Journal of Artificial Intelligence Research 74 (2022) 733-764 Submitted 11/2021; published 06/2022

Cooperation and Learning Dynamics under Wealth Inequality and Diversity in Individual Risk Perception

Ramona Merhej RAMONA.MERHEJ@TECNICO.ULISBOA.PT
INESC-ID and Instituto Superior Técnico, Lisbon, Portugal
ISIR, CNRS, Sorbonne University, Paris, France

Fernando P. Santos F.P.SANTOS@UVA.NL
Informatics Institute
University of Amsterdam, The Netherlands

Francisco S. Melo FMELO@INESC-ID.PT
Francisco C. Santos FRANCISCOCSANTOS@TECNICO.ULISBOA.PT
INESC-ID and Instituto Superior Técnico
Universidade de Lisboa, Portugal

Abstract

We examine how wealth inequality and diversity in the perception of risk of a collective disaster impact cooperation levels in the context of a public goods game with uncertain and non-linear returns. In this game, individuals face a collective-risk dilemma where they may contribute or not to a common pool to reduce their chances of future losses. We draw our conclusions based on social simulations with populations of independent reinforcement learners with diverse levels of risk and wealth. We find that both wealth inequality and diversity in risk assessment can hinder cooperation and augment collective losses. Additionally, wealth inequality further exacerbates long term inequality, causing rich agents to become richer and poor agents to become poorer. On the other hand, diversity in risk only amplifies inequality when combined with bias in group assortment—i.e., high probability that agents from the same risk class play together. Our results also suggest that taking wealth inequality into account can help to design effective policies aiming at leveraging cooperation in large group sizes, a configuration where collective action is harder to achieve. Finally, we characterize the circumstances under which risk perception alignment is crucial and those under which reducing wealth inequality constitutes a deciding factor for collective welfare.

1. Introduction

Global risks are potential crises that threaten the well-being and the common welfare of the human race. They concern economic, environmental, geopolitical, societal or even technological risks (World Economic Forum, 2021). Examples include global warming, pandemics, large-scale conflicts, nuclear proliferation, etc. Although the risk categories are diverse, they all share a common property: no global risk can be prevented without substantial collective cooperation and coordination.

While the importance of avoiding global risks is generally not contested, history abounds with examples of failed risk avoidance and inefficacy in reaching collective action. Notably, the human race has faced two World Wars, several economic crises, and serious environmental damage. This failure happens because collective action poses serious cooperation and coordination challenges that are often difficult to overcome. First, averting global risks typically poses a social dilemma where the most favorable individual outcome happens when the risk is avoided through other agents' efforts and personal contribution is not necessary. Second, on top of a cooperation dilemma, collective risks entail coordination difficulties also known as the problem of many hands (PMH) as they require the involvement of a large number of players (Van de Poel et al., 2012). The intricate dynamics of the problem result in a non-trivial decision-making process, when agents, in the face of an uncertain disaster, are torn between the urge to cooperate to ensure risk avoidance and the selfish preference to defect (Axelrod, 1980; Axelrod & Hamilton, 1981; Kollock, 1998). Moreover, solving collective action problems has additional challenges related with multiple sources of heterogeneity across societies. In climate action negotiations, wealth inequality and contribution capacity of the involved countries, as well as their differences in assessing risk, makes reaching a unified target more difficult.

The social dilemmas involved in averting global collective risks are usually modeled by a threshold game with uncertain returns, the collective risk dilemma (CRD) (Milinski et al., 2008). In a CRD, agents decide how much of their wealth to contribute to a common pool in order to avoid the risk of a future disaster. The future disaster is only avoided with certainty if the agents manage to collect contributions above a given target threshold. If the target is not achieved, all agents are subject, with a given risk probability, to a disaster modeled by large losses in wealth, independently of whether they contributed or not.

To assess the challenges faced by populations in the context of CRDs, the decision-making dynamics of the individual agents in the game must also be captured. Studies on CRDs have therefore resorted to behavioral experiments (Milinski et al., 2008; Dannenberg et al., 2011; Tavoni et al., 2011; Milinski et al., 2011; Domingos et al., 2020; Milinski et al., 2008; Szekely et al., 2021), or modeled the decision-making process of agents using evolutionary game theory (Santos & Pacheco, 2011; Chen et al., 2012; Hilbe et al., 2013; Abou Chakra et al., 2018; Vasconcelos et al., 2014, 2013; Santos et al., 2019a, 2021) and, more recently, reinforcement learning (Domingos et al., 2021a; Merhej et al., 2021). Most of the studies, however, assume homogeneous populations of agents and hence only study the original cooperation and coordination challenges of the game without the additional challenges introduced by heterogeneity. Nonetheless, a few behavioral experiments and theoretical studies did look into the effects of wealth inequality in that context and concluded that it has a significant impact on the target achievement capability of a population (Baland & Platteau, 1997; Milinski et al., 2011; Hauser et al., 2019; Tavoni et al., 2011; Vasconcelos et al., 2014; Merhej et al., 2021). Yet, heterogeneity in risk assessment remains an unexplored topic in collective risks despite its prevalence in real life threats (World Economic Forum, 2021; Lee et al., 2015).

In this work, we investigate the challenges of collective action in populations facing collective risks and featuring a) wealth inequality, b) risk assessment diversity, and c) a combination of wealth inequality and risk assessment diversity. We model the long-term adaptation process of agents with multi-agent reinforcement learning.

We proceed with Section 2 on related work then detail the collective risk game and the agents' learning dynamics in Section 3. Following that, we explore our results in Section 4 and conclude our work in Section 5.

2. Related Work

We investigate the challenges posed by heterogeneity on the efficacy of a population of reinforcement learners in achieving collective action and preventing common risks. In this context, we are given the choice of modeling agents either as joint-action learners or as independent learners. The former requires agents to be aware of all other agents' actions as well as their learning dynamics. We believe this configuration to be unsuitable for modeling human learning in large populations where such knowledge is often not available. Moreover, the computational complexity of joint-action learning increases exponentially with the number of players, making it not only unsuitable for modeling human actors, but also impractical. On the other hand, independent learners need only be aware of their own actions and returns. The approaches to model independent reinforcement learners can further be divided between gradient-based independent learners and non-gradient based independent learners, as suggested in (Bloembergen et al., 2015). The former set of algorithms requires a well-defined differentiable objective function and knowledge of other agents' policies, an assumption we find inadequate when modeling human decision-making. We therefore resort to the latter class of independent learners, i.e., non-gradient based learners. Given the absence of additional constraints to single reinforcement learning, classical RL algorithms such as Q-learning (Watkins & Dayan, 1992), Policy Hill Climbing (PHC), or Hedge (Auer et al., 1995), can be directly adopted (Domingos et al., 2021a). Nonetheless, the literature abounds with RL algorithms designed specifically for the multi-agent setting. These include, but are not limited to, the Roth-Erev algorithm (Roth & Erev, 1995), WoLF-PHC (Bowling & Veloso, 2001), Frequency Adjusted Q-learning (Kaisers & Tuyls, 2010), Lenient Frequency Adjusted Q-learning (Bloembergen et al., 2010), Regret Matching (Hart & Mas-Colell, 2000), or Counterfactual Regret Minimization (Zinkevich et al., 2007). While most of these algorithms have been designed to reach specific equilibria, the Roth-Erev algorithm was designed and shown to effectively describe human behaviors in social dilemmas (Roth & Erev, 1995). The authors argue that this is a result of its capturing of two distinct characteristics of human learning: the *law of effect* that encourages previously successful actions (Thorndike, 1898) and the *power law of practice* that makes learning and adaptation slower with experience (Newell & Rosenbloom, 1981). Therefore, and while all aforementioned algorithms are viable candidates, we choose to model human decision making using the Roth-Erev algorithm with the goal of understanding the behavioral differences between homogeneous and heterogeneous populations in the face of collective risks, and how these behavioral differences echo on the overall well-being of the population. We note that the convergence points can possibly vary under different learning dynamics. However, this remains out of the scope of our paper and we refer the interested reader to a survey on the qualitative differences of various learning dynamics (Bloembergen et al., 2015).

While the general and basic modeling of collective risk dilemmas considers homogeneous and symmetrical agents, a few studies do investigate the consequences of wealth inequality among agents. When considering populations of reinforcement learners, wealth inequality is shown to hinder target achievement and entices rich agents to contribute more than poor agents (Merhej et al., 2021). On the other hand, when modeling agents with evolutionary game theory, wealth inequality is found to help in achieving cooperation if rich/poor individuals can imitate each other regardless their wealth category. Otherwise, if individuals preferentially imitate peers belonging to the same wealth class, inequality is detrimental for cooperation (Vasconcelos et al., 2014). Additionally, rich agents are found to contribute more than poor agents who only choose to cooperate if all rich players cooperate (Wang et al., 2010; Chan et al., 2008).

The data found in behavioral experiments does not always confirm the predictions under evolutionary game theory or reinforcement learning. Here, rich individuals are found to under-contribute while poor individuals over-contribute (Chan et al., 2008), and cooperation and collective success are harder to accomplish under wealth inequality (Tavoni et al., 2011).

While only wealth inequality has been studied in collective risks, other types of heterogeneity have been considered in other collective action dilemmas, notably in continuous public goods games. These are non-threshold games where agents need to join efforts to create a common good instead of avoiding a common disaster. In that context, and under evolutionary game theory, strong inequality in wealth, productivity and benefits are found to inhibit cooperation (Hauser et al., 2019).

Importantly, the effective occurrence of global risks in the future can only be approximately estimated, and not accurately measured. As a result, agents make their decision based on how likely they believe a disaster to occur. Diversity in global risk assessment is therefore quite common (World Economic Forum, 2021). A survey of 119 countries confirms significant variance in public concern and risk assessment of the global climate change problem (Lee et al., 2015). It finds that diversity in risk assessment mainly results from a diversity in education and fundamental understanding of the climate change problem between countries.

Similarly, the risks of the COVID-19 pandemic were assessed differently across countries that, as a consequence, adopted different safety measures (Alanezi et al., 2021; Gibney, 2020). Some perceived the virus propagation as more threatening than an economic shutdown and opted for a full lockdown early on, while others were more worried about unemployment, stagnation etc. and tried to protect the economy first. The diversity in COVID-19 risk perception is not only observed on a national scale but also on an individual scale (Lamarche, 2020). Such risk assessment diversity also translates into a behavioral diversity in safety measure compliance (Thanh et al., 2020).

Although risk perception diversity is a fundamental feature of our society, and the risk factor has been shown to have substantial impact on a population's ability to achieve the target (Milinski et al., 2008; Santos & Pacheco, 2011; Santos et al., 2012; Domingos et al., 2021a), to the best our knowledge, diversity in risk perception has not yet been explored in the context of collective risk dilemmas.

The objective of our work is to highlight the consequences of heterogeneity on the cooperative aptitude of a population facing collective risks. As such, we use RL as a means to model human actors. Nonetheless, we believe that the results we show can be valuable for research on cooperative capabilities in multi-agent reinforcement learning, a very active topic of research (Dafoe et al., 2020). We therefore present some established cooperation challenges in MARL and the respectively designed solutions.

While n-player non symmetrical social dilemmas and games with mixed-motives are abundant in the real world, cooperation in multi-agent reinforcement learning has mainly focused on 2-player games. A study on sequential social dilemmas with deep RL (Leibo et al., 2017), identified coordination sub-problems that prevent proper cooperation of agents.

Coordination problems in MARL are quite common and are not only restricted to social dilemmas (Matignon et al., 2012). One of the reasons for coordination difficulty in MARL is the non-stationarity of the opponent and the simultaneous policy updates of the players (Balduzzi et al., 2018; Bloembergen et al., 2015; Tuyls et al., 2006). Suggested solutions try to increase agents' *understanding* of the opponent's dynamics and leverage these to achieve higher cooperation. One algorithm proposes predicting the opponent's policy changes before computing the agent's policy gradient (Zhang & Lesser, 2010). Another alternative suggests differentiating through the variations of the opponent to actively shape their learning (Foerster et al., 2018a). A third solution incorporates both policy prediction and opponent shaping to increase stability while simultaneously escaping saddle points (Letcher et al., 2018).

Alternative solutions to increase cooperation in MARL focus on enabling *communication* capacities between agents. Communication can take several forms. For example, agents may communicate by sending messages (Foerster et al., 2016; Cao et al., 2018), sharing intentions (Kim et al., 2021) or experiences (Christianos et al., 2020), advising actions (Omidshafiei et al., 2019) to one another etc. Enriching agents with communication capabilities has shown to improve performance (Foerster et al., 2016; Christianos et al., 2020), speed up learning (Omidshafiei et al., 2019; Foerster et al., 2016; Christianos et al., 2020) and enhance coordination (Kim et al., 2021; Omidshafiei et al., 2019). Implementing a centralized critic with decentralized actors is another form of indirect communication and information sharing among agents that can increase performance and cooperation (Baker et al., 2019; Foerster et al., 2018b; Lowe et al., 2017). However, in cooperative environments, a centralized approach can learn inefficient policies with only one agent active and the other being "lazy". This happens as the second agent is discouraged from learning because its exploration would hinder the first agent's success and lead to worse team reward (Sunehag et al., 2017). Value decomposition addresses this problem by learning to decompose the team value function into per-agent value functions and thus transforms a complex learning problem into local, more readily learnable sub-problems (Sunehag et al., 2017; Son et al., 2019; Rashid et al., 2018, 2020; Marchesini & Farinelli, 2021).

A third set of solutions to leverage cooperation in RL introduce *conditional commitment* in agents' policies. One example is an algorithm designed to always asymptotically behave as a Tit-for-Tat strategy by learning simultaneously a cooperative and a selfish Q-function and alternating between them to avoid exploitability (Jacq et al., 2019).

Finally, solutions modifying agents' motivations can be seen as *institutional* solutions (Dafoe et al., 2020). Notably, in MARL, intrinsic rewards can be engineered and added to environmental rewards to help agents solve a sub-problem of the game and facilitate the emergence of coordination (Liu et al., 2019).

We note that the advised solutions for increasing cooperation in MARL settings focus on 2-player games. Major computational challenges still inhibit the scaling of these algorithms to n-player games. Additionally, most solutions are developed to increase cooperation in purely cooperative settings. We propose a non-symmetrical n-player social dilemma. We describe the emergent behaviors of simple reinforcement learners in these settings. In the context of reinforcement learning, the paper contributes, not by proposing novel algorithms for solving social dilemmas, but by identifying novel cooperation and heterogeneity challenges in large RL populations facing collective risks.

We examine the impact of heterogeneity in a collective risk dilemma representative of the climate action problem. As such, we first consider heterogeneity in terms of wealth inequality and introduce the notion of rich and poor agents. Then, we examine heterogeneity as a risk assessment diversity and introduce the notion of agents with high and low risk perception. Finally, we rely on the findings of Lee et al. (2015) on the correlation found between wealth and risk awareness. In fact, the authors show that public awareness of the risks associated with the climate change problem is low in poor or developing countries while it is high in rich or developed countries. As such, we combine the two types of heterogeneities in a population where rich agents have a high risk perception of the problem and poor agents have a low risk perception of the problem . We emphasize that, contrary to previous approaches (e.g., Santos & Pacheco, 2011; Domingos et al., 2021a), we explicitly differentiate between risk perception and the effective risk exposure, an element of particular importance whenever risk diversity is considered.

3. Methods

In this section, we describe the methods used for answering our research question: how do different sources of heterogeneity affect the behaviors of agents in populations facing collective risks, and how do such behavioral differences affect social welfare? We first define in Section 3.1 the dynamics of the collective risk dilemma as well as the design of the introduced heterogeneities. Then, in Section 3.2, we present the decision-making process or learning algorithm of the agents facing the CRD. Finally, in Section 3.3, we specify the values and hyper-parameters used to perform the computer simulations.

3.1 Game Definition

Formally, in a population of finite size Z, we allocate for every player an initial endowment b. Players are then sampled into groups of size N to play CRDs. Every player must choose to either contribute nothing or a fraction c of their endowment to a common pool. The benefits gained by investing in the common pool are modeled by the increased chances of avoiding a disaster (risk) happening with probability r. Should the players manage to jointly collect a sum greater than a target threshold $\mathbf{t}$, then the disaster is avoided with certainty. Otherwise, with a disaster probability r, all players lose a fraction p of whatever they have left of their initial endowments. At the end of the game, each player i, who started with an initial endowment b, is left with

$$b_i^{final} = \begin{cases} (1-c_i)b & \text{if the disaster was avoided,} \\ (1-c_i)b - p(1-c_i)b & \text{otherwise.} \end{cases} \tag{1}$$

where c_i represents the binary choice of either contributing 0 or a fraction c of the initial endowment to the pool ($c_i \in \{0, c\}$).

In game theory, normal-form games are usually defined by a payoff matrix that represents the benefits of a joint action for a given player. While the game risk r and disaster impact p determine objectively the resulting wealth of an agent, in real life, the (perceived) benefit, often called *utility*, of such an outcome is not necessarily equivalent or even linearly dependent on said outcome. A loss of \$1000 is not equally damaging to a millionaire as it is to an employee earning the minimum-wage salary. The log-utility function has been

Table 1: Payoff matrix of the game based on player's action and the outcome of the game.

Strategy	Disaster avoided	Disaster faced
C	$x_C = \log(1-c)$	$\bar{x}_C = \log(1-c-p-pc))$
D	$x_D = 0$	$\bar{x}_D = \log(1-p)$

used in economics to capture this diminishing marginal utility (Peters & Gell-Mann, 2016). It assumes that a loss of 70% of one's possessions, for example, is equally painful for any individual even if, with different initial wealth, the losses are not equal in absolute value. A large body of literature exists about utility functions and their representative meanings. It can be interesting to compare results obtained under different utility functions or even introduce heterogeneity between agents' perceived utility. For the moment, however, this remains out of the scope of our paper, and we consider a homogeneous log utility for all players. Under a log-utility function, the payoffs of the game are expressed as the difference in the log of agents' wealth before and after a game was played. Hence, a successful game costs $x_C = \log\left(\frac{b-cb}{b}\right) = \log(1-c)$ for a cooperator and $x_D = \log\left(\frac{b}{b}\right) = 0$ for a defector. Similarly, we can derive that a failure of avoiding the disaster costs $\bar{x}_C = \log(1-c-p(1-c))$ for cooperators and $\bar{x}_D = \log(1-p)$ for defectors. In the two (discrete) actions game, the goal of each player i is to find a stochastic policy π_i^*—representing the probability of player i choosing to cooperate—that maximizes the payoff. Table 1 summarizes the payoffs of the game for the two game outcomes of interest (i.e., disaster avoided or faced).

Finally, we recall that social dilemmas arise from a misalignment of individual and collective interests generated by specific tensions in the payoff function (Macy & Flache, 2002). More precisely, mutual cooperation should always be preferred over a unilateral cooperation and over a mutual defection. However, there should also always be either greed or fear that drives agents to defect to either exploit their peers or protect themselves from exploitation. In our game, a disaster is faced with probability r if the group fails to achieve the target threshold. Total cooperation always results in target achievement while total defection always results in failure of target achievement. As such, mutual cooperation yields a payoff $x_C = \log(1-c)$, whereas mutual defection yields with probability r a payoff $\bar{x}_D = \log(1-p)$ and with probability $1-r$, a payoff $x_D = 0$. To satisfy the conditions for a social dilemma to occur, $x_C > (1-r)x_D + r\bar{x}_D$ which implies that $r > \frac{\log(1-c)}{\log(1-p)}$. Additionally, the threshold $\mathbf{t}$ needs to be lower-bounded by the contribution of a single cooperative agent $\mathbf{t} > cb$, otherwise such a unilateral cooperation would also avoid the disaster and hence be as good as a mutual cooperation. Moreover, to incentivize agents to defect, the threshold needs to be achievable with less than a total cooperation $\mathbf{t} < Ncb$, otherwise, agents would have no motivation to free-ride.

3.1.1 Wealth Inequality

When investigating the impact of wealth inequality in our CRD setting, we consider a population of finite size Z of which a fraction $z_R = 20\%$ is rich and holds $w_R = 50\%$ of the total wealth W. The remaining fraction $z_P = 1 - z_R = 80\%$ of the population is poor and holds the remaining $w_P = 50\%$ of the riches. The total wealth held by the rich/poor is equally distributed within the same wealth class. Heterogeneity is introduced as inequality in initial endowments. Instead of an equal initial endowment b, poor players now start with an initial endowment $b_P = \frac{W \times w_P}{Z \times z_P}$ and correspondingly rich players start with an initial endowment $b_R = \frac{W \times w_R}{Z \times z_R}$ such that $b_R > b_P$. We continue to define the threshold $\mathbf{t}$ as a function of the average wealth $b = W/Z$ in the population.

3.1.2 Risk Perception Diversity

When investigating the impact of risk perception diversity in our CRD setting, we consider the same population of finite size Z but go back to a homogeneous wealth distribution of the total wealth W among the agents. That is, every agent starts with an equal initial endowment $b = W/Z$. However, half of the agents now have a low risk perception, i.e., they perceive a potential disaster as less likely to occur than it actually does, while the other half have a high risk perception, i.e., they perceive a potential disaster as more likely to occur than it actually does. For an effective disaster occurrence probability r, agents with low risk perception view the disaster happening with probability $r_L = r - \delta$, while agents with high risk perception view the disaster happening with probability $r_H = r + \delta$, where δ is a diversity factor. As a result, the population maintains an average risk perception equal to the effective risk value. To model risk perception diversity, we assume that during learning, agents with low risk perception face a disaster with probability r_L while agents with high risk perception face a disaster with probability r_H. However, when evaluating the effective consequences of the learned policies, we consider the common and effective risk value r.

3.1.3 Wealth Inequality and Risk Perception Diversity

We consider the population with wealth inequalities of Section 3.1.1 and assume that the poor agents have a low risk perception while rich agents have a high risk perception. That is, poor agents perceive a risk r to occur with probability $r_L = r - \delta$ while rich agents perceive it with probability $r_H = r + \delta$. Compared to populations with only risk inequality, this maintains a setting where $w_P = 50\%$ of the wealth in the population is managed by agents with low risk perception (here also poor) while the other $w_R = 50\%$ of the wealth is managed by agents with high risk perception (here also rich). The difference to populations with only risk inequality, is that 50% of the wealth is now held by 20% of the population instead of 50% of the population.

3.2 Agent Learning Algorithm

To study the dynamics of cooperation under reinforcement learning, we train a population of independent RL learners with the Roth-Erev Algorithm (Roth & Erev, 1995). By independent learners we mean that the RL agents do not model the presence of other players and perceive the emerging dynamics as part of their environment's dynamics. Prior to any interaction, every agent i has an initial propensity to cooperate or defect, determined by the values of a propensity vector $\mathbf{q}_{i,0} = \left[q_{i,0}(C), q_{i,0}(D)\right]^T$. The propensity vector is updated with every learning interaction. At the end of game k, agent i, according to the selected action A and the received return x, updates the propensity vector as

$$\begin{aligned} q_{i,k}(A) &= (1-\phi)q_{i,k-1}(A) + x, \\ q_{i,k}(\neg A) &= (1-\phi)q_{i,k-1}(\neg A), \end{aligned} \tag{2}$$

where $\neg A$ represents the non-chosen action and ϕ, $0 < \phi < 1$, is a forgetting parameter that inhibits the propensities from growing to infinity.

We train agents of a population asynchronously using the update rule (2). The procedure is summarized in Algorithm 1. At every learning step k, a group of N agents is selected randomly from the population of Z agents. The agents in the group engage in a collective risk dilemma. Every player i in the group chooses one of the available actions following probabilities $\mathbf{p}_{i,k-1}$ that are derived by normalizing the propensity vector $\mathbf{q}_{i,k-1}$. Since payoffs are negative or zero, we use the soft-max function to normalize the propensity vector. At any step k of the learning process, player i will select action A with probability

$$p_{i,k-1}(A) = \frac{\exp(q_{i,k-1}(A))}{\sum_{A' \in \{C,D\}} \exp(q_{i,k-1}(A'))}. \tag{3}$$

The selected actions and the game risk factor r determine whether or not the game is successful (i.e. if agents avoided the disaster). The payoffs for each agent are calculated according to Table 1 after which all agents in the group update their propensity vectors. This is repeated for a total of K learning steps. Since the algorithm does not guarantee that all agents are chosen equally as many times, we define K', the minimum number of learning steps that every agent needs to have performed before training is done. After K total learning steps, if some agent still has not performed at least K' updates, then training continues until this condition is satisfied.

Algorithm 1: Roth-Erev RL algorithm in an adaptive population with asynchronous updates of propensities.

Init: *K total number of learning steps, K' minimum number of updates per agent*
for $i \leftarrow 1$ **to** *Z, population size* **do**
  $\mathbf{q}_i(0) \leftarrow$*random initialization*;
  $u_i \leftarrow 0$ `/* tracks number of learning steps per agent */`
for $k \leftarrow 1$ **to** K **do**
  1. *sample random group G of size N*;
  2. *sample actions $A_i \sim \mathbf{p}_{i,k-1}$ for $i \in G$ (Eq. 3)*;
  3. *evaluate game success*;
  4. *calculate payoff of $i \in G$ (Tab. 1)*;
  5. *update $\mathbf{q}_{i,k}$ (Eq. 2)*;
  6. $u_i \leftarrow u_i + 1$ *for $i \in G$*;
  7. $u_{min} \leftarrow$ *min*$(\mathbf{u})$
while $u_{min} < K'$ **do**
  repeat steps 1. to 7.

3.3 Computer Simulation

In this section we present the numerical values that define the CRD, as well as the learning algorithm hyper-parameters.

In all settings, we consider a population of $Z = 200$ individuals. The average wealth in the population is set to $b = 1$ yielding $W = Z$. A contribution represents 10% of an agent's wealth, i.e., $c = 0.1$. We define the threshold $\mathbf{t}$ as a function of the average wealth b in the population. We set the target to be achievable if at least $M = N/2$ agents in the group contribute, i.e., $\mathbf{t} = Mcb = Ncb/2$. If the threshold target is not achieved, agents lose an additional 70% of their remaining wealth, i.e., $p = 0.7$. We test varying risk values $r \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, varying group sizes $N \in \{2, 4, 6, 8, 10, 20\}$ and varying risk perception diversity factors $\delta \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$.

We sample $q_{i,0}(A)$ from a normal distribution $\mathcal{N}(\mu = -10, \sigma = 1)$. This generates players with a slight random initial preference to defect or cooperate such that $\log_e\left(\frac{q_{i,0}(C)}{q_{i,0}(D)}\right) \sim \mathcal{N}(\mu' = 0, \sigma' = \frac{2\sigma}{\mu})$, according to the log domain transformation of Katz (Katz et al., 1978). The forgetting parameter is set to $\phi = 0.001$.

4. Results

In this section, we provide three sets of experiments that investigate how different systemic inequality levels between independent RL agents affect cooperation levels in the population and how these behavioral differences influence social welfare in the system.

- In Section 4.1 we start by analyzing how wealth inequalities impact the success of a population in a CRD according to three criteria: 1) the probability of achieving the target threshold, 2) the collected contributions, and 3) the remaining welfare. Departing from previous work, here we also study explicitly the effects of varying groups sizes.
- Section 4.2 conducts a similar analysis, but now assessing the impact of the diversity in risk assessment along the same criteria, together with the impact of group assortment based on risk assessment.
- Finally, Section 4.3 considers the combined impact of wealth inequality and risk perception diversity.

In all sections, the evaluation proceeds by allowing the agents to train for a total of 2.5×10^5 learning steps, while imposing a minimum number of $K' = 3 \times 10^4$ learning steps for every agent. The values reported in the three criteria correspond to the values observed at the end of the training period, averaged over 5 independent runs.

The three metrics considered provide a different perspective on the "success" of the population: the collected contributions are a measure of cooperation in the population, while target achievement and remaining welfare are variables we use to quantify the well-being of the system.

As such, we define η, the percentage of groups in the population that achieve the target, ρ, the ratio between the collected contributions and the maximum possible total contributions, and ζ, the ratio between the remaining wealth (after contribution costs and disaster losses) and the initial wealth. After population evaluation, we evaluate the different classes of agents (rich/poor, high/low risk perception etc.) on similar criteria.

We look at the average group achievement η, the probability of cooperating, i.e., the policy π, and the percentage of secured welfare ζ. While π is directly obtained at the end of the learning procedure, the achieved contributions ρ, the group achievement η, and the remaining wealth ζ are not.

We can compute ρ as

$$\rho = \frac{\pi_1 Z_1 b_1 + \pi_2 Z_2 b_2}{Z_1 b_1 + Z_2 b_2} \tag{4}$$

where indices 1 and 2 represent the binary classes in the heterogeneous population, e.g., class 1 corresponds to poor/low-risk and class 2 corresponds to rich/high-risk; π_i, Z_i and b_i denote, respectively, the average cooperation probabilities, the number of agents, and the initial wealth of the agents in class i, $i = \{1, 2\}$. The numerator computes the collected contributions while the denominator computes the maximum possible contributions in the case of total cooperation, i.e., $\pi_1 = \pi_2 = 1$. For homogeneous populations, $\rho = \pi$. To compute η and ζ, for every setting, we rollout a game where the population is split into groups of N players. In each group, agents, following their learned policies, choose to either contribute or not. At the end of the game, we calculate the percentage of groups, as well as the percentage of rich/poor and high/low risk agents in the population that successfully reach the target. We also evaluate the remaining wealth of the population and of the different classes of agents after cooperation costs and disaster losses. The random variables are evaluated and averaged over 10^5 simulations.

We run our simulations on a homogeneous population P_0 as a baseline, and its heterogeneous counterparts P_1, P_2 and P_3—representing respectively wealth inequality, risk perception diversity and a combination of wealth and risk diversity.

4.1 Effect of Wealth Inequalities

To understand the consequences of wealth inequality in the context of a collective risk dilemma, we compare key performance metrics of populations with and without wealth inequality, i.e., P_0 and P_1, across different game settings of either varying risk factors r or varying group sizes N.

4.1.1 Wealth inequality and probability of risk occurrence

In a first experiment, we train the homogeneous population P_0 and the population with wealth inequality P_1 to play a collective risk dilemma in groups of $N = 6$ agents and under varying risk factors $r \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. Figure 1 shows the group achievement rates η, the achieved contributions ρ, and the secured welfare ζ for the two populations. We also explicit η, π and ζ for the different classes of agents within the populations.

First, as expected, Figure 1a shows that the performance of a population with and without inequalities increases with the risk factor r. Agents feel a stronger urge to achieve the target if the consequences of failure are larger. This result is also in accordance with those found under social learning rules (Vasconcelos et al., 2014). However, while under evolutionary dynamics and social learning, diversity has proven to increase cooperation rates (Santos et al., 2008; Vasconcelos et al., 2014), we find in Figure 1a that wealth inequality

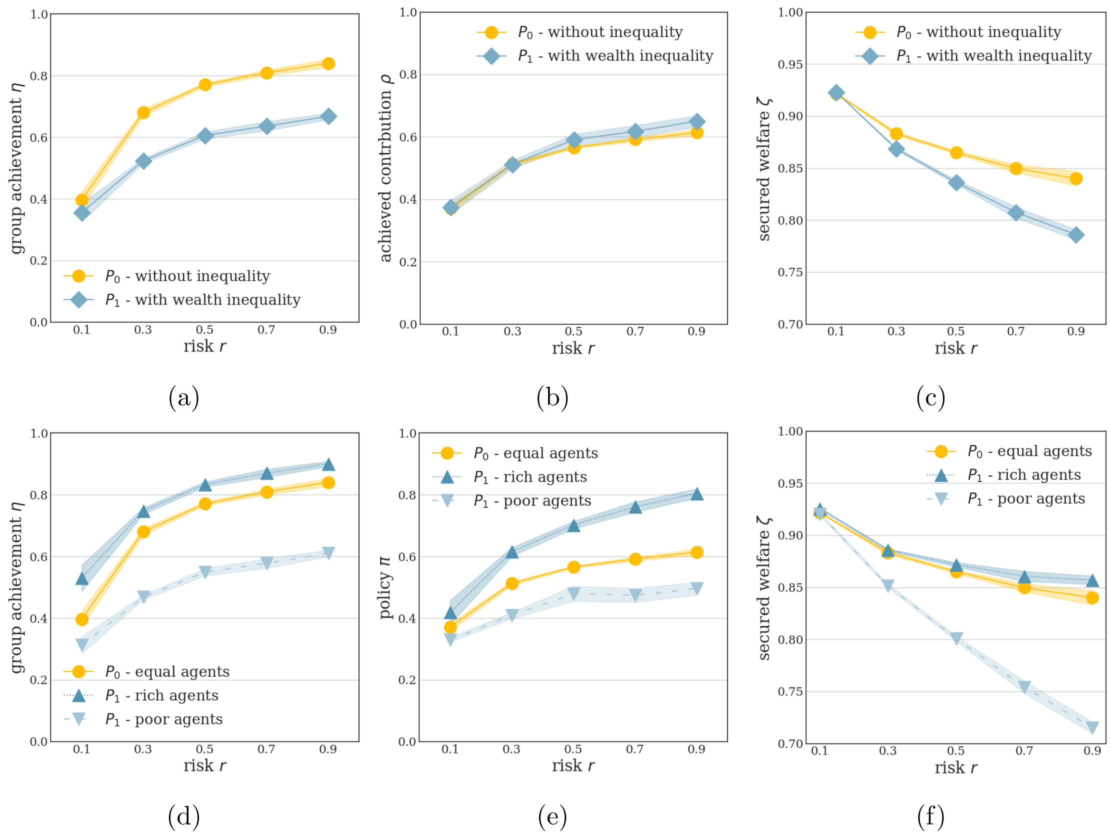


Figure 1: **Impact of risk under wealth inequality:** In the first row, we compare populations without inequality to populations with wealth inequality. We plot (a) the overall group achievement rate η, (b) the percentage of achieved contributions ρ, and (c) the percentage of remaining wealth in the populations ζ. In the second row, we take a closer look at the internal dynamics in the populations and compare results for rich, poor and equal agents (i.e., agents from the homogeneous populations). Again, we plot (d) the overall group achievement rate η, (e) the policy π or the average cooperation probability, and (f) the percentage of remaining wealth ζ for each class of agents. In all panels, shaded areas represent the standard deviation over 5 runs.

decreases the overall achievement of a population for all risk values (P_1 under-achieves with respect to P_0). This is in conformity with some experimental results (Tavoni et al., 2011), where wealth inequalities are found to inhibit group achievement. This result also reinforces the different types of dynamics created by individual-based learning (as in here) when compared to decision-making coupled with dynamics of peer-influence (Vasconcelos et al., 2014), absent in our model.

Second, Figure 1b reveals that the drop in overall achievement rate for population P_1 compared to P_0 is not caused by a reduction in total collected contributions. Populations with wealth inequality collect similar pools of contributions as their homogeneous counterparts but can fail up to two times more often in achieving the target.

Third, we note that while higher risk results in higher group achievement, this does not translate into higher secured wealth. In fact, Figure 1c shows that the secured wealth in the population continuously decreases with the risk. While populations adapt to higher risks by increasing their contributions and target achievement, they do not do that at a rate that effectively protects them from the ever more probable disasters. When comparing homogeneous and heterogeneous populations, we observe that the populations with wealth inequality, as a result of a decrease in group achievement, also lose a larger amount of their wealth.

To better understand how the introduction of wealth inequality can reduce group achievement and secured welfare despite consistent high contributions, we consider in Figure 1d the discrepancies in target achievement for rich vs. poor agents. While rich agents achieve the target more often than agents from homogeneous populations, poor agents that represent a majority of 80% of the population, achieve the target significantly less often than both equal or rich agents which explains the observed overall drop in Figure 1a.

Next, we examine the policies or average cooperation probability of rich, poor and homogeneous agents. We observe in Figure 1e that in populations with wealth inequality, the main cooperators are rich agents. Poor agents cooperate significantly less than rich agents and this gap in cooperation increases with the risk. We confirm that these behaviors also persist under a linear-utility function, suggesting that they are not specific to the chosen log-utility. While rich agents continue to adapt to higher risk values by increasing their contributions, poor agents are less reactive to the risk (flatter curve) and seem to stagnate for risk values greater than 0.5. This has detrimental effects on poor agents and the population as a whole for two main reasons. First, the large number of poor agents in the population (80%) makes it hard for all poor agents to interact in groups with rich agents, and hence benefit from their cooperation and increased group success. In fact, for $N = 6$, 25% of the groups in the population are purely poor groups. Second, the small share that a poor agent's contribution represents in terms of the target threshold, introduces higher coordination problems for this class of agents. While a contribution by a rich agent represents 83% of the needed target, a contribution of a poor agent only represents 20% of that same target. As such, the same collective dilemma requires higher coordination from poor agents, e.g., 5 out of 6 poor players need to cooperate in a purely poor group to achieve the target. With the learned cooperation rates of poor agents, for $r = 0.3$ for example, 95.5% of purely poor groups fail to reach the threshold. These are failures only suffered by poor agents and explain the gap in achievement observed in Figure 1d. In terms of population achievement, a 95.5% failure of 25% of all groups implies an overall achievement drop of 24% and explains the drop introduced by wealth inequality in Figure 1a.

Finally, Figure 1f shows that the achievement inequality introduced by wealth inequality generates further wealth inequality at higher risk values. While the gap in target achievement is relatively constant in Figure 1d, the difference in wealth losses is minimal for small risk values and increases for larger ones. At high risk, the consequences of target achievement failure are more prominent. We observe that poor agents lose significantly more than rich agents and deduce that the observed additional losses in population welfare introduced by wealth inequality in Figure 1c, are only incurred by the poor agents. In terms of wealth distribution, this causes poor agents to become poorer and rich agents to become richer.

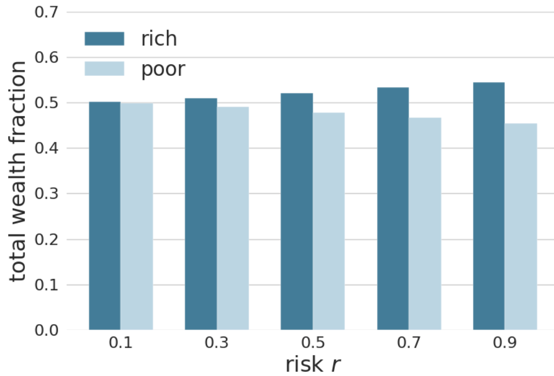


Figure 2: Resulting population wealth distribution among rich and poor agents after engaging in a collective risk dilemma with varying risk values. Recall that before any contribution costs and disaster losses, rich and poor agents each hold 50% of the overall population's wealth. As the risk of disaster occurrence increases, the consequences of differences in target achievement between the classes are magnified. Poor agents lose more than rich agents (see Figure 1f) and own even less of the resulting remaining wealth in the population. Rich agents become relatively richer and poor agents become relatively poorer.

Figure 2 illustrates this phenomena by plotting the distribution of wealth among rich and poor agents after engaging in collective risk dilemmas of varying risk values.

To conclude, our first findings suggest that wealth inequality decreases overall population achievement and secured wealth. However, not all agents suffer equally from these consequences. In fact, compared to homogeneous populations, wealth inequality slightly increases target achievement and welfare of rich agents while significantly decreasing achievement and welfare of poor agents. Wealth inequality is an unstable state that results in increasingly higher inequalities.

4.1.2 Wealth Inequality and Group Size

Group size has a considerable effect on overall population achievement under evolutionary dynamics (Szolnoki & Perc, 2011; Hauert et al., 2006; Kurokawa & Ihara, 2009; Peña & Nöldeke, 2018). Additionally, we demonstrated under reinforcement dynamics how a group size of $N = 6$, combined with wealth inequality, can increase coordination difficulties for poor agents and lead to a decrease in their wealth and target achievement.

In a second experiment, we train the homogeneous population P_0 and the population with wealth inequality P_1 to play a collective risk dilemma of risk $r = 0.3$ in varying group sizes of $N \in \{2, 4, 6, 8, 10, 20\}$. Figure 3 shows the group achievement, the total contributions and the wealth when engaging in different group sizes for the two populations and their corresponding classes of agents (equal and rich/poor) in a collective risk dilemma of average risk $r = 0.3$.

Figure 3a shows that for populations without inequality, larger group sizes imply an increased coordination difficulty and result in lower group achievements. In fact, group achievement of homogeneous populations decreases from 85% for $N = 2$ to around 40% for $N = 20$. This clear negative correlation between target achievement and group size is not observed for populations with wealth inequality. For the tested values, populations with wealth inequality are relatively robust to varying group sizes and always reach values of $\eta \simeq 50\%$. As such, for smaller group sizes with less coordination difficulties, homogeneous populations outperform populations with wealth inequality. However, as their performance continues to decline, and that of populations with wealth inequality remains stable, we observe a cross-point after which populations with inequality outperform their homogeneous counterparts.

Similar to results observed under varying risk values, Figure 3b confirms that the discrepancies in target achievement between the two populations are not a result of differences in contributions. Both populations collect similar pool contributions.

In Figure 3c we observe that the qualitative changes of the target achievement with respect to the group size are mimicked in the wealth. This was not the case for varying risk values (Figures 1a and 1c) where higher group achievements for larger risk values still resulted in larger losses. For a constant risk value, target achievement is highly correlated with the secured wealth.

When looking at the distributions of target achievement, cooperation and welfare between the different classes of agents, we find again in Figure 3d that poor agents under-achieve with respect to rich agents. However the difference in achievement decreases with larger group sizes. Larger group sizes result in a better mixing of rich and poor players as the probability of sampling purely poor or purely rich groups decreases. In the extreme case, when the group size is equal to the population size, the group achievements of rich and poor agents become the same. We conclude that in populations with wealth inequality, larger group sizes can decrease achievement inequalities without decreasing overall population achievement.

From Figure 3e we understand why populations with wealth inequality under-achieve with respect to homogeneous populations despite equal total contributions. While rich and poor agents each hold 50% of the wealth in the population, most of the contributions in the population come from rich agents. Rich agents cooperate more than equal agents while poor agents cooperate less than equal agents. The variations in cooperation are almost symmetrical among the two classes and result in population P_1 collecting as many contributions as population P_0. However, the rich agents who succeed more often, represent only 20% of the population, while the poor agents who fail more often represent 80% of the population. The symmetrical gains and losses in the contributions are translated into asymmetrical gains and losses in target achievement. However, this effect is less prominent for larger group sizes with higher mixing of rich and poor agents.

Finally, in Figure 3f we observe again, that the qualitative changes in target achievement are mirrored in wealth changes. We note that wealth changes are smoother than target achievement changes. This is because not meeting the target only results in disaster losses with a risk probability $r = 30\%$. A larger risk value would have caused more pronounced welfare losses for a similar target achievement rate.

Overall, Figure 3 reveals the qualitative difference in the impact that the group size has on populations with and without inequality. Yet, it does not explain the reason for this observed difference. While the risk factor increased group achievement for both populations, the group size induces two very different dynamics. It decreases achievement for homogeneous populations but has an ambiguous and non-monotonous impact on heterogeneous

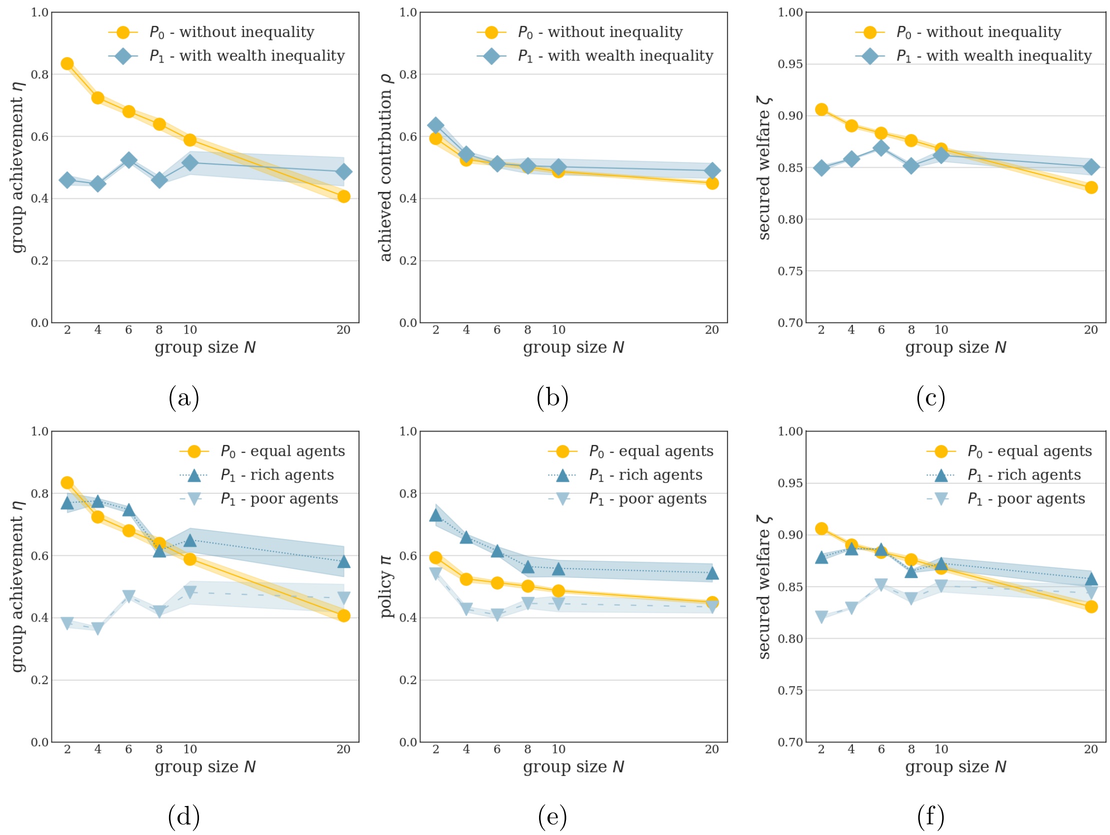


Figure 3: **Impact of group size under wealth inequality:** In the first row, we compare population P_0 without inequality to population P_1 with wealth inequality. We plot (a) the overall group achievement rate η, (b) the percentage of achieved contributions ρ, and (c) the percentage of remaining wealth in the populations ζ. In the second row, we take a closer look at the internal dynamics in the populations and compare results for rich, poor and equal agents (i.e., agents from the homogeneous populations). Again we plot (d) the overall group achievement rate η, (e) the policy π or the average cooperation probability, and (f) the percentage of remaining wealth ζ of each class of agents. In all panels, shaded areas represent the standard deviation over 5 runs.

populations. We investigate how introducing wealth inequalities can diminish the effect of the group size.

For $N = 6$, we saw that groups of purely poor agents that represented 25% of the groups in the populations, almost never achieve the target, causing a significant drop in overall achievement. To capture how group size modifies global target achievement, we present in Figure 4 a bar plot for $N = 2$, $N = 6$ and $N = 10$, showing for each group size, the different possible group configurations and their respective target achievement probability. For a group of size of N, we have $N + 1$ different configurations representing respectively groups with $n_R \in \{0, 1, \ldots N\}$ rich agents and $n_P \in \{N, N-1, \ldots 0\}$ poor agents. The first bar always represents purely poor groups. The width of the bars are proportional to the probability of said group in the population, while the height represents the probability of

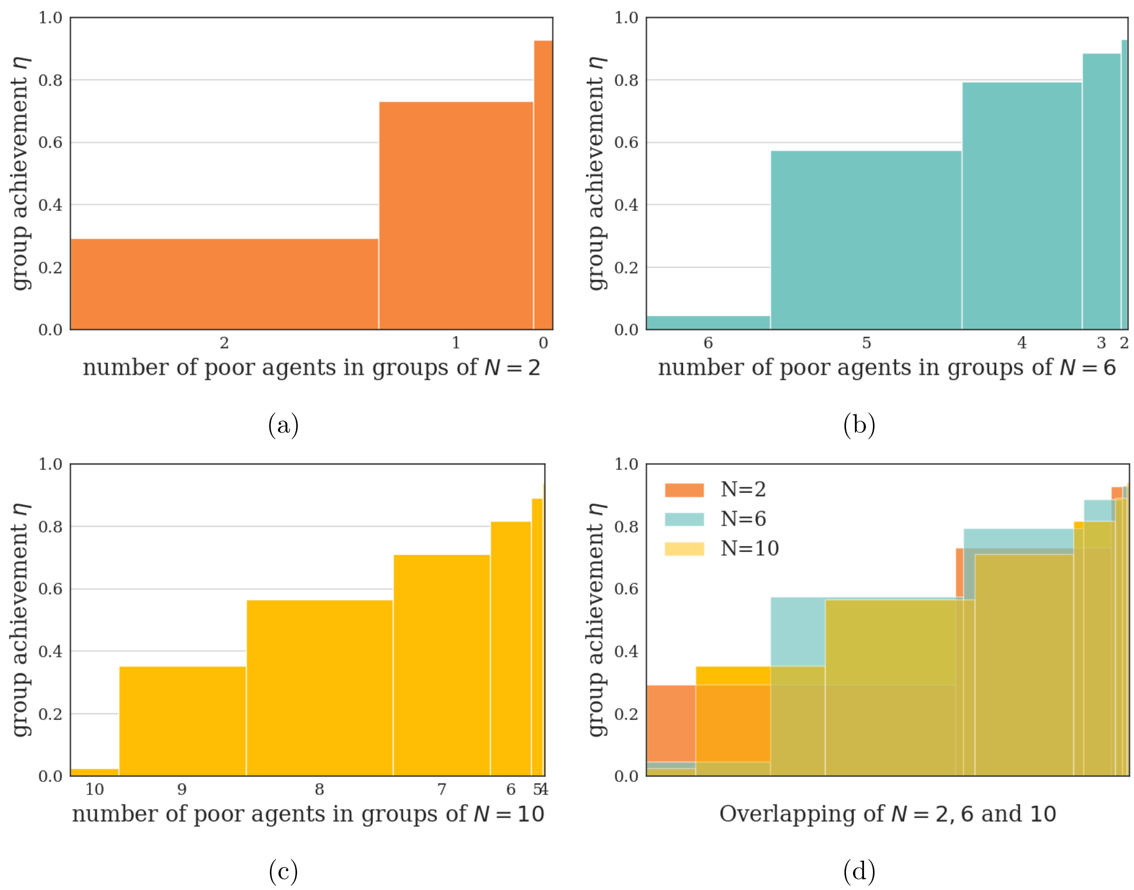


Figure 4: **Group achievement distribution:** We explicit for different group sizes of (a) $N = 2$, (b) $N = 6$, and (c) $N = 10$, the $N + 1$ possible group configurations with $N, N - 1, \ldots 0$ poor agents and their corresponding group achievements. The first bars always represent purely poor groups and are the least performing groups in the population. The width of the bars are proportional to the frequency of the groups in the population. As such, the area of the figure covered by the bar plots is proportional to the overall population achievement. For $N = 2$, more than 60% of the groups are purely poor groups. For larger group sizes, the mixing of rich and poor agents is more probable. Low performing purely poor groups become less frequent. However, larger group sizes also increase coordination difficulty. This is best seen in purely poor groups whose performance drops from 30% for $N = 2$ to 2% for $N = 10$. In (d) we overlap the three group size configurations and observe that they all cover around half of the figure area which explains the relatively stagnant group achievement of around 50% with respect to N observed in Figure 3a.

target achievement. For group sizes of $N = 6$ and $N = 10$, less than $N + 1$ bars are plotted. This is because the corresponding groups are almost non existent in the population. The total area of the figure is equal to 1 and the fraction occupied by the bar plot represents the total population target achievement.

Looking at the first bar in each plot, we see that for $N = 2$, groups of purely poor agents achieve the target with probability $\eta = 30\%$. As N increases, coordination problems increase and purely poor groups achieve the target less often. However, we also observe that the frequency of low performing groups (the width of the bar) decreases with N. As N increases, opposite dynamics seem to be at play. On one hand a better mixing of poor and rich agents decreases the probability of highly unsuccessful poor groups, and on the other hand, increasing coordination difficulties make target achievement more difficult. However, neither one these effects strictly outweighs the other. As such, when superposing the different group configurations in Figure 4d, we see that despite not fully overlapping, all three settings cover around 50% of the figure area. We deduce that the variations of the population achievement are neither strictly increasing nor strictly decreasing with the group size. Different group sizes can be either a little more or a little less advantageous for the population. This explains the small fluctuations observed around $\eta = 50\%$ in Figure 3a.

Our results highlight how, in large groups, wealth inequality can help escape strong coordination difficulties and improve collective target achievement by allocating more control to a few rich agents. On the other hand, when action is taken in smaller groups, wealth inequality can exclude the majority of poor agents from collective success and result in an overall decline in achievement. We note that the best way of solving collective risk dilemmas, is to work in small groups of homogeneous agents (P_0 at $N = 2$).

4.2 Effect of Risk Perception Diversity

In the previous section we have looked at the consequences of wealth inequality on populations involved in collective risk dilemmas of varying risk probability and group sizes. Here we investigate the consequences of introducing a symmetrical risk perception diversity in the population. We present a study comparing the effects of a constant diversity for different risk values r, and a study comparing small and large diversities in perception δ in a collective dilemma with constant average risk.

4.2.1 Risk Perception Diversity and Probability of Risk Occurrence

In a first experiment, we train a population P_2 with risk perception diversity $\delta = 0.1$ to play a collective risk dilemma in groups of $N = 6$ players and under varying risk factors $r \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. During the learning phase, half of the agents in P_2 perceive a risk of $r_L = r - \delta$, while the other half perceives a risk of $r_H = r + \delta$. We compare the performance of P_2 to the performance of our baseline homogeneous population P_0. In Figure 5, we plot achievement rates, contribution probabilities and the secured wealth of the two populations and their corresponding classes of agents (correct and high/low risk perception).

Figures 5a, 5b and 5c show no significant differences in target achievement, contributions or secured wealth for homogeneous populations versus populations with risk diversity for all values of $r \geq 0.3$. For $r = 0.1$, populations with risk diversity contribute less and hence achieve the target less often. This difference however does not seem to have an impact on the overall secured wealth. The increased costs of disaster occurrence are compensated by savings on cooperation costs.

In the second row of Figure 5, we look at the variations between the different classes of agents. In Figure 5d, we only observe minor differences between agents at high and low risk for $r = 0.1$ that completely disappear for higher risk.

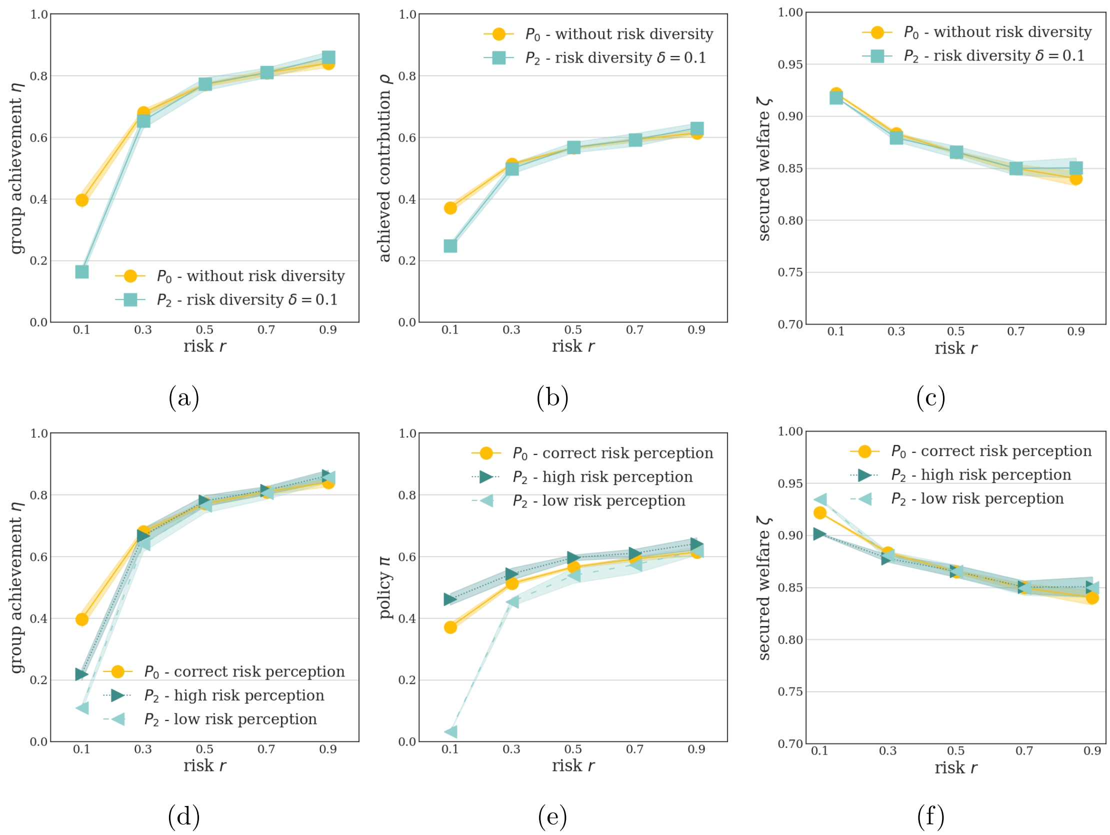


Figure 5: **Impact of risk in the presence of a diversity in risk perception:** In the first row, we compare populations without diversity to populations with symmetrical risk diversity of $\delta = 0.1$. We plot (a) the overall group achievement rate η, (b) the percentage of achieved contributions ρ, and (c) the percentage of remaining wealth in the populations. In the second row, we look at the same metrics for the agents with correct risk perception from population P_0, as well as agent with high and low risk perception from population P_2. Again we plot (d) the overall group achievement rate η, (e) the policy π or the average cooperation probability, and (f) the percentage of remaining wealth for each class of agents. In all panels, shaded areas represent the standard deviation over 5 runs.

Figure 5e shows that the similar group achievements of agents with high and low risk perceptions occur despite some differences in cooperation. While agents with high risk perception cooperate more than agents with low risk perception, the induced differences in target achievement between the two classes are negligible. This remains true for $r = 0.1$, where agents with high risk perception cooperate around 45% of the time while agents at low risk only cooperate 3% of the time. The large gap in cooperation only results in around 10% of difference in target achievement.

Finally, Figure 5f shows that both classes of agents secure similar wealth for $r \geq 0.3$. However, for $r = 0.1$, the costs of cooperation of agents with high risk perception outweigh the benefits of their slightly higher target achievement. Here we observe how agents with low risk perception manage to secure a higher fraction of their wealth than agents with a high risk perception.

We can conclude that a small risk perception diversity is easily absorbed by the population and, unlike wealth inequality, does not generate further inequalities. This is all the more so for larger values of baseline risk, as the relative impact of the diversity in perception δ/r becomes imperceptible.

4.2.2 Variations of Risk Perception Diversity

While the tested risk perception diversity of $\delta = 0.1$ was effectively absorbed by the population P_2 for most baseline risk values, we saw that for smaller risk, the impact of the diversity was more important. In a second experiment we propose to fix the baseline risk value to $r = 0.5$, and test the impact of increasing diversity factors δ. We plot in Figure 6, the group achievement, contributions and secured welfare of populations and their corresponding classes of agents with varying risk perception diversities $\delta \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$. We recall that $\delta = 0$ represents the homogeneous population P_0, while $\delta = 0.1$ represents the population P_2 from Section 4.2.1.

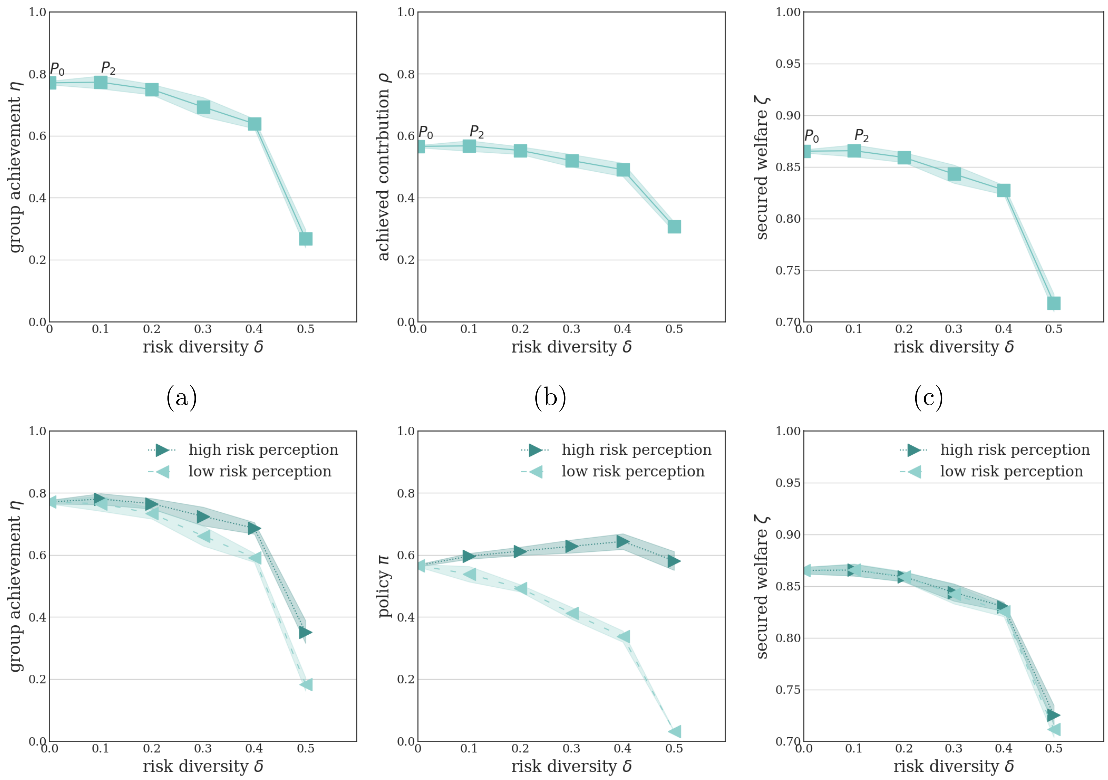
In the first row of Figure 6, we see that higher risk diversity results in a decline in target achievement, collected contributions as well as secured welfare. This is so despite a constant average risk value.

Looking at the second row, we observe a decline in target achievement for both classes of agents (Figure 6d), and agents with high risk perception slightly over-achieving in comparison to agents with low risk perception. This is so despite stronger differences in cooperation willingness between the two classes as δ increases (Figure 6e). We also observe similar secured welfare for both classes of agents (Figure 6f). The increased cooperation costs of agents with high risk perception are compensated by their smaller likelihood to face a disaster. The opposite is true for agents with low risk perception. The increased losses of disaster occurrence are compensated by decreased cooperation costs.

We conclude by comparing the effects of risk diversity to the effects of wealth inequality. While risk diversity and wealth inequality both reduce target achievement, we reconfirm that risk diversity, unlike wealth inequality, does not introduce large discrepancies in target achievement and secured welfare between the two classes of agents. We hypothesize that the $80 - 50$ wealth distribution is the main cause for the observed divergence in secured welfare and target achievement for rich and poor agents. This is because for $N = 6$, it does not allow a well mixing of rich and poor agents and a large number of poor players end up in purely poor groups with low target achievements. In the case of a symmetrical risk diversity with 50% of agents having a high and 50% of agents having a low risk perception, the probability of sampling a group of only agents with low risk perception is less than 2%. The better mixing of agents from the two classes, reduces the observed differences in target achievement and welfare.

4.2.3 Effect of Assortment Bias with Risk Perception Diversity

To validate our hypothesis, we introduce an assortment bias that privileges agents from a same risk perception class to interact in a same group. Recall that without any sampling bias, each class of agents represents 50% of the population. We assume that agents with a



(a) (b) (c)

(d) (e) (f)

Figure 6: **Impact of risk diversity intensity:** In the first row, we look at the overall population performance when engaged in groups of $N = 6$ in a collective risk dilemma of $r = 0.5$. We plot (a) the overall group achievement rate η, (b) the percentage of achieved contributions ρ, and (c) the percentage of remaining wealth in the populations. In the second row, we take a closer look at these metrics and compare results for agents with high and low risk perception. Again we plot (d) the overall group achievement rate η, (e) the policy π or the average cooperation probability, and (f) the percentage of the remaining wealth of agents. In all panels, shaded areas represent the standard deviation over 5 runs.

sampling bias perceive other agents of the same class as more frequent than they actually are. We denote this perceived frequency as α where $0.5 \leq \alpha \leq 1$. We use the perceived frequency α to compute a bias weight β for agents of a same class, and $\bar{\beta}$ for agents of opposite classes. The bias weight represents the ratio between the perceived frequency of a class in a population and its effective weight in the population (here 50%). We have

$$\beta = \frac{\text{perceived weight of same class in the population}}{\text{effective weight of same class in the population}} = \frac{\alpha}{0.5} \tag{5}$$

and respectively

$$\bar{\beta} = \frac{\text{perceived weight of opp class in the population}}{\text{effective weight of opp class in the population}} = \frac{1-\alpha}{0.5}. \tag{6}$$

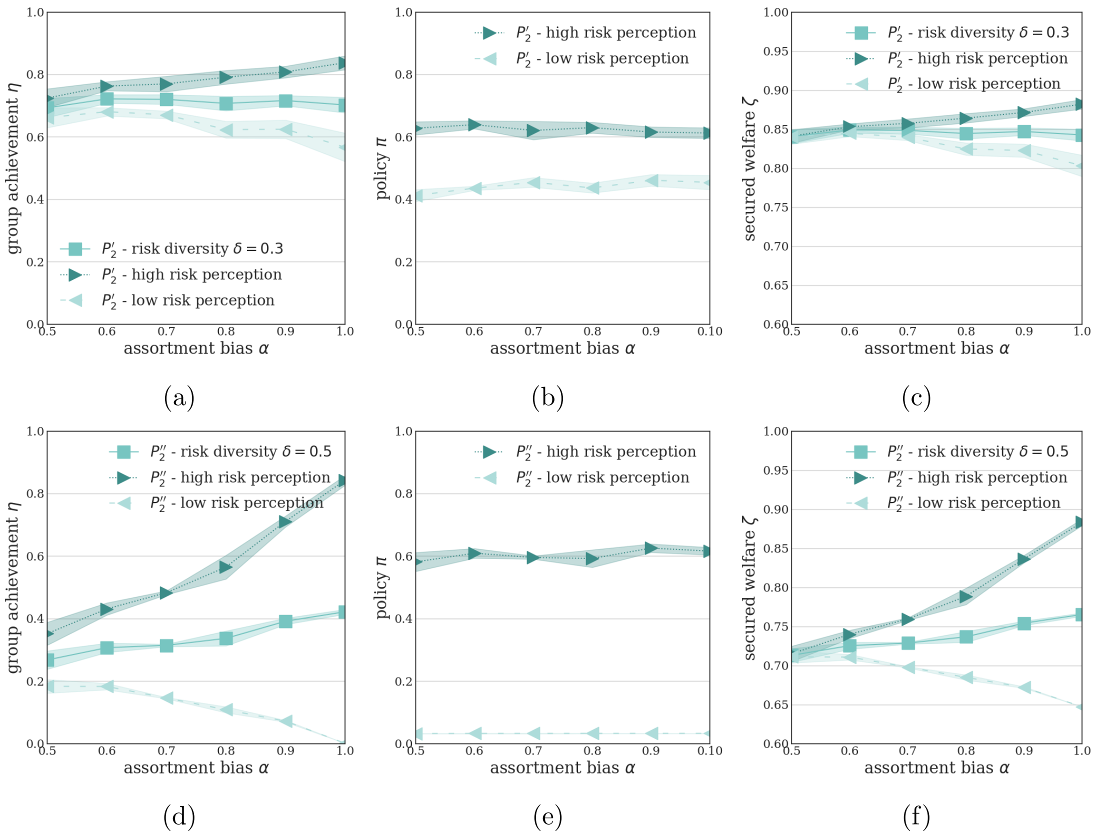


Figure 7: **Impact of assortment bias in the presence of a diversity in risk perception:** The first row corresponds to the population P_2' with risk perception diversity $\delta = 0.3$. We plot for agents with high and low risk perception (a) the average group achievement, (b) the learned cooperation policy, and (c) the secured welfare. In the second row, we display the same metrics for the population P_2'' with risk diversity $\delta = 0.5$. We show in (d) the average group achievement, in (e) the learned cooperation policy, and in (f) the secured welfare. $\alpha = 0.5$ represents a setting with no assortment bias, while $\alpha = 1$ represents a setting where agents only interact with agents with a same risk perception. In all panels, shaded areas represent the standard deviation over 5 runs.

As such, we have $\bar{\beta} = \beta = 1$ for no assortment bias or $\bar{\beta} < 1 < \beta$ for settings with assortment bias. To implement the assortment bias in our training algorithm, we modify the group sampling approach in Algorithm 1 (step 1.) and replace it by the procedure described in Algorithm 2. We begin by randomly sampling an agent i from the population. Agent i then defines a weight vector $\mathbf{w}$ that allocates for every agent $j \neq i$ in the population, a weight bias β if that agent belongs to the same risk class, and $\bar{\beta}$ if that agent belongs to the opposite risk class. The weight vector is normalized and used to sample $N - 1$ agents that will interact with player i in the collective risk dilemma.

We run simulations on two populations with different degrees of risk perception diversity, P_2' with $\delta = 0.3$ and P_2'' with $\delta = 0.5$, and evaluate the group achievement, cooperation probability and secured welfare of agents with high and low risk perception under different

Algorithm 2: Sampling with assortment bias.

Init: α, *perceived frequency of same class agents*,
Sample random agent i;
Calculate β, $\bar{\beta}$ *(Eq. 5 and 6)*;
Initialize weight vector $\mathbf{w} \leftarrow \mathbf{0}$;
for $j \leftarrow 1$ **to** Z *and* $j \neq i$ **do**
  if j *same class as* i **then**
    $\mathbf{w}_j = \beta$
  else
    $\mathbf{w}_j = \bar{\beta}$
Normalize $\mathbf{w}$: $\mathbf{w} \leftarrow \mathbf{w}/|\mathbf{w}|$;
Sample random group $G' \sim \mathbf{w}$ *of size N-1*;
Return: *group* $G = G' \cup \{i\}$

assortment biases $\alpha \in \{0.6, 0.7, 0.8, 0.9, 1.\}$. We display the results in Figure 7. In both cases for $\delta = 0.3$ and $\delta = 0.5$, stronger assortment bias results in a stronger achievement divergence between the two classes of agents (Figures 7a and 7d). This is so despite relatively constant cooperation policies with respect to the assortment bias (Figures 7b and 7e). As a result, while cooperation costs remain constant for both agent classes across all assortment arrangements, the disaster losses increase for agents with a low risk perception and decrease for agents with a high risk perception. We observe divergence in secured welfare (Figures 7c and 7f) which, similarly to wealth inequality, causes agents with a high risk perception to become richer, and agents with a low risk perception to become poorer.

We note that while stronger risk diversity only induced small inequalities in target achievement between the two classes (Figure 6d), the effects are magnified with the introduction of an assortment bias. For a same assortment bias α, the population P_2'' with higher risk perception diversity ($\delta = 0.5$), generates stronger class divergence and inequalities.

The results we show under assortment bias validate our hypothesis that insufficient mixing and integration of diverse agents in a same group generates further diversity and inequality in a population.

4.3 Effect of Wealth Inequality with Risk Perception Diversity

As a final analysis, we propose to investigate the impact of introducing both wealth inequality and risk perception diversity in a population facing collective risks. In a population with wealth inequality, where 20% of agents are rich and hold 50% of the total wealth, we consider that rich agents have a high risk perception, while poor agents have a low risk perception.

In a first experiment, we train a population P_3 with wealth inequality and risk perception diversity $\delta = 0.1$ in groups of $N = 6$ and under varying risk factors $r \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. To assess if any network effects emerge from the combination of the two heterogeneities, we compare the performance of population P_3 to that of population P_1 presenting only wealth inequality and plot the results in Figure 8.

Across all metrics, we observe that the introduction of risk diversity ($\delta = 0.1$) is easily absorbed by the population with wealth inequality for all risk values $r \geq 0.3$ but has some

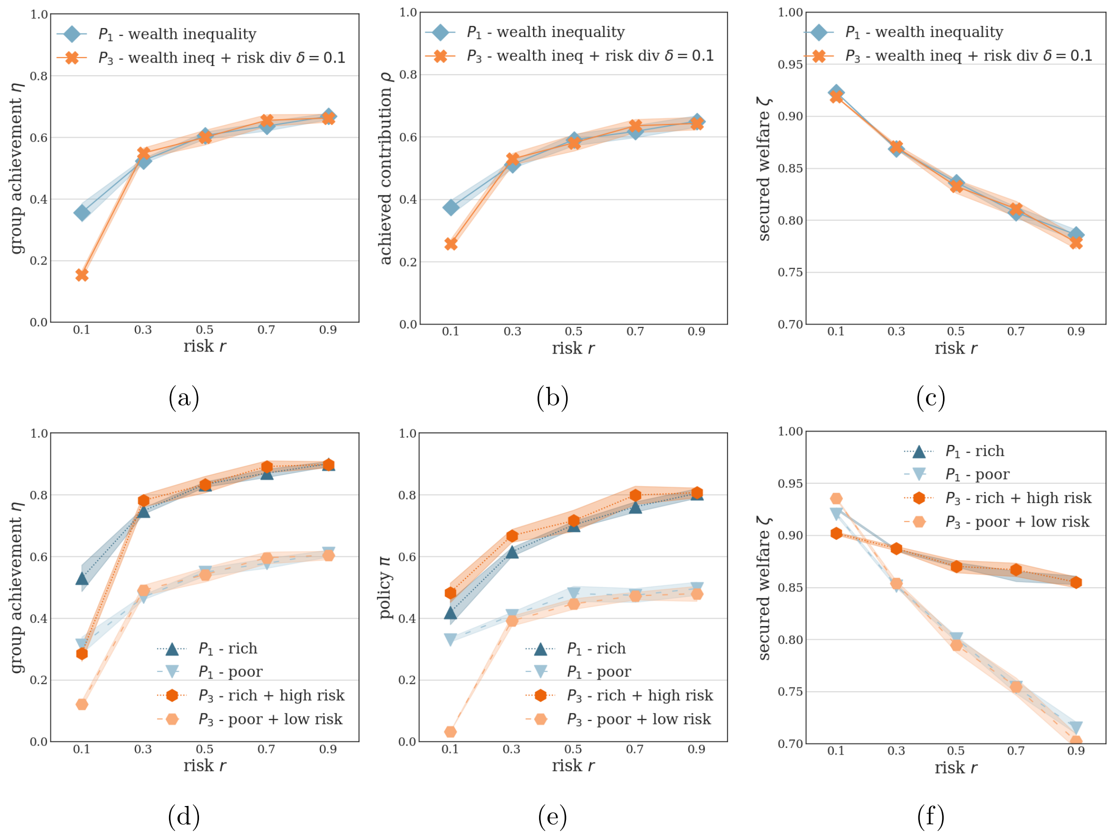
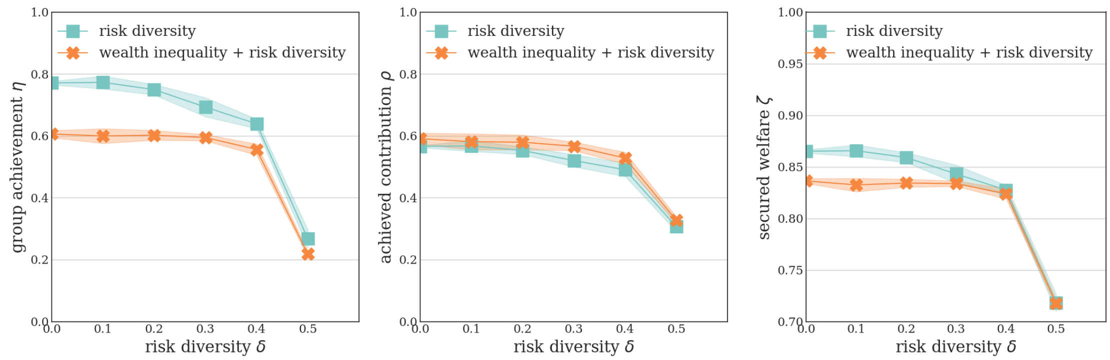


Figure 8: **Impact of risk under wealth inequality and diversity in risk perception:** In the first row, we compare a population with wealth inequality to a population with both wealth inequality and risk diversity. We plot (a) the overall group achievement rate η, (b) the percentage of achieved contributions ρ, and (c) the percentage of remaining wealth in the populations. In the second row, we take a closer look at the internal dynamics in the populations and compare results for rich and poor agents, with and without risk perception bias. Again we plot (d) the overall group achievement rate η, (e) the policy π or the average cooperation probability, and (f) the percentage of the remaining wealth of agents. In all panels, shaded areas represent the standard deviation over 5 runs.

negative effects for $r = 0.1$, where the relative strength δ/r of the diversity is large. The impact of introducing risk perception diversity on a population with wealth inequality is similar to the impact of introducing risk diversity in a homogeneous population. We deduce that the network effects of introducing both a risk diversity of $\delta = 0.1$ and wealth inequality are insignificant.

We recall that the introduction of wealth inequality in a homogeneous population reduces group achievement and total secured welfare. Additionally it generates inequalities between rich and poor agents in target achievement, cooperation and secured welfare. In a second experiment, we investigate how the intensity of these emerging effects from introducing wealth inequality varies under different risk diversities.



(a) (b) (c)

Figure 9: **Impact of risk diversity intensity with and without wealth inequality:** We compare populations with risk perception diversity to populations with both risk perception diversity and wealth inequality. We plot (a) the overall group achievement rate η, (b) the percentage of achieved contributions ρ, and (c) the percentage of remaining wealth in the populations. In all panels, shaded areas represent the standard deviation over 5 runs.

We compare in Figure 9, the performances of populations with risk diversity and populations with both risk diversity and wealth inequality engaging in groups of $N = 6$ in a collective risk dilemma of average risk $r = 0.5$ and presenting varying perception diversities $\delta \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$.

By comparing the two populations at $\delta = 0$, we observe the impact that wealth inequality has when introduced in a population without risk perception diversity. We see that wealth inequality reduces achievement by 20%. We observe in Figure 9a that this gap in achievement between populations with and without wealth inequality decreases for larger risk diversities δ. For higher δ values, the detrimental consequences of risk perception diversity on group achievement outweigh the negative effects of wealth inequality that eventually become insignificant.

Figure 9b shows that wealth inequality has no impact on the populations' contributions whether that population presents risk perception diversity or not. The effect of wealth inequality on populations' contributions is conserved for all risk diversity values.

Figure 9c shows similar results as Figure 9a. The negative effects of wealth inequality slowly fade when outweighed by the negative effects of high risk perception diversity (the two lines eventually merge).

We deduce that no network effects emerge when combining wealth inequality with small risk perception diversity δ. However, as δ increases, the population becomes more robust to the introduced wealth inequality. Only the disadvantages of risk perception diversity are significant.

5. Conclusion

In this work we examined the challenges posed by wealth inequality and risk assessment diversity on a population of reinforcement learners facing a collective risk dilemma. First, our results show that introducing wealth inequality in populations of learning agents significantly decreases collective group achievement in small group sizes and for all risk values. This result is aligned with existing empiricial results obtained with behavioral experiments and the same dilemma (Tavoni et al., 2011). Additionally, we show that initial wealth inequality can generate further wealth inequality, specially in high risk circumstances. We also found that the negative consequences of wealth inequality fade for large group sizes where wealth inequality can actually be beneficial as it decreases coordination difficulties among agents.

In a second experimental layout, we assessed the impact of diversity in risk perception on the chances of collective success. We observe that high diversity in risk assessment can severely hinder achievement and decrease overall contributions and welfare for all agents. Our results suggest that risk perception diversity, by itself, and unlike wealth inequality, does not generate further inequality among agents as all agents suffer equally from the resulting losses. Inequality in group achievement and welfare between the two risk-classes is however introduced when considering interaction assortment (i.e., having individuals with similar risk assessment playing together).

Overall, our results highlight that both wealth inequality and risk assessment diversity place challenges to group achievement, social welfare, and the emergence of egalitarian outcomes in a population of learning agents trying to solve collective risk dilemmas. We find that all studied inequalities have strong unfavorable effects on a population's efficiency and that some initial inequalities can contribute to augment disparities in the agents' population. Our simulations reveal that larger group sizes, able to combine rich and poor agents, can help to limit the emergence of additional inequality. These findings can be helpful to policy makers to implement urban policies and mechanisms that mitigate segregation based on wealth.

On the other hand, our study confirms that introducing mechanisms to level risk assessment is key to improve both group achievement and fairness. Again, the results can be used by world organizations or authorities to argument the need for a world wide aligned awareness of the risks associated with the climate change problem. As pointed in other works, education can prove crucial to ensure a homogeneous assessment of the risks involved in collective action failure (Lee et al., 2015). Our work also affirms the need for knowledge transfer and communication capabilities in AI agents to ensure effective multi-agent cooperation in the presence of conflicting beliefs across agents.

Several extensions can be considered in future works. The work developed here leaves does not explore important decision-making constraints that may interact in non-trivial ways with wealth inequality and risk diversity. Individuals may have the chance to commit with particular choices, being monitored and sanctioning if they do not abide by their pledges (Carattini et al., 2021; Han et al., 2017). Individuals may also adopt conditional behaviors based on the past contributions of other group or class members (Domingos et al., 2020, 2021a). We also expect that social norms (Santos et al., 2018; Szekely et al., 2021; Carattini et al., 2020) and communication (de Melo & Terada, 2020; Foerster et al., 2016; Kim et al., 2021; Omidshafiei et al., 2019) may facilitate cooperation in this context.

Finally, our work discusses binary risk and wealth classes. Real life diversity usually follows more fine-grained (or continuous) distributions, leading to larger set of possible classes. Wealth may follow power-law distributions, while risk diversity may follow a normal distribution, among several other possibilities. Moreover, additional inequalities can be examined. A relevant next step would be to study the role of heterogeneity in the effective impact of collective disasters (the value p in our model). Indeed, when facing climate change risk, rich countries may be more prepared to adapt to the consequences of extreme weather by building flood defenses, planning for heatwaves, among other measures (NASA, 2021). Lastly, future approaches may also consider the inclusion of agents with predetermined behaviors, and test how these individuals can profit from diversity to nudge others to behave more cooperatively (Santos et al., 2019b; Domingos et al., 2021b). This is possibly a relevant research avenue, where the combination of behavioral experiments and theoretical work may lead to valuable insights on cooperation among humans, but also in populations comprising humans and machines.

Acknowledgments

This work was partially supported by FCT-Portugal (UIDB/50021/2020, PTDC/MAT-APL/6804/2020, and PTDC/CCI-INF/7366/2020). This work has also received funding from the European Union's H2020 program (grant 76595).

References

Abou Chakra, M., Bumann, S., Schenk, H., Oschlies, A., & Traulsen, A. (2018). Immediate action is the best strategy when facing uncertain climate change. *Nature Communications*, *9*(1), 1–9.

Alanezi, F., Althumairi, A., et al. (2021). A comparative study of strategies for containing the covid-19 pandemic in gulf cooperation council countries and the european union. *Informatics in Medicine Unlocked*, *23*, 100547.

Auer, P., Cesa-Bianchi, N., Freund, Y., & Schapire, R. E. (1995). Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of IEEE 36th annual foundations of computer science*, pp. 322–331. IEEE.

Axelrod, R. (1980). Effective choice in the prisoner's dilemma. *Journal of Conflict Resolution*, *24*(1), 3–25.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396.

Baker, B., Kanitscheider, I., Markov, T., Wu, Y., Powell, G., McGrew, B., & Mordatch, I. (2019). Emergent tool use from multi-agent autocurricula. *arXiv preprint arXiv:1909.07528*.

Baland, J.-M., & Platteau, J.-P. (1997). Wealth inequality and efficiency in the commons part i: the unregulated case. *Oxford Economic Papers*, *49*(4), 451–482.

Balduzzi, D., Racaniere, S., Martens, J., Foerster, J., Tuyls, K., & Graepel, T. (2018). The mechanics of n-player differentiable games. In *International Conference on Machine Learning*, pp. 354–363. PMLR.

Bloembergen, D., Kaisers, M., & Tuyls, K. (2010). Lenient frequency adjusted q-learning. In *Proc. of 22nd Belgium-Netherlands Conf. on Artif. Intel.* Citeseer.

Bloembergen, D., Tuyls, K., Hennes, D., & Kaisers, M. (2015). Evolutionary dynamics of multi-agent learning: A survey. *Journal of Artificial Intelligence Research*, *53*, 659–697.

Bowling, M., & Veloso, M. (2001). Rational and convergent learning in stochastic games. In *International joint conference on artificial intelligence*, Vol. 17, pp. 1021–1026. Citeseer.

Cao, K., Lazaridou, A., Lanctot, M., Leibo, J. Z., Tuyls, K., & Clark, S. (2018). Emergent communication through negotiation. *arXiv preprint arXiv:1804.03980*.

Carattini, L., Levin, S. A., & Tavoni, A. (2021). How tangible environmental commitments spur cooperative behaviour in local and global dilemmas. *Europe*.

Carattini, S., Levin, S., & Tavoni, A. (2020). Cooperation in the climate commons. *Review of Environmental Economics and Policy*.

Chan, K. S., Mestelman, S., & Muller, R. A. (2008). Voluntary provision of public goods. *Handbook of Experimental Economics Results*, *1*, 831–835.

Chen, X., Szolnoki, A., & Perc, M. (2012). Risk-driven migration and the collective-risk social dilemma. *Physical Review E*, *86*(3), 036101.

Christianos, F., Schäfer, L., & Albrecht, S. V. (2020). Shared experience actor-critic for multi-agent reinforcement learning. *arXiv preprint arXiv:2006.07169*.

Dafoe, A., Hughes, E., Bachrach, Y., Collins, T., McKee, K. R., Leibo, J. Z., Larson, K., & Graepel, T. (2020). Open problems in cooperative ai. *arXiv preprint arXiv:2012.08630*.

Dannenberg, A., Löschel, A., Paolacci, G., Reif, C., & Tavoni, A. (2011). Coordination under threshold uncertainty in a public goods game. *ZEW-Centre for European Economic Research Discussion Paper*.

de Melo, C. M., & Terada, K. (2020). The interplay of emotion expressions and strategy in promoting cooperation in the iterated prisoner's dilemma. *Scientific Reports*, *10*(1), 1–8.

Domingos, E. F., Grujić, J., Burguillo, J. C., Kirchsteiger, G., Santos, F. C., & Lenaerts, T. (2020). Timing uncertainty in collective risk dilemmas encourages group reciprocation and polarization. *iScience*, *23*(12), 101752.

Domingos, E. F., Grujić, J., Burguillo, J. C., Santos, F. C., & Lenaerts, T. (2021a). Modeling behavioral experiments on uncertainty and cooperation with population-based reinforcement learning. *Simulation Modelling Practice and Theory*, *109*(102299).

Domingos, E. F., Terrucha, I., Suchon, R., Grujić, J., Burguillo, J. C., Santos, F. C., & Lenaerts, T. (2021b). Delegation to autonomous agents promotes cooperation in collective-risk dilemmas. *arXiv preprint arXiv:2103.07710*.

Foerster, J., Chen, R. Y., Al-Shedivat, M., Whiteson, S., Abbeel, P., & Mordatch, I. (2018a). Learning with opponent-learning awareness. In *Proc. of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 122–130. IFAAMS.

Foerster, J., Farquhar, G., Afouras, T., Nardelli, N., & Whiteson, S. (2018b). Counterfactual multi-agent policy gradients. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

Foerster, J. N., Assael, Y. M., De Freitas, N., & Whiteson, S. (2016). Learning to communicate with deep multi-agent reinforcement learning. *arXiv preprint arXiv:1605.06676*.

Gibney, E. (2020). Whose coronavirus strategy worked best? scientists hunt most effective policies. *Nature*, *581*(7806), 15–17.

Han, T., Pereira, L., & Lenaerts, T. (2017). Evolution of commitment and level of participation in public goods games. *Autonomous Agents and Multi-Agent Systems*, *31*(3), 561–583.

Hart, S., & Mas-Colell, A. (2000). A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, *68*(5), 1127–1150.

Hauert, C., Holmes, M., & Doebeli, M. (2006). Evolutionary games and population dynamics: maintenance of cooperation in public goods games. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1600), 2565–2571.

Hauser, O. P., Hilbe, C., Chatterjee, K., & Nowak, M. A. (2019). Social dilemmas among unequals. *Nature*, *572*(7770), 524–527.

Hilbe, C., Abou Chakra, M., Altrock, P. M., & Traulsen, A. (2013). The evolution of strategic timing in collective-risk dilemmas. *PloS one*, *8*(6), e66490.

Jacq, A., Perolat, J., Geist, M., & Pietquin, O. (2019). Foolproof cooperative learning. *ArXiv:1906.09831*.

Kaisers, M., & Tuyls, K. (2010). Frequency adjusted multi-agent q-learning. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pp. 309–316.

Katz, D., Baptista, J., Azen, S., & Pike, M. (1978). Obtaining confidence intervals for the risk ratio in cohort studies. *Biometrics*, 469–474.

Kim, W., Park, J., & Sung, Y. (2021). Communication in multi-agent reinforcement learning: Intention sharing. In *International Conference on Learning Representations*.

Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology*, *24*(1), 183–214.

Kurokawa, S., & Ihara, Y. (2009). Emergence of cooperation in public goods games. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1660), 1379–1384.

Lamarche, V. M. (2020). Socially connected and covid-19 prepared: The influence of sociorelational safety on perceived importance of covid-19 precautions and trust in government responses.. *Social Psychological Bulletin*.

Lee, T. M., Markowitz, E. M., Howe, P. D., Ko, C.-Y., & Leiserowitz, A. A. (2015). Predictors of public climate change awareness and risk perception around the world. *Nature Climate Change*, *5*(11), 1014–1020.

Leibo, J. Z., Zambaldi, V., Lanctot, M., Marecki, J., & Graepel, T. (2017). Multi-agent reinforcement learning in sequential social dilemmas. In *Proc. of the 16th Conference on*

Autonomous Agents and MultiAgent Systems, pp. 464–473. International Foundation for Autonomous Agents and Multiagent Systems.

Letcher, A., Foerster, J., Balduzzi, D., Rocktäschel, T., & Whiteson, S. (2018). Stable opponent shaping in differentiable games. *arXiv preprint arXiv:1811.08469*.

Liu, S., Lever, G., Merel, J., Tunyasuvunakool, S., Heess, N., & Graepel, T. (2019). Emergent coordination through competition. *ArXiv:1902.07151*.

Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, O. P., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. In *Advances in Neural Information Processing Systems*, pp. 6379–6390.

Macy, M. W., & Flache, A. (2002). Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences*, *99*(suppl 3), 7229–7236.

Marchesini, E., & Farinelli, A. (2021). Centralizing state-values in dueling networks for multi-robot reinforcement learning mapless navigation. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4583–4588. IEEE.

Matignon, L., Laurent, G. J., & Le Fort-Piat, N. (2012). Independent reinforcement learners in cooperative markov games: a survey regarding coordination problems. *The Knowledge Engineering Review*, *27*(1), 1–31.

Merhej, R., Santos, F. P., Melo, F. S., & Santos, F. C. (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*, pp. 898–906.

Milinski, M., Röhl, T., & Marotzke, J. (2011). Cooperative interaction of rich and poor can be catalyzed by intermediate climate targets. *Climatic change*, *109*(3), 807–814.

Milinski, M., Sommerfeld, R. D., Krambeck, H.-J., Reed, F. A., & Marotzke, J. (2008). The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proceedings of the National Academy of Sciences*, *105*(7), 2291–2294.

NASA (2021). Responding to Climate Change. Available at: https://climate.nasa.gov/solutions/adaptation-mitigation/. Last accessed: 06.10.2021.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive Skills and Their Acquisition*, *1*(1981), 1–55.

Omidshafiei, S., Kim, D.-K., Liu, M., Tesauro, G., Riemer, M., Amato, C., Campbell, M., & How, J. P. (2019). Learning to teach in cooperative multiagent reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, pp. 6128–6136.

Peña, J., & Nöldeke, G. (2018). Group size effects in social evolution. *Journal of Theoretical Biology*, *457*, 211–220.

Peters, O., & Gell-Mann, M. (2016). Evaluating gambles using dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *26*(2), 023103.

Rashid, T., Farquhar, G., Peng, B., & Whiteson, S. (2020). Weighted qmix: Expanding monotonic value function factorisation for deep multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, *33*, 10199–10210.

Rashid, T., Samvelyan, M., Schroeder, C., Farquhar, G., Foerster, J., & Whiteson, S. (2018). Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning. In *International Conference on Machine Learning*, pp. 4295–4304. PMLR.

Roth, A. E., & Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, *8*(1), 164–212.

Santos, F. P., Mascarenhas, S. F., Santos, F. C., Correia, F., Gomes, S., & Paiva, A. (2019a). Outcome-based partner selection in collective risk dilemmas. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 1556–1564. International Foundation for Autonomous Agents and Multiagent Systems.

Santos, F. P., Pacheco, J. M., Paiva, A., & Santos, F. C. (2019b). Evolution of collective fairness in hybrid populations of humans and agents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, pp. 6146–6153.

Santos, F. P., Pacheco, J. M., Santos, F. C., & Levin, S. A. (2021). Dynamics of informal risk sharing in collective index insurance. *Nature Sustainability*, 1–7.

Santos, F. P., Santos, F. C., & Pacheco, J. M. (2018). Social norm complexity and past reputations in the evolution of cooperation. *Nature*, *555*(7695), 242–245.

Santos, F. C., & Pacheco, J. M. (2011). Risk of collective failure provides an escape from the tragedy of the commons. *Proceedings of the National Academy of Sciences*, *108*(26), 10421–10425.

Santos, F. C., Santos, M. D., & Pacheco, J. M. (2008). Social diversity promotes the emergence of cooperation in public goods games. *Nature*, *454*(7201), 213.

Santos, F. C., Vasconcelos, V. V., Santos, M. D., Neves, P., & Pacheco, J. M. (2012). Evolutionary dynamics of climate change under collective-risk dilemmas. *Mathematical Models and Methods in Applied Sciences*, *22*(supp01), 1140004.

Son, K., Kim, D., Kang, W. J., Hostallero, D. E., & Yi, Y. (2019). Qtran: Learning to factorize with transformation for cooperative multi-agent reinforcement learning. In *International Conference on Machine Learning*, pp. 5887–5896. PMLR.

Sunehag, P., Lever, G., Gruslys, A., Czarnecki, W. M., Zambaldi, V., Jaderberg, M., Lanctot, M., Sonnerat, N., Leibo, J. Z., Tuyls, K., & Graepel, T. (2017). Value-decomposition networks for cooperative multi-agent learning. *arXiv preprint arXiv:1706.05296*.

Szekely, A., Lipari, F., Antonioni, A., Paolucci, M., Sánchez, A., Tummolini, L., & Andrighetto, G. (2021). Evidence from a long-term experiment that collective risks change social norms and promote cooperation. *Nature Communications*, *12*(1), 1–7.

Szolnoki, A., & Perc, M. (2011). Group-size effects on the evolution of cooperation in the spatial public goods game. *Physical Review E*, *84*(4), 047102.

Tavoni, A., Dannenberg, A., Kallis, G., & Löschel, A. (2011). Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proceedings of the National Academy of Sciences*, *108*(29), 11825–11829.

Thanh, P. T., et al. (2020). Survey data on government risk communication and citizen compliance during the covid-19 pandemic in vietnam. *Data in Brief*, *33*, 106348.

Thorndike, E. L. (1898). Animal intelligence: an experimental study of the associative processes in animals.. *The Psychological Review*, *2*(4), i.

Tuyls, K., Hoen, P. J., & Vanschoenwinkel, B. (2006). An evolutionary dynamical analysis of multi-agent learning in iterated games. *Autonomous Agents and Multi-Agent Systems*, *12*(1), 115–153.

Van de Poel, I., Fahlquist, J. N., Doorn, N., Zwart, S., & Royakkers, L. (2012). The problem of many hands: Climate change as an example. *Science and Engineering Ethics*, *18*(1), 49–67.

Vasconcelos, V. V., Santos, F. C., & Pacheco, J. M. (2013). A bottom-up institutional approach to cooperative governance of risky commons. *Nature Climate Change*, *3*(9), 797–801.

Vasconcelos, V. V., Santos, F. C., Pacheco, J. M., & Levin, S. A. (2014). Climate policies under wealth inequality. *Proceedings of the National Academy of Sciences*, *111*(6), 2212–2216.

Wang, J., Fu, F., & Wang, L. (2010). Effects of heterogeneous wealth distribution on public cooperation with collective risk. *Physical Review E*, *82*(1), 016102.

Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3), 279–292.

World Economic Forum (2021). The Global Risk Report. Available at: `www.weforum.org/reports/the-global-risks-report-2021`. Last accessed: 22-09-2021.

Zhang, C., & Lesser, V. (2010). Multi-agent learning with policy prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 24.

Zinkevich, M., Johanson, M., Bowling, M., & Piccione, C. (2007). Regret minimization in games with incomplete information. *Advances in Neural Information Processing Systems*, *20*.