived as a loss and thus an economic burden. Although the material outcomes may be the same, these framings evoke different psychological responses. The role of reference points in shaping decision-making under climate risk can be systematically examined within this framework.

(iii) **Diminishing sensitivity**: PT suggests that people are more sensitive to changes near the reference point than to those further away. This feature can be examined to assess how marginal changes in the probability of collapse elicit varying degrees of willingness to cooperate, depending on their distance from the reference point

(iv) **Probability distortion**: Individuals tend to misperceive probabilities, over weighting low-probability events and under weighting high-probability ones. This distortion can be examined to understand how perceptions of the uncertain risk of catastrophic climate change affect mitigation efforts.

2 Literature Overview

Several models of collective action under climate risk have been proposed [3, 2]. Many of these models rely on the assumption of rational preferences. Evidence from behavioral experiments shows that framing public goods games in terms of losses, rather than gains, can lead to different outcome, aligned with PT indicating the need for behaviourally grounded models. [6]. Some efforts have been made to incorporate PT into climate action, such as incorporating loss aversion into coalition formation models of international environmental agreements [8, 7].

However, the *dynamics* of collective behavior when socio-ecological feedback is explicitly considered remains an important area of exploration

3 Method

Similar to [1], we plan on developing analogous deterministic limits for reinforcement learning models in stochastic games, that incorporate prospect theory–based preferences (e.g., cumulative prospect theory based Q-learning [4]). Further, we aim to review derivations of equilibrium solutions of games when agents have PT based preferences.

We will simulate ecological public goods games (EcoPG) involving prospect theory–based agents and evaluate the conditions under which cooperation emerges. Further, we will examine how the likelihood of cooperation changes with parameters of risk aversion, probability distortion, and changes in reference points. We will further compare these outcomes with predictions from models involving rational agents to identify the key consequences of incorporating these behaviours.

References

[1] Wolfram Barfuss, Jonathan F. Donges, and Jürgen Kurths. Deterministic limit of temporal-difference reinforcement learning for stochastic games. *Physical Review E*, 99(4):043305, 2019.

[2] Wolfram Barfuss, Jonathan F. Donges, V. V. Vasconcelos, Jürgen Kurths, and Simon A. Levin. 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):201916545, 2020.

[3] Scott Barrett and Astrid Dannenberg. Climate negotiations under scientific uncertainty. *Proceedings of the National Academy of Sciences*, 109(43):17372–17376, 2012.

[4] Dominic Danis, Parker Parmacek, David Dunajsky, and Bhaskar Ramasubramanian. Multi-agent reinforcement learning with prospect theory, 2023.

[5] Elisabeth Gsottbauer and C. J.M̃. van den Bergh, J.˙ Bounded rationality and social interaction in negotiating a climate agreement. *International Environmental Agreements: Politics, Law and Economics*, 13(3):225–249, 2013.

[6] Iñigo Iturbe-Ormaetxe, Giovanni Ponti, Josefa Tomás, and Luis Ubeda. Framing effects in public goods: Prospect theory and experimental evidence. *Games and Economic Behavior*, 72(2):439–447, 2011.

[7] Doruk İriş. Economic targets and loss-aversion in international environmental cooperation. *Journal of Economic Surveys*, 30(3):624–648, 2016.

[8] Doruk İriş and Alessandro Tavoni. Loss aversion in international environmental agreements. *Environmental and Resource Economics Review*, 27(2):363–397, 2018.