The Role of Information in Modulating Cooperation Under the Risk of Ecological Collapse

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Akilan R

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

December, 2025

Supervisor: Dr Wolfram Barfuss

Certificate

This is to certify that this dissertation entitled 'The Role of Information in Promoting Cooperation Under the Risk of Ecological Collapse' towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Akilan R at the Center for Development Research (ZEF), University of Bonn under the supervision of Dr Wolfram Barfuss, Junior Professor, during the academic year 2024-2025.

Dr Wolfram Barfuss

Committee:

Dr Wolfram Barfuss

Dr Joy Merwin Monteiro

This thesis is dedicated to my grandmother, Mrs Sarada DS, for her selfless care throughout my growing years, which laid the foundation of everything I have achieved.

Declaration

I hereby declare that the matter embodied in the report entitled 'The Role of Information in Promoting Cooperation Under the Risk of Ecological Collapse', are the results of the work carried out by me at the Center for Development Research (ZEF), University of Bonnunder the supervision of Dr Wolfram Barfuss and the same has not been submitted elsewhere for any other degree.

Akilan.

Akilan R

Acknowledgments

I am grateful to all those who have supported and influenced me; the following is by no means comprehensive, but an attempt to acknowledge a few of them.

I would first like to thank Wolf, whom I've thoroughly enjoyed working with. I am very fortunate to have an exceptionally supportive supervisor. I've had a lot of fun working on this project; his encouragement motivated me to engage with it deeply. His guidance and feedback have always been precise and insightful, and I feel it has improved my abilities to do better science. I appreciate his enthusiasm, accommodating nature and support, which made this an excellent thesis experience. I would also like to extend my thanks to the others in the BarfussLab - Jack, Balu, Ali, Paul, Shu and Clemence for their suggestions that greatly improved the project and coffee chats that made my time more enjoyable.

Beyond the thesis, I've had the fortune of learn from great mentors in the past, Tom Keaney, Hanna Kokko, Akshay Malwade, Kasturi Saha, Milind Watve and GN, each of whom played a role in nurturing my abilities to do good science. I owe special thanks to Sutirth Dey, who has greatly influenced my perspectives about science, academics and life. His advice has guided me on more occasions than I can count. I would also like to thank Joy Monteiro, the expert for this thesis, for kindly supporting my diverse interests by giving me the opportunity to work on multiple projects under his guidance, which allowed me to develop the right foundation for my research. I'm grateful for the education at IISER Pune, whose interdisciplinary approach and supportive faculty and administration let me follow my interests and build the right skills along the way.

Looking back, it feels like a privilege to work on an original research question in game theory, a topic I first discovered through The Selfish Gene back in high school. I'm thankful to the organisers of the game theory workshops that helped me explore this field further. And to my aunt Dr. Bharti, who nurtured my love for science early on. I extend my thanks to all the brilliant people whose work in and around the field has inspired me. It amazes me how much of science today is built on the generosity of people who freely share their knowledge and creations online - through open-source software, Wikipedia, preprint archives, and countless other community efforts

Above all, it is my friends whom I will miss the most. They have been the ones providing encouragement, companionship, and going above and beyond to support me during some of the toughest times. Nakul has been my go-to person to talk about almost anything, from scientific ideas to personal problems to random light hearted banter. Epsita is an exceptionally smart and compassionate person, whose company I greatly enjoy. Saransh is a person whose dedication reminds me to keep striving for the things that really matter. And I would like to thank Vihang, for always being there when I needed him, especially when my mental health was not at its best. I would also like to thank Vivek, Samarth, Iyer, Sumant(h), Avanthika, Anish, Ayush, Shashwati, Keshav and Vinayak for their friendship. And so many others -especially the ones I spent time with in room 116, room 604, and the second-floor corridor. Thank you for the warmth, laughter, and memories.

I am deeply grateful to my parents for their constant presence, love, and care. It is their support that created the perfect environment for me to thrive. I also extend my thanks to my brother, relatives in Germany, colleagues at ZEF, Kokkonuts and IISER, and dormmates in Augustinushaus and anyone I've unintentionally overlooked who've helped in numerous ways, big and small over the years.

Financial support from ZEF and KVPY-INSPIRE is also acknowledged

Abstract

Crossing critical climate tipping points can trigger large-scale ecological collapse with severe consequences. Mitigating such an outcome requires coordinated efforts from global stakeholders, such as countries, major corporations, and cities. Yet achieving such cooperation is hindered by short-sighted incentives. Moreover, individual willingness to contribute is complicated by uncertainty about others' responses and limited awareness of ecosystem dynamics. Despite a growing body of research on this global collective action problem, the role of information in shaping outcomes remains poorly understood. In this thesis, I implement a game-theoretic framework to systematically evaluate how access to social and ecological cues shapes long-term cooperation under the threat of ecological collapse. I find that ecological information is essential for sustaining cooperation, whereas social information alone is largely ineffective in the settings examined. However, adding social transparency alongside ecological awareness can be double-edged: it can either promote or hinder cooperation, depending on the context. This study advances the theoretical understanding of how various forms of information influence cooperative behaviour under climate risk.

Contents

Chapter 1

Introduction

1.1 Sustainability as a Collective-Action Problem

The Earth system is approaching critical thresholds due to human activities that destabilise ecological and climatic processes [2]. Crossing certain Earth system thresholds could initiate irreversible changes with impacts that unfold over millennia. These consequences include intensified heatwaves, rising sea levels, large-scale biodiversity loss, disruption of food and water systems, thus heightening risks to livelihoods and life across the globe [16] . The challenge is immediate: there needs to be stewardship policies that must be established in the coming years to steer the Earth system away from such destabilising pathways and instead retain the Earth within a safe and habitable state, where societies can sustainably balance ecological balance for long-term well-being [31, 35].

A robust body of scientific evidence demonstrates that reducing carbon emissions and adopting sustainable practices can effectively mitigate these risks [15]. However, the benefits of coordinated efforts, such as a stable climate and resilient ecosystems, are shared globally and manifest only in the long term, whereas the costs of clean energy investments or restrictions on resource exploitation are immediate and borne by individual actors. From a self-interested economic standpoint, actors could free-ride, making others bear the costs of emission reductions while still benefiting from the resulting improvements in environmental quality [8].This situation is known as a social dilemma [11], as what appears as rational to the individual is at odds with what is good for the collective. Appropriate action, such as investing in clean energy research or enforcing regulations on resource use, often involves short-term economic costs. These efforts are further undermined by short-term pressures, such as election cycles causing myopic decision making. [17]

The ongoing overuse and degradation of the Earth's resources is a case of 'tragedy of the commons' [14], where self-interested actions of individuals is ultimately leading to ruin of all. Averting large-scale ecological collapse and achieving effective stewardship of the Earth system requires a large-scale shift human behaviour away from the current status quo [10].

This requires recognising humanity as embedded within a complex adaptive system with by dynamic feedbacks and co-evolutionary processes between social and ecological components [32]. Effective interventions must therefore be grounded in an understanding of these feedbacks and designed to leverage mechanisms that foster and sustain long-term cooperation [20]. Insights into these mechanisms are relevant to collective risks across domains. These include the overharvesting of fisheries, the spread of pandemics where effective collective responses require overcoming short-term costs to secure long-term shared benefits

1.2 Literature Review

There has been a growing body of research on drivers of cooperation under the risk of climate change, through both theoretical models and laboratory experiments [37]. For instance, Milinski et al. [21] experimentally demonstrated that a higher perceived risk of collapse substantially increases cooperative investments in mitigation. Barfuss et al. [6] modelled climate change as a dynamic game and showing that stronger concern for the future can promote cooperation. Barrett and Dannenberg [9] demonstrated that ambiguity about the location of ecological tipping points discourages contributions to mitigation. Recent work has extended game-theoretic models by integrating behavioural insights on fairness [13, 25] and loss aversion [41].Pacheco et al. [29] showed how social interaction structures modulate cooperation under climate risk.

1.3 Research Focus: Information as a modulator of cooperation

While there has been substantial progress in research on the climate dilemma, the role of information in collective–risk dilemmas remains underexamined. Decision makers rarely act in isolation [33, 38]; they draw on feedback from two sources: *ecological information* (signals about environmental states and transitions) and *social information* (signals about others' behavior). Deciding to engage in sustainable practices first requires understanding that ecological system is degrading, and understand potential impacts [39]. Although such awareness has grown over recent decades, understanding remains uneven across societies, and the planetary crises are not yet fully recognised by all stakeholders. Climate policy choices are further shaped by broader social and political contexts [18], particularly by how actors perceive the efforts of others [26]. At the international level, climate agreements rely on mutual expectations and reciprocity: countries' willingness to adopt necessary mitigation targets often depends on whether others are seen to contribute their fair share. This holds true for individuals as well as firms, whose motivation contribute depends both on their understanding of ecological degradation and on information about how peers are responding.

Independent lines of research have highlighted these mechanisms in how they influence cooperation. Kleshnina et al. [19] show that access to environmental state information in dynamic games can alter cooperative behaviour depending on the underlying transition structure. [22] has demonstrated that noisy in perception of of the environmental state can influence cooperation. Likewise, social cues are essential for direct reciprocity a long established mechanism to support cooperation [3]. Building on these insights, we hypothesize that combining social and ecological information fosters cooperation more effectively than either source alone. Parks et al. [30] has experimentally examined how signals of others' behavior and environmental cues shape cooperation in a collective-pool harvesting game. However, a unified theoretical treatment that *jointly* examines ecological cues and social feedback in coupled socio-ecological systems is still missing. Identifying the minimal information requirements for sustaining cooperation can be very useful for effective stewardship. Our analysis could also have policy implications, as it clarifies how different measures, such as promoting environmental awareness and monitoring individual actions, affect outcomes.

1.4 Objectives

This thesis addresses the following research question:

> *How does access to social and ecological information influence the emergence of cooperation in collective-risk dilemmas?*

To investigate this, we develop an idealized model in which access to social and ecological information can be systematically varied. This design allows us to isolate the effects of each information type and examine their combined influence on cooperative behavior. Specifically, our aims are to:

(i) compare the relative effectiveness of social and ecological information in fostering cooperation;

(ii) investigate the interaction between the two types of information; and

(iii) examine how these effects depend on underlying incentive structures and ecological dynamics.

Chapter 2

Preliminaries

2.1 Multi-Agent Markov Environment

We model interaction among the decision-makers as a multi-agent Markov environment [1]

Formally, in such a system there are N agents, where each takes action a^i from the action set $A^i = A^i_1, \ldots, A^i_M$. M denotes the number of possible actions. The environment has Z possible states $\mathcal{S} = \{S_1, \ldots, S_Z\}$. The environmental dynamics are defined by the probability to transition from state s to s':

$$T_{sas'} = P(s' \mid s, a) \in [0, 1]$$

where a denotes the joint action of all agents. The transition probabilities over all possible current states, joint actions, and next states can be represented with a transition tensor T with dimensions $Z \times M \times \cdots$ (N times) $\cdots \times M \times Z$.

The assumption that the next state only depends on the current state and joint action makes the system Markovian. The rewards obtained by the agents across all possible transitions are represented by a reward tensor R with dimensions $Z \times M \times \cdots$ (N times) $\cdots \times M \times Z$.

The strategies (also referred to as policies or behavioural profiles) of all agents can be combined into a single object called the *joint policy tensor*, which serves as the dynamical variable of the system.

The strategy of each agent i at s:

$$X_s^i = \left[X_{sa_1}^i, X_{sa_2}^i, \dots, X_{sa_M}^i \right]$$

where $X_{sa}^i = P(a \mid i, s)$ is the probability of agent i selecting action a in state s. The individual strategies (also referred to as policies or behavioral profiles) of all agents can be combined into a single object called the *joint policy tensor*, with dimensions $N \times Z \times M$, which serves as the dynamical variable of the system. In the simplest case, the strategy is a fixed probability distribution $p_i(a_i)$ over the possible actions $a_i \in A_i$. More generally, an agent's strategy uses information available to them such as including actions taken by previously, or the current state of the environment.

The rewards for each agent depend on the depends on the actions chosen by all players and the state transition

$$R^i_{s,a_1,..,a_n,s'} \text{ or } R^i_{s,a,s'} \text{ in short}$$

where, s and s' are the current and next state respectively.

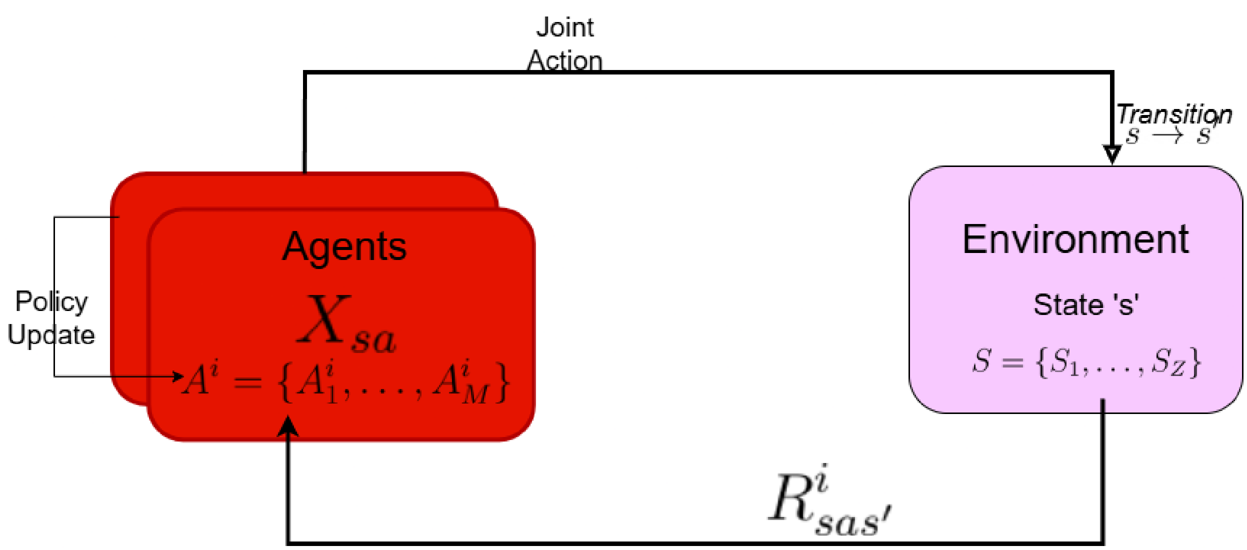


Figure 2.1: Schematic representation of a multi-agent Markov environment: N agents select actions according to their behavioral profiles X_{sa}^i. The resulting joint action determine the transition of the environment from state s to s' according to the transition probability $T_{sas'}$, with rewards $R^i_{sas'}$ assigned to the agent, depending on action and the state transition. The policy is a dynamical variable, updated with time.

2.2 Partially Observable Environments

In many systems, agents do not have direct access to the true state of the environment. Instead, they receive *observations* that provide only partial information about that state [40]. Let $\mathcal{O}$ be the set of possible observations. Since agents have an incomplete perception of the environment every observation $o \in \mathcal{O}$, corresponds to a *distribution* over possible true states of the system. $O(o \mid s)$ is the likelihood of receiving observation o when the true state is $s \in \mathcal{S}$

In the case of *deterministic partial observability*, which is the scenario we consider in our current analysis. The agent has a limited perception of the environment with $|\mathcal{O}| \leq |\mathcal{S}|$. Here, we can define a projection function

$$\omega : \mathcal{S} \to \mathcal{O},$$

which maps each true state $s \in S$ to a corresponding observation $o \in O$. The observation model can be expressed as

$$O(o \mid s) = \kappa(o, \omega(s)) \, ,$$

where $\omega(\cdot)$ denotes the observation (or projection) function, and κ represents the Kronecker-delta. Here the agent's policy X is defined over its observations rather than true states, i.e. $X^i_{oa} = P(a|i, o)$ with $\sum_a X^i_{oa} = 1$ for each agent i and observation $o \in \mathcal{O}$.

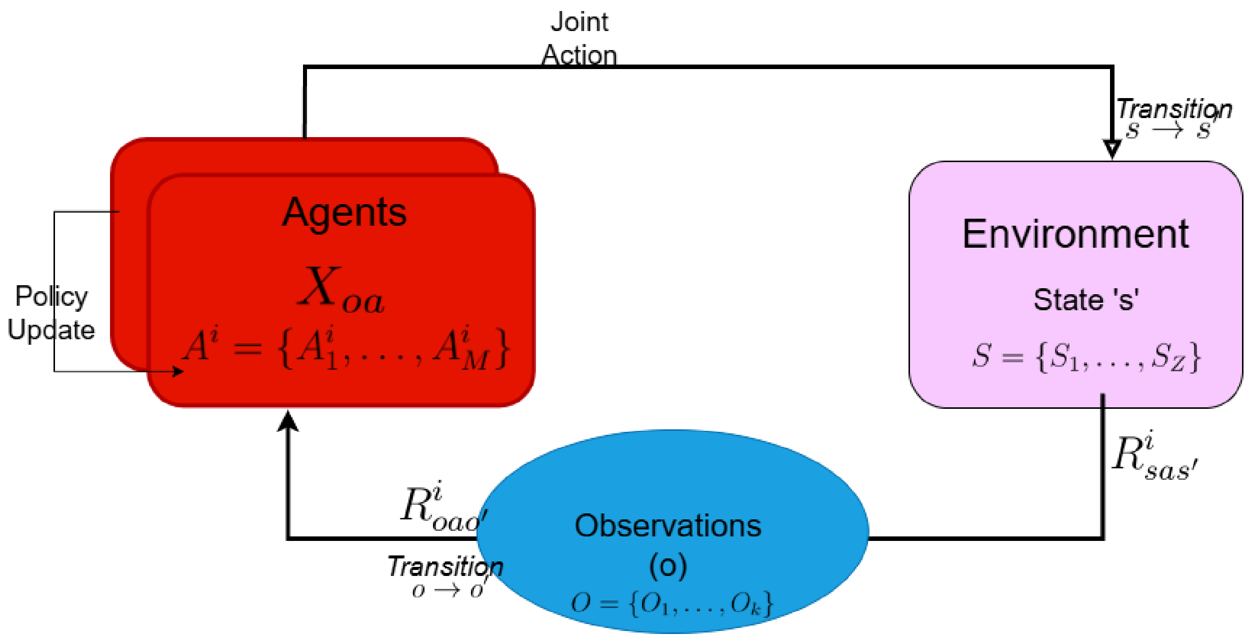


Figure 2.2: In a partially observable multi-agent environment, the underlying system still evolves over states $S = \{S_1, \ldots, S_z\}$, which transition according to the agents' joint actions. However, each agent perceives only partial information, observing the current state and its transitions through a reduced observation set $O = \{O_1, \ldots, O_k\}$. While the true system reward $R^i_{sas'}$ reflects the actual outcome of the state transition, the effective reward signal $R^i_{oao'}$ available to the agent is observation-based and serves as the learning signal for policy updates. In this case, the policy is maintained over observations X_{oa}, which is iteratively updated.

2.3 Games: Fundamental Concepts

A *game* is a formal representation of strategic interactions [28]. A game specifies a set of players, where rewards depends on the actions chosen by all players. It is said to be in *normal form* when all players choose their strategies simultaneously.

We can also consider *repeated games*, where the same game is played multiple times and rewards are accumulated over successive rounds. In a *stochastic game*, however, there are multiple environmental states which change probabilistically, each with its own reward structure. Note that the mathematical object of a *stochastic game* is synonymous with a *multi-agent Markov environment* described above, the former term being more common in economics and mathematics literature, while the latter is more widely used in dynamical systems and machine learning.

2.4 Model: Ecological Public Goods Game

Public Goods Game The *Public Goods Game* is a classic model that captures the tension between individual incentives and collective welfare [34]. It has been widely used to study situations involving shared resources, ranging from neighborhood cleanliness to national welfare schemes. In this game, each player decides whether to contribute to a common pool or to free-ride on the contributions of others. The total contribution is multiplied by an enhancement factor, which is > 1, to reflect the collective benefits of the contribution. The resulting amount is distributed equally among all participants, regardless of their individual contributions, reflecting the nature of a public resource. The socially optimal outcome occurs when everyone contributes; however, from an individual perspective, withholding one's contribution while benefiting from the efforts of others yields a higher personal payoff, leading to the social dilemma

The model we use in study is the *Ecological Public Goods (EcoPG) Game* [6]. This model is an extension of the standard Public Goods Game by incorporating an ecological tipping element to reflect that environmentally harmful actions is not only socially suboptimal but could lead to catastrophic consequences in the future. The actors in this framework may represent individuals, cities, corporations, or even countries. For simplicity, we begin with a two-actor case, although the model can be generalized to larger groups.

Each actor has two available actions: *Cooperation* (C) and *Defection* (D). Cooperation corresponds to climate-friendly actions, such as lowering carbon emissions and investing in renewable energy. Defection, on the other hand, represents opting for business-as-usual behavior that pollutes the environment. Cooperators bear a cost c while adopting of sustainable practices. The collective benefit for the group of $f \cdot \Sigma c_i$, which is divided among the 2 agents.

The net reward for each agent is simply the Benefit - Cost. For the two agent case, we will have three cases

- **Both defect:** No one contributes, so there is neither cost nor benefit, and each receives

$R = 0$

- **One cooperates:** The cooperator pays a cost c while the total benefit fc is shared equally, giving $R_{\text{coop}} = \frac{fc}{2} - c$ and $R_{\text{def}} = \frac{fc}{2}$.
- **Both cooperate:** Each contributes c, and the benefit $2fc$ is shared equally, yielding $R = fc - c$. for both agents.

The above scenarios apply as long as the Earth remains in the relatively safe operating zone, which we term the 'prosperous state'. Defection also carries with it another probabilistic risk of crossing a planetary tipping point, pushing the system into a 'degraded' state (corresponding to the Hothouse Earth scenario [35]). The probability of collapse grows marginally with each additional defector, at rate q_c/N. Once the system enters the degraded state, all actors suffer a negative environmental payoff $m < 0$, irrespective of individual behaviour at that point. Cooperation can marginally increase the chance of recovery, although it may take substantially longer.

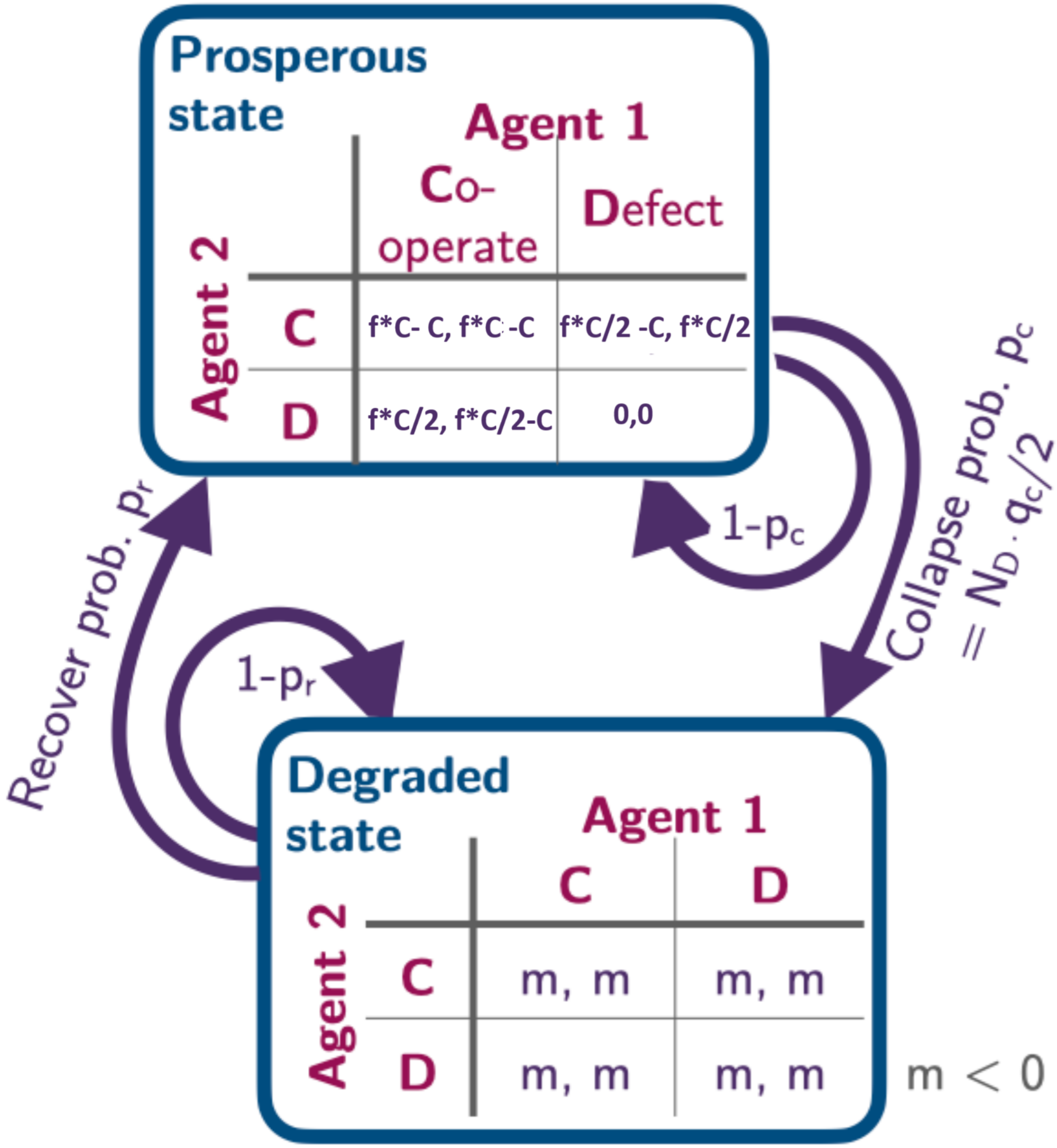


Figure 2.3: The Ecological Public Goods Game (EcoPG) for two actors. Unless otherwise specified, the parameters are set as: $f = 1.2$ (enhancement factor), $c = 5$ (mitigation cost), $q_c = 0.02$ (probability of collapse), and $q_r = 0.0001$ (probability of recovery).

For a comprehensive analysis of the Ecological Public Goods Game, see Barfuss et al. [6].

2.5 Multi-agent Reinforcement Learning (MARL) Dynamics

Multi-agent reinforcement learning (MARL) models how multiple agents learn and adapt their strategies through interactions within a Markov environment [1, 36]

This framework allows us to formalise how agents adapt in complex environments [7]. The agents take actions, experience *reinforcements* in the form of rewards, and iteratively adjust their policies to improve performance. Through this process of trial and error, agents aim to optimize their policies to maximize their expected cumulative (or average) rewards. Note that this optimization is taking place in a non-stationary environment; i.e., the reward each agent receives depends on the evolving policies of others. Nevertheless, each agent learns in an independent manner without any central coordination. Our goal is to understand how agents adapt their strategies over time and under what conditions such adaptation gives rise to cooperation.

Formally, each agent has an objective function, called the return (G) that it seeks to maximize through interaction with the environment.

$$G_i(t) = (1 - \delta_i) \sum_{k=0}^{\infty} (\delta_i)^k \, r_i(t + k)$$

which is the exponentially discounted sum of future rewards starting from time t.

where $r_i(t + k)$ is the reward received by agent i at time $t + k$, and $\delta_i \in [0, 1)$ is the *discount factor*. The discount factor determines how strongly the agent values the future relative to the present: $\delta_i \to 0$ makes the agent myopic, favoring immediate rewards, whereas $\delta_i \to 1$ makes it far-sighted, giving nearly equal weight to long-term outcomes.

2.5.1 Temporal Difference Learning Updates

The class of algorithms we use is temporal-difference (TD) learning [36]. TD learning is particularly suitable for our framework, as the policies are updated continuously in the learning process, allowing us to track the evolution over time. It is a model-free method, meaning that it does not require making explicit assumptions about how agents perceive their world. There is also research suggesting that TD learning approximates learning mechanisms in humans [27]

We present here the actor-critic algorithm, a specific implementation of Temporal Difference Learning. Alternate implementations such a Q-Learning or SARSA can also be used

2.5.1.1 Actor Critic Learning

In the actor-critic learning framework, there are two coupled components: the **actor**, which updates the agent's policy, and the **critic**, which evaluates the current policy by estimating the value function and providing feedback via TD errors on how the policy should be adjusted.

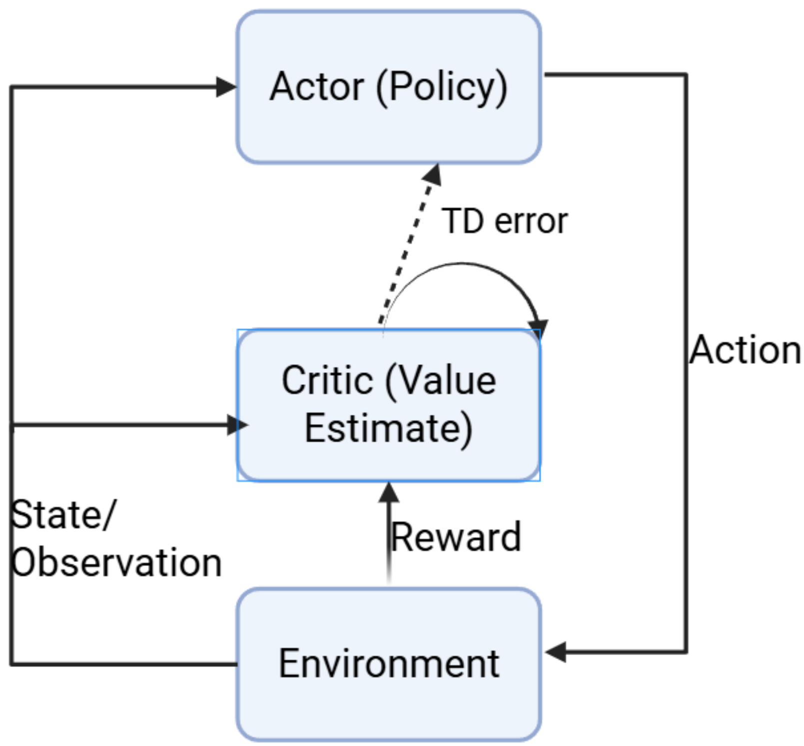


Figure 2.4: Schematic representation of the Actor–Critic algorithm. The actor updates the agent's policy, while the critic estimates the value function. The reward received from the environment after taking an action in a given state (or observation, in the case of partial observability) is used to compute the temporal-difference (TD) error, which provides a feedback signal to update both the value estimates and the policy.

Critic We first define a *state-value function*

$$v_s(X) = \mathbb{E}_\pi[G_t \mid S_t = s] \, ,$$

the expected return starting from state s and under policy X.

However, since the agent does not have prior knowledge of the environment's transition dynamics or reward distributions, it cannot compute $v_s(X)$ directly. Instead, the agent maintains *estimates*. $V(s)$ is used as an estimate of the true state-value $v_s(X)$. The estimates are successively updated by comparing the prediction for the current state with a more informed prediction obtained after an interaction step. This difference between successive estimates is called the *temporal-difference (TD) error*.

The *TD error* $D^i_{sa}(t)$ for a given state and action policy is:

$$D^i_{sa}(t) \;=\; R^i_{s(t)a(t)s(t+1)} + \delta V^i(s(t+1)) - V^i(s(t)). \tag{2.1}$$

Intuitively, $R^i_{s(t)a(t)s(t+1)} + \delta V^i(s(t+1))$ is the *target estimate* i.e. the immediate reward received upon taking action a at state s, combined with the estimated value of the next state. $V(S_t)$ is the *current estimate* of the value of the present state.

Deterministic Limit In the standard critic updates, error are estimated after each individual experience. Alternatively, we can imagine that agents interact with the environment many times before updating their estimates. The number of interactions here is called the batch size. If the batch size tends to infinity, then the randomness of individual samples averages out. This yields the *deterministic limit of temporal-difference (TD) learning* [5], where the learning dynamics become fully deterministic maps rather than stochastic updates, making them more tractable for analysis. Conceptually, this corresponds to introducing a separation of timescales - The agents explore and gather experience on a *fast timescale*, while updating their behaviour on the *slow timescale*.

This gives us equations for each possible state and action, with the single-sample rewards and next state value function replaced by their averages given the polices and transition probabilities under this deterministic limit.

$$^{\infty}D^i_{sa} = \underbrace{\langle R\rangle^i_{sa}}_{\text{avg. immediate reward}} + \delta^i \underbrace{\text{next}V^i_{sa}(X)}_{\text{avg. next-state value}} - \underbrace{V^i_s(X)}_{\text{avg. current-state value}} \tag{2.2}$$

$$\langle R\rangle = \sum_{a_{-i}} X^{-i}_{sa_{-i}} \sum_{s'} T_{saa_{-i}s'}\, R^i_{saa_{-i}s'} \tag{2.3}$$

As there are an infinite interactions before an update, the expected reward for agent i, taking action a at time t can be expressed as an average over the current policy of other agents X^{-i}, the transition probabilities T, and the reward function R^i for each possible joint action and resulting state transition.

$$\text{next}V^i_{sa}(X) = \sum_{a_{-i}} \sum_{s'} X^{-i}_{sa_{-i}}\, T_{saa_{-i}s'}\, v^i_{s'}(X) \tag{2.4}$$

In other words, the next V_{sa} represents the average of the true state values under the given distribution in the infinite–interaction limit. For the simulations, to compute this quantity, we make use of the fact that $v(s)$ satisfies the so-called Bellman consistency condition [36].

The current value estimate remains constant across actions, so it can be safely set to zero for tracking the policy evolution.

For a detailed derivation of the temporal difference updates under this deterministic limit, see Barfuss et al. [5]

Actor The **actor** represents the policy X, which is adjusted according to the critic's estimated error.

We define $Q^i_{sa}(t)$ as the *state–action value*, representing agent i's estimate of the expected return obtained by taking action a in state s and subsequently following its policy. In other words, $Q^i_{sa}(t)$ quantifies how rewarding each action is expected to be in a given state. The behavioral policy of agent i is thus modeled as a softmax distribution over these Q-values:

$$X^i_{sa}(t) \;=\; \frac{\exp\big(\beta^i\, Q^i_{sa}(t)\big)}{\sum_b \exp\big(\beta^i\, Q^i_{sb}(t)\big)} \tag{2.5}$$

The parameter β^i is the *intensity of choice*, which controls the exploration–exploitation trade-off.

The Q values are updated using the same temporal difference error computed by the critic,

$$Q^i_{sa}(t+1) = Q^i_{sa}(t) + \alpha^i D^i_{sa}(t). \tag{2.6}$$

where $\alpha^i \in (0,1)$ is the *learning rate.*

On substituting (2.6) into (2.5), we get

$$X^i_{sa}(t+1) \; = \; \frac{X^i_{sa}(t) \, \exp\bigl(\alpha^i \beta^i D^i_{sa}(t)\bigr)}{\sum_b X^i_{sb}(t) \, \exp\bigl(\alpha^i \beta^i D^i_{sb}(t)\bigr)}. \tag{2.7}$$

This expression governs the learning dynamics.

2.6 Learning Dynamics in Partially Observable Environments

In a partially observable environment (Section 2.2), the policy is defined over observations rather than states i.e as X_{so}. However, the rewards and transition probabilities are still dependent on the effective *state policy*. We derive the effective state policy $Y^j_{sa_j}$ by averaging the observation-based policy $X^j(o_j, a_j)$ over the observation likelihood for each state:

$$Y^j_{sa_j} = \sum_{o_j \in \mathcal{O}^j} O^j(o_j|s) \, X^j_{o_j a_j} \tag{2.8}$$

This allows us to obtain policy-averaged transition kernels $T_{ss'}$:

$$T_{ss'} = \sum_a Y^i_{sa} \, T_{sas'} \tag{2.9}$$

To relate observation-based actions to rewards, we further compute a posterior distribution over system states for the given observation. The posterior $B(o, s)$ is defined according to Bayes' rule.

$$B(o, s) = \frac{O(o|s) \, \bar{P}(s)}{\sum_{u \in \mathcal{S}} O(o|u) \, \bar{P}(u)} \tag{2.10}$$

where $\bar{P}(.)$ is the stationary probability of states under the current joint policy, serves as the prior.

This stationary probability is computed using the normalised eigenvector corresponding to the eigenvalue-1 of Tss' (2.9) for the given policy.

Using equations 2.10 and 2.8, we can reformulate rewards and transitions over observations rather than states.

$$R^i_{oa_i} = \sum_{s,\,a_{-i},\,s'} B(o,s) \left(\prod_{j \neq i} \bar{Y}^j_{sa_j} \right) T_{sas'}\, R^i_{sas'} \tag{2.11}$$

$$T^i_{oa_io'} = \sum_{s,\,a_{-i},\,s'} B(o,s) \left(\prod_{j \neq i} \bar{Y}^j_{sa_j} \right) T_{sas'}\, O^i(o'|s') \tag{2.12}$$

Intuitively, the rewards 2.11 and transition probabilities of the observations corresponding to the observations are derived by averaging over the posterior state distribution $B(o,s)$ and the effective state policies $\bar{Y}^j(s,a_j)$.

Value functions and temporal-difference updates follow the same formulations as derived in Section 2.5.1, except with observations substituted for states.

Under partial observability, the learning dynamics are thus:

$$X_{oa}(t{+}1) = \frac{X_{oa}(t)\, \exp\bigl[\alpha\beta\, D_{oa}(t)\bigr]}{\sum_b X_{ob}(t)\, \exp\bigl[\alpha\beta\, D_{ob}(t)\bigr]}.$$

Please see [4] for a detailed explanation of learning dynamics under partial observability.

Remark: Non-Ergodicity If all states s that map to an observation $\tilde{o}$ have stationary probability zero, the posterior belief $P(s \mid \tilde{o})$ is undefined. Computationally, however, in such a case, $P(s \mid \tilde{o})$ is treated as 0 to avoid such errors. This results in the expected reward and transition distributions for that observation ($\tilde{o}$) reducing to zero as well

$$\mathbb{E}[R \mid o] = \sum_s B(o,s)\, R(s,\cdot) = 0, \qquad \mathbb{E}[T \mid o] = \sum_s B(o,s)\, T(s,\cdot,\cdot) = 0,$$

which implies that no effective learning occurs for that observation. In other words, the agent's learning dynamics completely ignore any observation if it isn't observed in the steady state.

2.7 Simulations (Illustrative Examples)

To provide a reference for interpreting the phase portraits and learning trajectories in Chapter 4, we first outline the learning dynamics in simpler games.

2.7.1 Single-State Game (Repeated Prisoner's Dilemma)

To illustrate the basic dynamics, we begin with the simple case of a repeated Prisoner's Dilemma. This is a single-state, two-player game with the following payoff matrix:

	C	D
C	$(0.8, 0.8)$	$(-0.5, 1)$
D	$(1, -0.5)$	$(0, 0)$

The entries show the payoffs in the form (Player 1, Player 2). In this setup, for both players, unilateral defection yields the highest payoff. However, mutual cooperation is better than mutual defection.

In this game, the learning trajectories tell how agents adapt their strategy based on the reward feedback. The evolution of agents' strategies with time based on the learning updates described above can be visualising as a phase portrait.

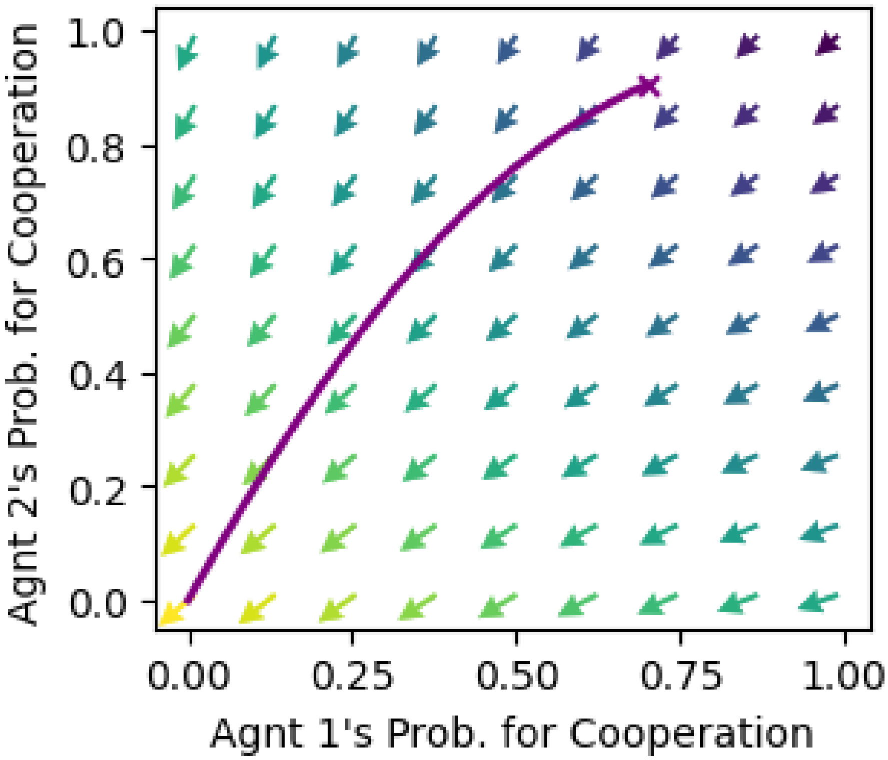


Figure 2.5: Learning update phase portrait: Prisoner's dilemma

Each point in Fig. 2.5 represents a unique joint strategy, in this case, defined by the cooperation probabilities of the two agents (p_c^1, p_c^2). The arrows depict the TD errors: their direction indicates how each agent's cooperation probability is updated during learning, while their length reflects the magnitude of the update. The purple trajectory shows an idealised learning path starting from randomly chosen initial conditions. Over time, these trajectories converge towards the attractor of the system, in this case complete mutual defection (zero cooperation).

The attractor corresponds to the stable Nash equilibrium, as neither agent has any incentive to unilaterally deviate.

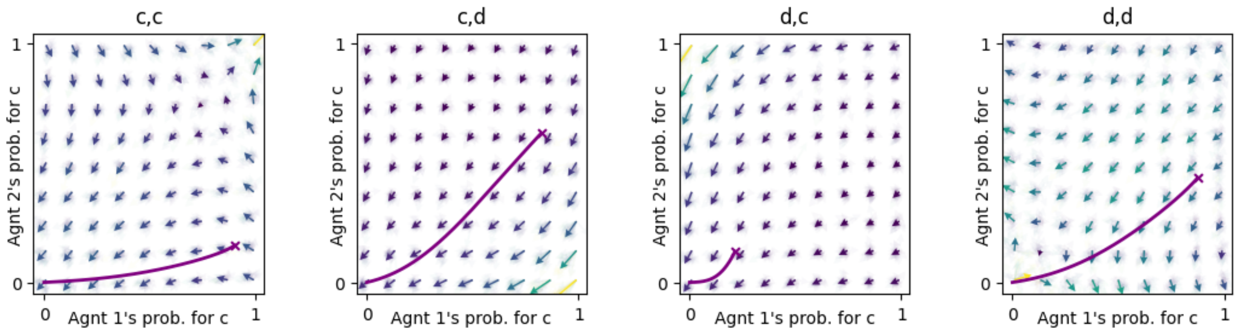


Figure 2.6: Learning update phase portrait: Prisoner's dilemma. Projections for each of the 4 states are shown separately.

2.7.2 Multi-State Game (Repeated Prisoner's Dilemma -Memory-1)

We now consider a version of the Prisoner's Dilemma with the same payoff matrix, but in which the agents possess memory, allowing them to adopt distinct strategies for different interactions. Since each agent could have either cooperated (C) or defected (D) in the preceding round, the system comprises four possible states CC, CD, DC, and DD representing all combinations of the agents' prior actions. Accordingly, the learning problem is defined over the 4-dimensional state space.

For interpretable visualisation, we plot the phase portrait projections for each state separately. However, the temporal-difference update within a given state depends on the agents' strategies in the other states. To account for this interdependence while plotting, we approximate the update by averaging over many randomly sampled strategies in the remaining states.

In Figure 2.6 the faint grey arrows represent the raw updates obtained from individual samples, while the colored arrows indicate the averaged update. The coloured indicate the expected direction of learning.

Chapter 3

Methods

In this section, we describe the model implementation along with the treatments, as well as the simulation procedure for the main outputs we studied.

3.1 Information regimes

In our model, we are interested in how the information available to agents influences the emergence of cooperation. To capture this systematically, we distinguish between two sources of information: *ecological information* and *social information*

3.1.1 Ecological information

Ecological information refers to the knowledge agents possess about the state of the Earth system. In the Ecological Public Goods Game (Section 2.4), we consider a simplified setting with two possible environmental states: a *prosperous* state, in which ecological conditions remain favorable, and a *degraded* state, representing a "Hothouse Earth" scenario that emerges once planetary boundaries are crossed. When ecological information is available, agents can adopt strategies that are conditional on the environmental state. This implies that their adaptation explicitly incorporates the possibility of ecological degradation and recovery, making ecological dynamics a factor in decision-making. By contrast, in the absence of ecological information, the ecology is represented as a single undifferentiated state.

3.1.2 Social information

Social information refers to the awareness agents have of the behavior of the interaction between players. Specifically, this information allows an agent to observe the actions taken by other group members and makes it possible to adopt strategies that explicitly depend on the actions of others. For simplicity, we restrict our analysis to strategies that depend only on the most recent round of play. In this case, an agent may assign different probabilities of cooperation or defection depending on its actions and the opponent's response in the previous round. An agent's behaviour depends on the previous interaction—that is, on the combination of its own action and the action of its partner Since each player can either cooperate (C) or defect (D), a two-player interaction yields four possible joint actions: $(C, C), (C, D), (D, C), (D, D)$.

A strategy in the presence of social information therefore specifies a probability of cooperation for each of these four cases. For example, one possible strategy is:

$$(p_C^i(CC),\ p_C^i(CD),\ p_C^i(DC),\ p_C^i(DD)) \ = \ [1,\ 0,\ 0,\ 1], \tag{3.1}$$

where $p_C^i(\cdot)$ denotes the probability that agent i cooperates, given the previous joint action. This can be read as: agent i always cooperates after mutual cooperation (CC) or mutual defection (DD), and always defects if the previous interaction was (CD or DC).

This is often referred to as a *memory-1* strategy. By contrast, when social information is absent, agents cannot condition on past actions at all. Their behavior is then independent of social feedback and depends only on other available information, which is called memory-0 strategies in the literature [34].

3.1.3 Experimental Treatments

We employed a 2×2 factorial design to systematically isolate and evaluate the effects of social and ecological information, as well as their interaction. These information modes were implemented as distinct regimes within the Ecological Public Goods (EcoPG) framework. We assumed that the environmental state in the EcoPG game is defined jointly by the ecological and social contexts. Specifically, we represent the state as

$$s_t = (e_t,\ a^1_{t-1},\ a^2_{t-1}) \tag{3.2}$$

where $e_t \in \{p, g\}$ denotes the ecological condition - either prosperous (p) or degrade (g), and $a_{t-1}, a'_{t-1} \in \{C, D\}$ correspond to the previous actions of agent 1 and agent 2 respectively.

To implement the different information conditions, we followed the framework of partial observability where agents lack direct access to the true system state s_t, but instead rely on observations o_t (see Section 2.2). For each information regime r, the observation is determined by a deterministic mapping,

$$o_t = \omega_r(s_t); \quad \omega_r : S \to O_r. \tag{3.3}$$

which specifies which components of the underlying state are visible to the agent under that regime. The probability of receiving o_t in state s_t is therefore

$$O(o_t \mid s_t) = \delta^k(o_t,\, \omega_r(s_t)) \tag{3.4}$$

here δ^k is the Kronecker delta function.

We thus obtain a factorial design with four treatments differing in terms of the information the agents possess.

I. **Complete:**

$$\omega_{\text{Com}}(e_t, a^1_{t-1}, a^2_{t-1}) = (e_t, a_{t-1}, a'_{t-1}).$$

In the complete information case, the agent has access to both ecological and social cues. This results in 8 (4 action histories x 2 ecological conditions) distinct observations:

$$\{\, p_C^i(pros., CC),\ p_C^i(deg., CC),\ p_C^i(pros., CD),\ p_C^i(deg., CD),$$
$$p_C^i(pros., DC),\ p_C^i(deg., DC),\ p_C^i(pros., DD),\ p_C^i(deg., DD) \,\}$$

where "pros." = prosperous state, "deg." = degraded state.

II. **Ecological-Only:**

$$\omega_{\text{Eco}}(e_t, a_{t-1}^1, a_{t-1}^2) = e_t.$$

This yields two possible observations of the state: prosperous (p) and degraded (g).

III. **Social-Only:**

$$\omega_{\text{Soc}}(e_t, a_{t-1}^1, a_{t-1}^2) = (a_{t-1}^1, a_{t-1}^2),$$

resulting in 4 observations: $(C, C), (C, D), (D, C), (D, D)$, corresponding to possible joint actions in the previous round.

IV. **No Information:**

$$\omega_{\text{No-Info}}(e_t, a_{t-1}^1, a_{t-1}^2) = o_0.$$

All system states map to a single undifferentiated observation, thus agents receive no information about the environment or past actions.

The strategies of an agent i under each information regime are shown below (Table 3.1). $p_C^i(\cdot)$ denotes the probability of cooperation given the corresponding observational state.

Regime	**Strategy**	**Number of States**
Complete Information	$\{p^i_C(pros., CC),\, \ldots,\, p^i_C(deg., DD)\}$	8 (4 prosperous)
Ecological-Only	$\{p^i_C(p),\, p^i_C(g)\}$	2 (1 prosperous)
Social-Only	$\{p^i_C(CC),\, p^i_C(CD),\, p^i_C(DC),\, p^i_C(DD)\}$	4
No Information	p^i_C	1

Table 3.1: Strategy representation and number of observable states under each information condition.

3.2 Simulations

We conducted learning dynamics simulations (Section 2.6) under each of the four information conditions. Each simulation was initialised from a given policy and terminated once convergence was achieved, numerically defined as successive updates becoming smaller than a tolerance threshold of 10^{-5}, i.e., the simulation was terminated after a maximum of 25,000 time steps. The learning rate α was set to 0.05, chosen deliberately small in order to minimize oscillations during the update process. The converged policy, or the final policy obtained at the termination of a run, was recorded for evaluation. Note that in our simulations, since our primary interest was in the adaptation processes that mitigate the risk of ecological collapse, learning updates were disabled in the degraded states to avoid confounding effects. We show the phase portraits along with example trajectories for the selected parameter settings in Section 4.1 for reference.

3.2.1 Code Implementation

All simulations were implemented in `Python`. The game environments and reinforcement learning dynamics were implemented using the `pyCRLD` package. Code required to reproduce the figures will be made open source upon publication.

3.2.2 Evaluating Cooperation Levels

The quantities of interest from these simulations were the final strategies X^* learned by the agents and the resulting effective cooperation level. The effective cooperation probability ($\bar{p}_c^*$), was taken as an average over the stationary distribution of observational states ($\sigma_X(o)$) and over the two agents.

$$\bar{p}_c = \frac{1}{2} \sum_i \sum_o p_{i,c}(o)\, \sigma_X(o) \tag{3.5}$$

The outcomes were classified with the following thresholds that allowed classification of final polices as being cooperative, defective, or mixed with a margin of deviation.

$$\text{Outcome} = \begin{cases} \text{Cooperate} & \text{if } \bar{p}_c^* \geq 0.95, \\ \text{Mixed} & \text{if } 0.05 < \bar{p}_c^* < 0.95, \\ \text{Defect} & \text{if } \bar{p}_c^* \leq 0.05. \end{cases} \tag{3.6}$$

where $\bar{p}_c^*(.)$ is the average cooperation levels at the final point of the dynamics.

3.2.2.1 Basin of Attraction Size

We simulated $N = 500$ initial conditions for each of the four information regimes. Initial conditions were generated using Latin hypercube sampling (LHS), a stratified sampling method for uniform coverage of high-dimensional spaces. This sampling procedure is as follows: for each dimension, the uniformly distributed range is divided into N equal-width intervals, and one sample is drawn from each interval. The samples from different dimensions are then combined using independent random permutations, as in Latin hypercube sampling. In this case, the total number of dimensions is $|\mathcal{O}| \times$ number of agents. Because LHS provides an uniform sampling of the policy space, the size of the basin of attraction for cooperation can be estimated as the percentage of sampled initial conditions that converge to cooperative equilibria, expressed as

$$B_{\text{coop}} = \frac{N_{\text{coop}}}{N} \times 100\% \tag{3.7}$$

where N_{coop} is the number of simulations that converge to cooperative outcomes and N is the total number of simulations. Thus, B_{coop} represents the fraction of the joint policy space that evolves toward cooperation, thus serving as a measure of the cooperation levels under a chosen parameter set.

3.2.3 Strategy Analysis

In addition to cooperation levels, we looked at the strategies that emerged from the simulations to gain a more mechanistic understanding of the dynamics under each information condition

3.2.3.1 Local Stability Analysis

To systematically identify which policies are stable, we exhaustively evaluated all deterministic strategies under each of the four information regimes. In a deterministic strategy, the probability of cooperation in every possible state is fixed to either 0 (always defect) or 1 (always cooperate). If an information condition has k states, this results in 2^k distinct deterministic strategies. Since the interaction involves two agents, there are $\binom{2^k}{2}$ possible pairs of initial strategies to test. A strategy X was classified as *stable* if the learning dynamics, when initialized with policy X, converged to a final policy X^* within a tolerance

$$\text{stable if } |X^* - X\| < 10^{-2} \tag{3.8}$$

3.2.3.2 Strategy-Specific Basin of Attraction

We also characterised the distinct strategies that emerged from the randomly initialised simulations and quantified the frequency of their occurrence. To facilitate interpretation and visualisation, the strategies obtained from the simulations were rounded to one decimal place and aggregated. In the only-ecological and complete information case, only the strategies in the prosperous state were considered for analysis. Similar to Equation 3.7 we estimated its *strategy-specific basin of attraction* s, the basin size is computed as

$$B_s = \frac{N_s}{N} \times 100\%,$$

where N_s denotes the number of simulations whose final strategy matched s, and N is the total number of simulations for that information regime.

3.3 Comparison across Game Regimes

We repeated the cooperation levels evaluation and strategy analysis across different parameter regimes to ensure that our results are robust. In Barfuss et al. [6], the authors highlighted that in the ecological public goods game, three distinct regimes emerge depending on how strongly agents value the future (operationalised as the discount factor): a tragedy regime where mutual defection is the only stable outcome, a coordination regime where both mutual cooperation and mutual defection are bistable, and a comedy regime where cooperation is the only stable outcome irrespective of initial conditions. Using these results as a basis, we aimed to investigate the role of information across each of these regimes. Note that the original EcoPG model in [6] did not impose partial observability. Instead, it assumed that agents had access only to ecological information. This is equivalent to the *"only ecological information"* condition in our framework.

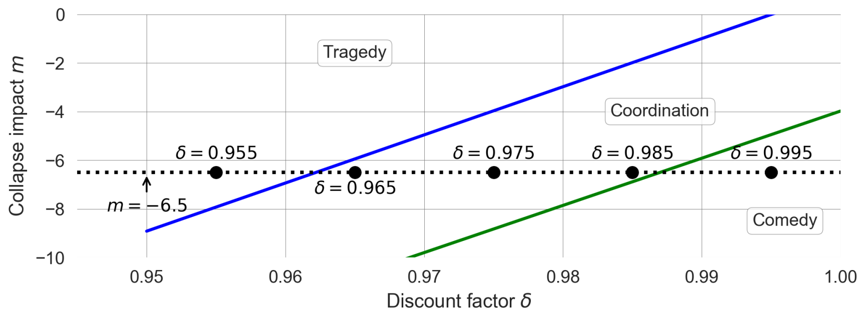


Figure 3.1: Figure reproduced from Barfuss et al. [6]. The plot illustrates the three regimes of the ecological public goods game as a function of the discount factor (δ) and the collapse impact (m). The blue and green lines mark the transitions separating the tragedy,coordination, and comedy regimes. The dotted line indicates the fixed collapse impact used for this analysis ($m = -6.5$), and the marked points show the sampled discount factors, $\delta \in \{0.955, 0.965, 0.975, 0.985, 0.995\}$

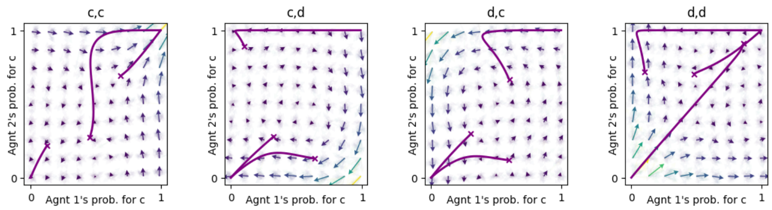
In our work, we chose a gradual and uniform sweep across δ allows us to examine how foresight modulates the emergence of cooperation under each information condition. As shown in the Figure 4.3, the parameters were $\delta \in \{0.955, 0.965, 0.975, 0.985, 0.995\}$ for a fixed $m = -6.5$. Thus, the first point lying in the tragedy regime, three in different positions within the coordination regime, and the last in the comedy regime.

Chapter 4

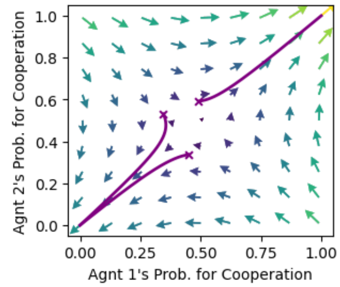
Results

4.1 Learning Dynamics Visualisation

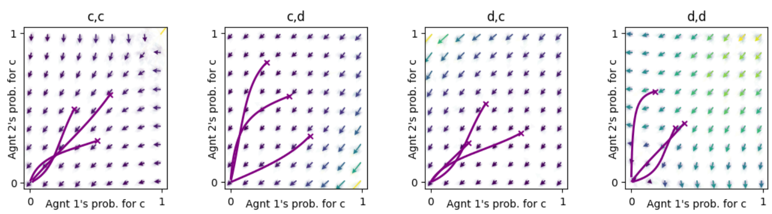
We first illustrate the phase portraits along with example trajectories from the simulation in each of the four information conditions (Fig. 4.2). Note that only the prosperous state projections are shown for the complete and only ecological information case.



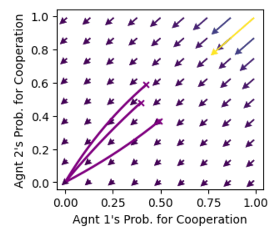
(a) Complete Information



(b) Ecological Information



(c) Only Social Information



(d) No Information

Figure 4.2: Phase portrait projections for each information regime. Three trajectories (in purple) are visualized, each starting from a random initial condition, $m = -6.5, \delta = 0.975$ for all four regimes.

By inspecting Fig. 4.2, we gain a qualitative understanding of the system's dynamics under different information regimes. When both ecological and social information (Fig.4.2 a) are available, we see complex patterns with bistability for c,c, mutual cooperation as the global attractor for d,d and limit cycles when the agent's previous actions were asymmetric. Under ecological-only information (Fig. 4.2b), the system exhibits bistability. When only social information (Fig. 4.2c) is available, we see convergence towards mutual defection, with minor local deviations in the flow. Finally, in the absence of information (Fig. 4.2d), the learning updates are uniformly towards mutual defection, regardless of the initial condition.

4.2 Comparison of Cooperation Levels Across Information Regimes

4.2.1 Baseline Comparison

To illustrate the role of information, we first compare the results for a specific case, where we assume intermediate collapse impact (payoff at the degraded state, $m = -6.5$) and moderate concern for the future (discount factor $\delta = 0.975$), while keeping all other parameters at their default values as specified in the model description (Section 2.3). In subsequent sections, we also systematically vary these parameters to examine the robustness of the observed patterns.

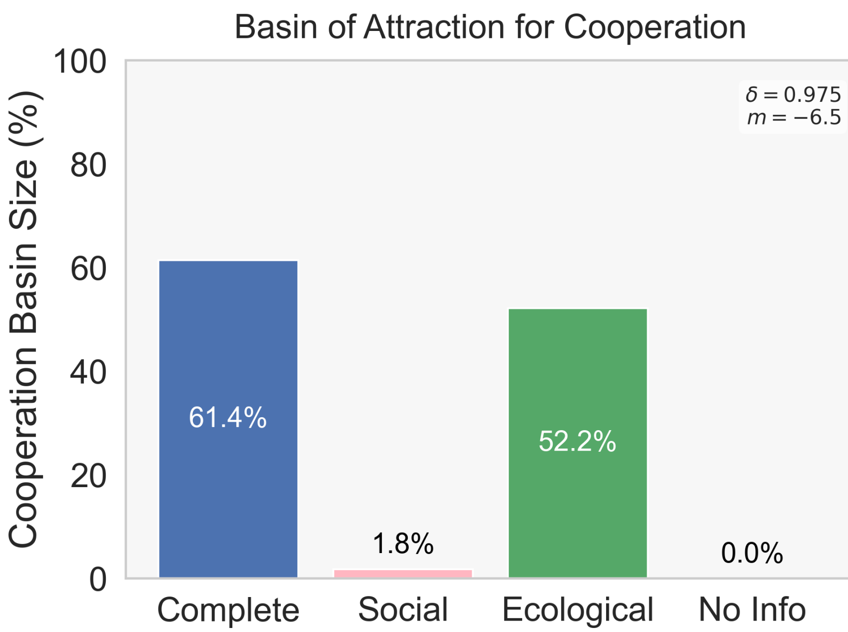


Figure 4.3: Cooperation Levels Across Each Information Regime. The basin size is calculated as % of the initial polices that converged to a final policy with average cooperation over 0.95

Figure 4.3 depicts the relative size of the cooperation basin of attraction. Under complete information, we have the largest basin of attraction for cooperation, with 61.4% of initial conditions converging to cooperative outcomes. This indicates that access to both social and ecological cues promotes cooperation substantially. When only ecological information is available, the basin of attraction remains large at 52.2%, suggesting that ecological feedback alone can sustain cooperation in a significant fraction of cases, albeit slightly less effectively than under complete information. In contrast, under the only social-information condition, the basin of attraction is reduced to 1.8%, showing that social cues in the absence of ecological awareness is mostly insufficient to maintain cooperative behaviour. Finally, in the no-information condition, cooperation does not emerge at all.

4.2.2 Comparison across Game Regimes

As described in Section 3.3 we compared the cooperation levels for each of the sampled discount factors $\delta \in \{0.955, 0.965, 0.975, 0.985, 0.995\}$

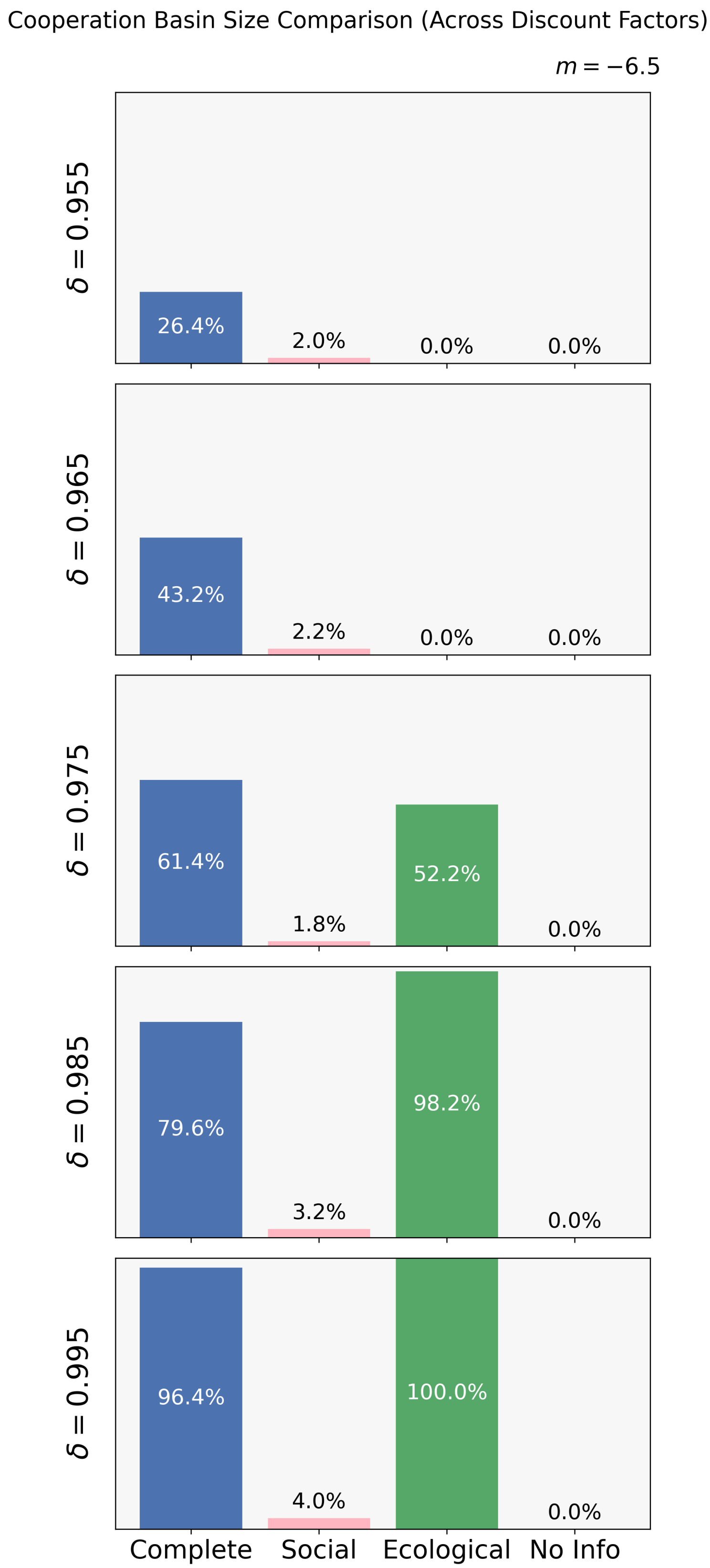


Figure 4.4: Cooperative basin size across information conditions for different discount factors (δ), with the collapse impact fixed at $m = -6.5$. Each panel shows the percentage size of the basin of attraction for the given parameter choice of the discount factor.

At low values of δ (0.955–0.965), cooperation is maintained almost exclusively under *complete information*. Under other conditions, we see minimal or no cooperation. With intermediate values of δ (around 0.975), *ecological information* begins to play a stronger role—its cooperative basin becomes comparable to that under *complete information*, both reaching roughly 50% cooperation. At high δ (0.985–0.995), the ordering reverses: *ecological information* produces the largest cooperative basin, approaching full cooperation, while it is slightly lower in case of complete information. *Social information* contributes minimally across all discount factors. For the *no information* case there is no cooperation irrespective of the discount factors.

Overall, we see two trends: (i) cooperation increases monotonically with higher δ when ecological information is available, and (ii) the relative ordering of cooperation levels under each information regimes is not fixed: complete information > ecological information at lower δ, but for higher δ the opposite is true.

4.3 Local Stability Analysis

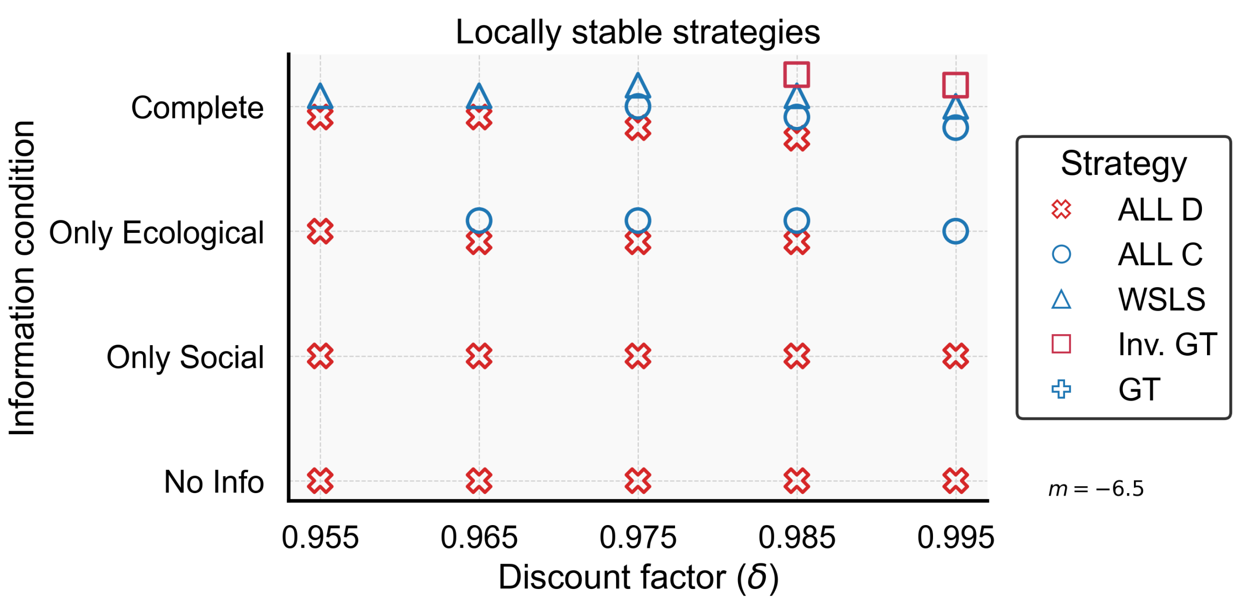


Figure 4.5: Set of locally stable strategies (indicated by different symbols) for each information condition (y-axis) and discount factor (x-axis). Each marker corresponds to a strategy that remains locally stable under that combination of parameters. Blue represents cooperative strategies, while red corresponds to defective strategies $m = -6.5$.

Explanation of Strategy Labels To make the strategies interpretable, we refer to several canonical examples from the literature, [24, 34], by their standard shorthand names. When social information is present, a deterministic strategy specifies an action (cooperate = 1, defect = 0) for each of the four possible outcomes of the previous round: (C, C), (C, D), (D, C), and (D, D) .

Name	**Policy**	**Explanation**
ALLC (Always Cooperate)	**1111**	Cooperates in every possible state; never defects.
ALLD (Always Defect)	**0000**	Always defects regardless of the partner's previous move.
WSLS (Win–Stay, Lose–Shift)	**1001**	Repeats the previous action if the last outcome was favourable (CC or DC) for the first agent, and switches otherwise (CD or DD).
GT (Grim Trigger)	**1000**	Cooperates until the partner defects once; thereafter defects forever.
*Inv. GT** (Inverse Grim Trigger)	**0001**	Cooperates only after mutual defection in the previous round; defects otherwise

Table 4.1: Explanation of legend entries denoting stable strategies for ecological and complete information. Apart from Inv. GT which emerged as a peculiar outcome in our simulations, the other strategies are commonly seen in the game theoretic analysis

In the no information case, there is only a single state, and in the ecological-only information condition, only one prosperous state. In these cases, the only possible outcomes correspond to uniform strategies either mutual defection **0** or mutual cooperation **1**. For convenience, are still denoted as **ALL D** and **ALL C** in the figures.

. . . .

Under *no information*, the always defect (ALL D) strategy is consistently the only stable equilibrium, indicating that cues are necessary for conditional cooperation. With only *social information*, the outcome remains similar; ALL D is the only attractor for these parameters.

Under only *ecological information*, as shown in [6], we have bistability between cooperative and defective equilibria at intermediate values (e.g., $\delta = 0.965$), with transitions toward cooperative outcomes at higher discount factors. *Complete information*, which combines both social and ecological cues, enables the emergence of cooperation even in regimes that would otherwise favour defection. In this case, ALL C becomes stable for higher δ - this indicates that ecological feedback can reinforce unconditional cooperation that is otherwise prone to exploitation. In addition to ALL D, both ALL C and WSLS remain stable over a broad range of discount factors, and at the highest discount factor ($\delta = 0.995$), the inverse Grim Trigger (Inv. GT) strategy is also a locally stable equilibrium. Inv. GT represents a peculiar case in which agents cooperate only following mutual defection, leading to oscillatory dynamics where the stationary state is dominated by defection with brief cooperative episodes. Overall, we find that local stability analysis accounts for the patterns observed in the basin of attraction size trends discussed previously.

4.4 Strategy-Specific Basin of Attraction

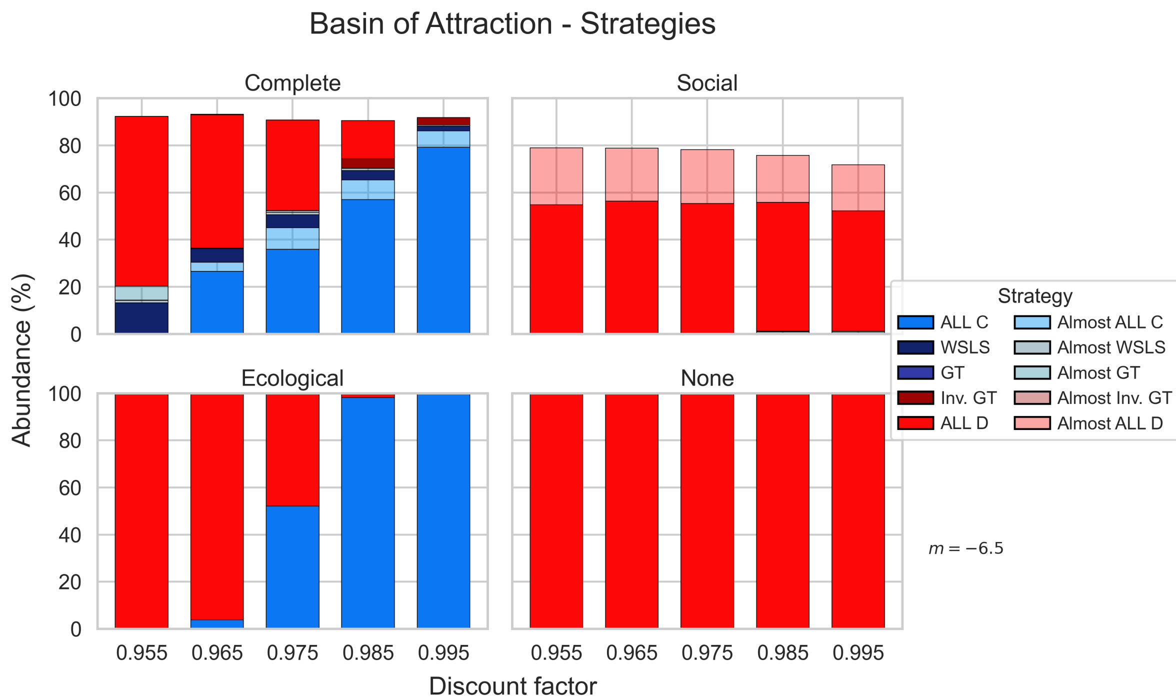


Figure 4.6: Basin of attraction of individual strategies across information conditions and discount factors (δ). Each stacked bar shows the proportion of initial conditions converging to the basin of a given strategy. Distinct colours represent pure strategies (ALL C, WSLS, GT, Inv. GT, ALL D), while lighter tints correspond to their "Almost" variants, which are mixed strategies that can be approximated as to the pure forms by rounding off to the policy to the nearest integer. Bluish shades correspond to cooperative strategies, while reddish correspond to defection. Asymmetric or mixed final strategies that could not be approximated by any of the canonical types are classified as "others" and omitted; consequently, the stacked bars do not sum to 100 % for the complete and social information conditions.

Here, we have a finer-grained representation of the results observed in the cooperative basin of attraction. Notice that the strategies with significant abundance are mostly the locally stable ones. Under complete information, increasing the discount factor δ leads to a smooth redistribution, with the always-cooperate strategy growing steadily, while always defecting correspondingly declines. Under social information only, apart from (ALL D), we see a non-trivial fraction of mixed and asymmetric strategies which are classified as "Others", this is likely due to the non-ergodic nature of learning dynamics under partial observability Remark 2.6.

Chapter 5

Discussion

The results establish information as an important modulator of cooperation in socio-ecological dilemmas. In particular, reliable access to ecological cues are shown to be essential for sustaining collective action. We see a significant difference in cooperation levels observed with and without explicit environmental feedback; this highlights an important theoretical insight: sustained cooperation arises only when agents actively model the environmental state, whereas mere implicit penalties associated with environmental degradation are insufficient.

For the parameter combinations examined, when both social and ecological reciprocity operate, we see a context-dependent pattern. When agents are relatively short-sighted, social information enhances coordination and promotes cooperative behaviour. But when agents are sufficiently future-oriented and ecological information is already sufficient, the same social feedback reduces cooperation levels. This is because of the presence of multiple attractors across parameters, including cooperative ones (e.g., WSLS, ALLC) and defective ones (e.g., ALLD, Inv GT). As a result, a fraction of initial conditions with either mutually high or mutually low levels of cooperation become trapped in their respective attractors. By contrast, when only ecological information is available, we can have monostable equilibria (defection at low discount factors, and cooperation at high discount factors) leading to more extreme cooperation levels.

Our model conceptually aligns with the conditional return game proposed by Kleshnina et al. [19], though it differs in the transition dynamics. Their analysis indicated that information about the ecological state could have either positive or negative effects on cooperation, depending on the recovery rate. It would be interesting to test whether we can replicate similar results in our model by varying the recovery rate

We find that when only *social information* is available, defection becomes the only stable outcome, which is a counter-intuitive result, since direct reciprocity is expected to promote cooperation [23]. We hypothesise that under higher f - the public goods enhancement factor, access to social information may again allow for the emergence of cooperation,

The results may also offer an alternative yet broadly compatible interpretation of the experimental findings by Parks et al. [30], who reported that ecological information consistently supports cooperative behaviour, whereas social cues can increase defection.

In light of these results, our model offers insight into why awareness about climate change plays a very significant role in collective outcomes. Policies that enhance access to reliable ecological information—such as transparent indicators of degradation, early-warning systems for tipping points, or locally visible feedback—should be prioritised to foster sustained cooperation. By contrast, one could speculate that social reciprocity-based enforcement mechanisms in international agreements may yield heterogeneous outcomes, i.e. while such mechanisms can stabilise cooperation among actors who are not sufficiently future-oriented, they can also undermine trust. In some cases, this may trap parties in cycles of mutual defection—situations where cooperation might have emerged more easily without social enforcement. This was also noted by [12] who argued that improved transparency in environmental efforts does not always translate to higher accountability.

Although the present model offers a useful stylised representation, it abstracts from real-world complexities, for example in asymmetries in leverage and resources among actors. We discuss potential extensions to address these factors in Chapter 6.

Chapter 6

Future Work

Building on the results presented here, future extensions could refine the analysis and yield a more comprehensive understanding of cooperative behaviour under the risk of ecological collapse. The natural next step is to extend the model to larger groups with $N > 2$ agents, given that real-world climate dilemmas typically involve many interacting actors. As discussed previously, varying parameters such as the enhancement factor (which modulates how favourable mutual cooperation is), the ecological collapse impact m, and the collapse leverage could be insightful. In the present work, partial observability was implemented as a deterministic reduction of the state space. We could extend this analysis to include noisy observations, allowing us to investigate the effects of graded uncertainty, similar to the approach of Mondal and Chakraborty [22]. Finally, collective action in response to environmental challenges is complicated by asymmetries among actors in their risk exposure, resource endowments, and influence over shared outcomes. Incorporating such inequalities into the present framework could provide insights into how these factors shape the emergence of cooperation.

Chapter 7

Conclusion

This thesis extends the Ecological Public Goods framework to investigate how information availability shapes cooperation in intertemporal collective-risk dilemmas. Using a partially observable, deterministic limit of temporal-difference learning, we simulated adaptive behaviour of agents under four informational regimes: complete, ecological-only, social-only, and none. We systematically examined the strategies that emerge, their stability, and the resulting levels of cooperation across these regimes.

For the conditions examined, we found that when agents are short-sighted, neither ecological nor social information alone can sustain cooperation, but their combination can. In contrast, when agents possess sufficient foresight, ecological cues alone are enough to maintain cooperation, while the addition of social feedback could lead to coordination failures.

Bibliography

[1] Stefano V Albrecht, Filippos Christianos, and Lukas Schäfer. Multi-Agent Reinforcement Learning: Foundations and Modern Approaches.

[2] David I. Armstrong McKay, Arie Staal, Jesse F. Abrams, Ricarda Winkelmann, Boris Sakschewski, Sina Loriani, Ingo Fetzer, Sarah E. Cornell, Johan Rockström, and Timothy M. Lenton. Exceeding 1.5 °c global warming could trigger multiple climate tipping points. *Science*, 377(6611):eabn7950, 2022. doi: 10.1126/science.abn7950.

[3] Robert Axelrod and William D. Hamilton. The evolution of cooperation. *Science*, 211 (4489):1390–1396, 1981. doi: 10.1126/science.7466396.

[4] Wolfram Barfuss and Richard P. Mann. Modeling the effects of environmental and perceptual uncertainty using deterministic reinforcement learning dynamics with partial observability. *Physical Review E*, 105(3):034409, March 2022. ISSN 2470-0045, 2470-0053. doi: 10.1103/PhysRevE.105.034409. URL `https://link.aps.org/doi/10.1103/PhysRevE.105.034409`.

[5] Wolfram Barfuss, Jonathan F. Donges, and Jürgen Kurths. Deterministic limit of temporal difference reinforcement learning for stochastic games. *Physical Review E*, 99(4): 043305, April 2019. ISSN 2470-0045, 2470-0053. doi: 10.1103/PhysRevE.99.043305. URL `https://link.aps.org/doi/10.1103/PhysRevE.99.043305`.

[6] Wolfram Barfuss, Jonathan F. Donges, Vítor 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):12915–12922, June 2020. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1916545117. URL `https://pnas.org/doi/full/10.1073/pnas.1916545117`.

[7] Wolfram Barfuss, Jessica Flack, Chaitanya S. Gokhale, Lewis Hammond, Christian Hilbe, Evan Hughes, Joel Z. Leibo, Tom Lenaerts, Nicky Leonard, Simon Levin, U. Madhushani Sehwag, Alex McAvoy, Jan M. Meylahn, and Fernando P. Santos. Collective cooperative intelligence. *Proceedings of the National Academy of Sciences*, 122(25), 2025. doi: 10.1073/pnas.2319948121. URL `https://doi.org/10.1073/pnas.2319948121`.

[8] Scott Barrett. *Environment and Statecraft: The Strategy of Environmental Treaty-Making*. Oxford University Press, 2003.

[9] Scott Barrett and Astrid Dannenberg. Sensitivity of collective action to uncertainty about climate tipping points. *Nature Climate Change*, 4(1):36–39, January 2014. ISSN 1758-678X, 1758-6798. doi: 10.1038/nclimate2059. URL `https://www.nature.com/articles/nclimate2059`.

[10] J. B. Coleman, M. Alfano, W. Barfuss, et al. Stewardship of global collective behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 118 (27):e2025764118, 2021. doi: 10.1073/pnas.2025764118. URL `https://doi.org/10.1073/pnas.2025764118`.

[11] Robyn M. Dawes. Social dilemmas. *Annual Review of Psychology*, 31:169–193, 1980. doi: 10.1146/annurev.ps.31.020180.001125.

[12] Aarti Gupta and Harro van Asselt. Transparency in multilateral climate politics: Furthering (or distracting from) accountability? *Regulation & Governance*, 13(3):18–34, 2017. doi: 10.1111/rego.12159.

[13] Costas Hadjiyiannis, Doruk İriş, and Chrysostomos Tabakis. International Environmental Cooperation under Fairness and Reciprocity. *The B.E. Journal of Economic Analysis & Policy*, 12(1), August 2012. ISSN 1935-1682. doi: 10.1515/1935-1682.2917. URL `https://www.degruyter.com/document/doi/10.1515/1935-1682.2917/html`.

[14] Garrett Hardin. The tragedy of the commons. *Science*, 162(3859):1243–1248, 1968. doi: 10.1126/science.162.3859.1243.

[15] IPCC. *Climate Change 2022: Mitigation of Climate Change*. Cambridge University Press, 2022. doi: 10.1017/9781009157926. URL `https://www.ipcc.ch/report/ar6/wg3/`.

[16] IPCC. Climate change 2023: Synthesis report of the ipcc sixth assessment report (ar6), 2023. URL `https://www.ipcc.ch/report/ar6-syr/`.

[17] Alan M. Jacobs. *Governing for the Long Term: Democracy and the Politics of Investment*. Cambridge University Press, 2011. ISBN 9780521173837. doi: 10.1017/CBO9780511921456.

[18] Andrew Kennard and Katharina Schnakenberg. From gridlock to ratchet: Conditional cooperation on climate change. *International Organization*, 79(2):257–280, 2025. doi: 10.1017/S0020818325000037.

[19] Maria Kleshnina, Christian Hilbe, Štěpán Šimsa, Krishnendu Chatterjee, and Martin A. Nowak. The effect of environmental information on evolution of cooperation in stochastic games. *Nature Communications*, 14(1):4153, July 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-39625-9. URL `https://www.nature.com/articles/s41467-023-39625-9`.

[20] Simon A. Levin. Public goods in relation to competition, cooperation, and spite. *Science*, 341(6144):122–123, 2013. doi: 10.1126/science.1236404.

[21] Manfred Milinski, Ralf D. Sommerfeld, Hans-Jürgen Krambeck, Floyd A. Reed, and Jochem Marotzke. The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proceedings of the National Academy of Sciences*, 105(7): 2291–2294, February 2008. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0709546105. URL `https://pnas.org/doi/full/10.1073/pnas.0709546105`.

[22] Samrat Sohel Mondal and Sagar Chakraborty. Noisy information channel mediated prevention of the tragedy of the commons, August 2024. URL `http://arxiv.org/abs/2408.08744`. arXiv:2408.08744 [q-bio].

[23] Martin A. Nowak. Five rules for the evolution of cooperation. *Science*, 314(5805): 1560–1563, 2006. doi: 10.1126/science.1133755.

[24] Martin A. Nowak and Karl Sigmund. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, 364(6432):56–58, 1993. doi: 10.1038/364056a0.

[25] Karine Nyborg. Reciprocal climate negotiators. *Journal of Environmental Economics and Management*, 92:707–725, 2018.

[26] Karine Nyborg, John M. Anderies, Astrid Dannenberg, et al. Social norms as solutions to collective action problems. *Science*, 354(6308):42–43, 2016. doi: 10.1126/science.aaf6632.

[27] John P. O'Doherty, Peter Dayan, Karl Friston, Hugo Critchley, and Raymond J. Dolan. Temporal difference models and reward-related learning in the human brain. *Nature Neuroscience*, 6(9):1253–1259, 2003.

[28] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. MIT Press, Cambridge, MA, 1994.

[29] Jorge M. Pacheco, Vítor V. Vasconcelos, and Francisco C. Santos. Climate change governance, cooperation and self-organization. *Physics of Life Reviews*, 11(4):573–586, 2014. doi: 10.1016/j.plrev.2014.02.003.

[30] Craig D. Parks, Xiaojing Xu, and Paul A. M. Van Lange. Does information about others' behavior undermine cooperation in social dilemmas? *Group Processes & Intergroup Relations*, 20(2):260–274, March 2017. ISSN 1368-4302, 1461-7188. doi: 10.1177/1368430215612220. URL `https://journals.sagepub.com/doi/10.1177/1368430215612220`.

[31] Katherine Richardson, Will Steffen, Timothy M. Lenton, Carl Folke, Johan Rockström, et al. Earth beyond six of nine planetary boundaries. *Science Advances*, 9 (37):eadh2458, 2023. doi: 10.1126/sciadv.adh2458. URL `https://doi.org/10.1126/sciadv.adh2458`.

[32] Johan Rockström, Will Steffen, Kevin Noone, et al. A safe operating space for humanity. *Nature*, 461(7263):472–475, 2009. doi: 10.1038/461472a.

[33] Caroline Schill, John M. Anderies, Therese Lindahl, et al. A more dynamic understanding of human behaviour for the anthropocene. *Nature Sustainability*, 2(12):1075–1082, 2019. doi: 10.1038/s41893-019-0419-7.

[34] Karl Sigmund. *The Calculus of Selfishness*. Princeton University Press, Princeton, NJ, 2010. ISBN 978-0-691-14275-2. URL `https://press.princeton.edu/books/paperback/9780691142752/the-calculus-of-selfishness`.

[35] Will Steffen, Johan Rockström, Katherine Richardson, Timothy M. Lenton, Carl Folke, Diana Liverman, Colin P. Summerhayes, Anthony D. Barnosky, Sarah E. Cornell, Michel

Crucifix, Jürgen F. Donges, Ingo Fetzer, Stuart J. Lade, Marten Scheffer, Ricarda Winkelmann, and Hans Joachim Schellnhuber. Trajectories of the earth system in the anthropocene. *Proceedings of the National Academy of Sciences*, 115(33):8252–8259, 2018. doi: 10.1073/pnas.1810141115.

[36] Richard S. Sutton and Andrew Barto. *Reinforcement learning: an introduction*. Adaptive computation and machine learning. The MIT Press, Cambridge, Massachusetts London, England, second edition edition, 2020. ISBN 978-0-262-03924-6.

[37] Alessandro Tavoni and Doruk İriş. Tipping and reference points in climate change games. In Graciela Chichilnisky and Armon Rezai, editors, *Handbook on the Economics of Climate Change*. Edward Elgar Publishing, June 2020. ISBN 978-0-85793-906-7 978-0-85793-905-0. doi: 10.4337/9780857939067.00018. URL `https://china.elgaronline.com/view/edcoll/9780857939050/9780857939050.00018.xml`.

[38] Sander van der Linden. The social-psychological determinants of climate change risk perceptions: Towards a comprehensive model. *Journal of Environmental Psychology*, 41:112–124, 2015. doi: 10.1016/j.jenvp.2014.11.012.

[39] Elke U. Weber. Experience-based and description-based perceptions of long-term risk: Why global warming does not scare us (yet). *Climatic Change*, 77(1-2):103–120, 2006. doi: 10.1007/s10584-006-9060-3.

[40] Åström, Karl Johan. Optimal Control of Markov Processes with Incomplete State Information I. 10:174–205, 1965. ISSN 0022-247X. doi: {10.1016/0022-247X(65)90154-X}. URL `https://lup.lub.lu.se/search/files/5323668/8867085.pdf`.

[41] Doruk İriş and Alessandro Tavoni. Tipping points and loss aversion in international environmental agreements. Technical Report Working Paper 239, Grantham Research Institute on Climate Change and the Environment, London School of Economics, 2016. URL `https://www.lse.ac.uk/granthaminstitute/publication/tipping-points-and-loss-aversion-in-international-environmental-agreements/`.