Learning in Public Goods Games with Non-Linear Utilities: a Multi-Objective Approach

Work in Progress

Nicole Orzan
University of Groningen
Groningen, Netherlands
n.orzan@rug.nl

Erman Acar
University of Amsterdam
Amsterdam, Netherlands
erman.acar@uva.nl

Davide Grossi
University of Groningen
Groningen, Netherlands
n.orzan@rug.nl

Roxana Rădulescu
Vrije Universiteit Brussels, Belgium
Utrecht University, Netherlands
r.t.radulescu@uu.nl

ABSTRACT

Addressing the question of how to achieve optimal decision-making under risk and uncertainty is crucial to both understanding human decision-making processes, and enhancing the capabilities of artificial agents that collaborate with or support humans. In this work, we address this question in the context of Public Goods Games. We study learning in a novel extended version of the Public Goods Game where agents have different risk preferences, by means of multi-objective reinforcement learning. We introduce a parametric non-linear utility function to model risk preferences at the level of individual agents. These attitudes are represented as preferences over the rewards received from the game. We study the interplay between such preference modeling and environmental uncertainty, which is constructed as noise over the level of incentive alignment in the game the agents play. We observe that different combinations of individual preferences and environmental uncertainties sustain the emergence of cooperative patterns in non-cooperative environments (i.e., where competitive strategies are dominant), while others sustain competitive patterns in cooperative environments (i.e., where cooperative strategies are dominant).

KEYWORDS

Multi-Objective Reinforcement Learning; Public Goods Games; Non-Linear Utility Functions

1 INTRODUCTION

How can cooperation emerge and sustain itself in situations where agents do not necessarily have a direct motive for cooperation? This is a fundamental question in various research areas, such as evolutionary biology [23, 30, 37], political sciences [8, 9], cognitive sciences [43] and physics [11]. To answer this question, researchers developed and studied models of real-world scenarios involving tension between the collective and personal motives, called social dilemmas [13, 28]. The main characteristic of social dilemmas is that players are better-off defecting at the individual level, while, at the group level, the best outcome is mutual cooperation.

This work focuses on a specific class of social dilemmas known as Public Goods Games (PGG), extensively studied in literature [2, 4, 5, 46]. A PGG describes situations where cooperation by all agents is Pareto optimal, but because of the profitability of free-riding [4], rational agents fail too cooperate: defection by all agents is a Nash equilibrium [36]. We refer to this kind of games as mixed-motives, since the incentives of the agents are partially misaligned.

In addition to incentive misalignment, other factors influencing cooperation emergence in many real-world scenarios include uncertainty and different individual attitudes towards risk [21, 29]. Uncertainty can have different sources: we refer to environmental uncertainty when actors are unsure about the amount of goods they can receive from the environment [3, 50], and to social uncertainty when referring to ambiguity regarding the opponents' possible actions [10, 17]. Individual preferences denote a personal inclination toward one choice over another. In the specific context of PGGs, we are interested in modeling individuals which, in conditions of uncertainty, are biased towards participating in the production of the collective good, also called risk seeking agents, and individuals which are more inclined to not participate in it, also called risk averse agents. We refer to these attitudes towards risk as 'preferences' and we model them using a parametric nonlinear utility function of the reward received by individuals from the result of their investment in the collective good.

Since we are working with non-linear utility functions in the PGG, we need to distinguish our perspective from the literature on the PGGs that addresses non-linear public good productions. Specifically, this branch focuses on settings where the resulting public good product comes from a non-linear production process [41]. These are called non-linear public good games, and allow one to model certain real-world situations (populations of bacteria, viruses, or cooperative hunting [12, 40]). In contrast, we shift our focus to an individual level, and capture settings where potentially different attitudes towards risk can occur within a population.

To model preferences, in our work we explicitly decouple the collective versus the individual incentives experienced by the agents, and parameterize the collective incentive at the individual level. This choice allows us to model settings in which individuals in a population can have different perceptions regarding these incentives. Furthermore, we take a multi-objective approach on the optimization of these two levels of rewards, drawing on multi-objective

Proc. of the Adaptive and Learning Agents Workshop (ALA 2024), Avalos, Müller, Wang, Yates (eds.), May 6-7, 2024, Online, https://ala2024.github.io/. 2024.

reinforcement learning (MORL) methods. This way we can investigate learned behaviours that emerge from individually preferred trade-offs between the cooperative and competitive objectives.

Contributions. We investigate learning in PGGs where agents have different risk preferences, modeled as non-linearities over the utility function. We study the interplay between this mechanism and environmental uncertainty from a multi-objective perspective. More specifically, we present the three following contributions. *First*, we propose a novel multi-agent multi-objective environment based on the Extended Public Goods Game (EPGG) [39], called the Multi-Objective EPGG (MO-EPGG). This environment, next to facilitating training agents on games with different levels of incentive alignment, also allows one to explicitly model the trade-off between the individual and the cooperative components of PGGs. Moreover, it enables decoupling environmental and social uncertainties, allowing for the analysis of their impact, both when occurring concurrently or in isolation. *Second*, we propose a non-linear utility function that allows one to combine the collective and individual rewards, parameterized at the agent level. The selected shape of the utility function allows us to model risk averse and risk seeking agents, by operating a convex or concave transformation over the collective game reward. Moreover, it allows to model a population of agents with various attitudes towards risk. *Third*, we perform preliminary experiments on the dynamics of a population of independent multi-objective reinforcement learning agents trained on the MO-EPGG. We show that risk-averse utility functions strongly diminish cooperation in cases with and without uncertainty. In the presence of environmental uncertainty, risk-seeking utilities improve cooperation in non-cooperative environments, which are environments with defection being a dominant strategy.

2 RELATED WORK

2.1 Non-Linear Utilities in Public Goods Games

Although the PGG with linear utility functions is the most known and used, various models of non-linear PGGs have been proposed in the literature as well. In the *threshold public goods game* for example, the resulting public good is given by a step function of the number of cooperators: the resource is created only if a minimum fraction of actors participates in the production of the public good [14]. When the minimum number of participants is 1, this is called the *Volunteer's Dilemma* [6, 18]. A *sigmoid public goods* function closely models many biological systems where the output production is small for low input levels and bigger for intermediate inputs, decreasing again for even bigger ones [7, 12]. In other paradigms, the public good production is modeled by applying a concave (convex) function over the total good accumulated by the agents, whenever the produced good is lesser (greater) than the good provided by a linear function of the good.

Several papers focused on analyzing non-linear public good games, by different means. In [35] authors employ non-linear PGG with different incentive structures to analyze behavioral subtyping, i.e. if cooperative behavior in one task can predict cooperative behavior in another. In [51], evolutionary dynamics techniques are employed to study the role of different non-linear production functions on the evolution of cooperation in finite populations, while in [41], the evolutionary dynamics of two different populations collaborating for the production of a non-linear public good is investigated. In [16] authors explore the effects of different non-linear PGGs on the evolution of cooperation using Darwinian dynamics.

In the aforementioned literature, non-linearities in PGGs are typically functions that influence the production of the public good. In our work, however, we take a different perspective by introducing non-linearities at the level of the individual utilities extracted from rewards. More specifically, our goal is to model individuals' attitudes towards risk. The study of risk and uncertainty has been a central focus in decision theory, which seeks to understand human decisions processes and derive optimal decision-making strategies [26, 31, 49]. Some studies have shown that people make decisions based on some subjective function of the investment they made [19, 47]. For instance, individual's risk attitudes are often described as functions of the investment made (x) by means of a utility function shaped as $u(x) = x^{\beta}$. Here, the parameter β governs the risk preference of the individual: if $0 < \beta < 1$, the function is concave, signifying risk-aversion; if $\beta > 1$, the function is convex, indicating a risk-seeking attitude [26]. In our work, we draw on this idea in order to formulate a utility function that allows us to model individual preferences for actors participating in the PGG.

2.2 Multi-Objective Reinforcement Learning

In the field of reinforcement learning, the main focus is often to solve single-objective problems, by determining the agent's best policy to reach a specific goal. However, real-world challenges are of multi-objective nature most of the time [45]. Autonomous agents, whether human or artificial, need to optimize for multiple goals simultaneously, or find a trade-off between them. This is the central concept of multi-objective reinforcement learning (MORL) [25, 32, 45], a relatively recent field that demonstrated a significant progress in the last years [44]. In MORL, the core idea is to receive vector rewards from the environment instead of scalar rewards. Those are usually combined by means of a scalarization function that should serve the final objective of multiple objective optimization. Often, a linear scalarization function is employed, which allows the employment of single-objective RL methods. Alternatively, other choices include monotonically increasing non-linear scalarization functions [1, 45]. These are of particular interest for our work since non-linear functions are often used to model utilities under uncertainty and risk, especially in economics literature, which aims at modeling human behavior [1, 22, 48].

Another part of this field of research focuses on fairness, i.e., how to optimize the trade-off among the objectives of different individuals under particular fairness constraints [20, 24, 48]. For example, in [48], authors employ deep RL techniques to learn a policy that treats users equitably. We build on this framework, but rather than focusing on the fair treatment of a set of users, we investigate the effect of uncertainty and individuals' attitudes towards risk. To this end, we extend their approach to work with a different scalarisation function customized for our scenario, which allows us to model individual preferences, and train independent reinforcement learning agents in a multi-objective setting. We thus adopt a multi-agent multi-objective reinforcement learning (MOMARL) [42] perspective, which extends MORL to multi-agent scenarios.

3 PRELIMINARIES

In this section we present the formal definitions and the background knowledge supporting our work. These include the Extended Public Goods Game, multi-objective stochastic games and the multi-objective optimization criteria.

3.1 The Extended Public Goods Game

Following [38, 39], we model the Public Goods Games as a tuple $\langle N, \boldsymbol{c}, \boldsymbol{A}, f, \boldsymbol{r} \rangle$. N is the set of players, and $|N| = n$. Every player i is endowed with some amount of wealth (or coins) $c_i \in \mathbb{R}^{\geq 0}$, and $\boldsymbol{c} = (c_1, \ldots, c_n)$ is the tuple containing all agents' coins. Each agent decides whether to invest in the public good (cooperate) or keep the endowment for themselves (defect); therefore, the set of allowed actions for every agent i consists of cooperate (C) and defect (D) i.e., $A_i = \{C, D\}$. $\boldsymbol{A} = A_0 \times ... \times A_n$, and the tuple $\boldsymbol{a} = (a_1, \ldots, a_n) \in \boldsymbol{A}$ represents the action profile of the agents. The *multiplication factor* $f \in (1, n)$ represents the quantity by which the total investment is multiplied to produce the public good, which is then evenly distributed among all agents. The reward function for each agent $r_i : \boldsymbol{A} \times (1, n) \times (\mathbb{R}^{\geq 0})^n \to \mathbb{R}$ is defined as follows:

$$r_i(\boldsymbol{a}, f, \boldsymbol{c}) = \frac{1}{n} \sum_{j=1}^{n} c_j I(a_j) \cdot f + c_i(1 - I(a_i)), \tag{1}$$

where a_j is the $j-$th entry of the action profile $\boldsymbol{a}$ and $I(a_j)$ is the indicator function, equal to 1 if the action of the agent j is cooperative, and 0 otherwise, and c_j denotes the $j-$the entry of $\boldsymbol{c}$. For the sake of simplicity, in the following we assume all endowments to be equal, namely $c_i = c \; \forall \; i \in N$. Since $1 < f < n$, group defection is a dominant strategy equilibrium. Yet, this profile is Pareto dominated by the profile in which all agents cooperate.

Following [38, 39], we define the class of Extended Public Goods Games (EPGG) by letting the value of f range over $(0, R_+)$, where $R_+ > n$ is an arbitrary value. When $1 < f \leq n$, the EPGG models mixed-motives scenarios like the classic PGG. When $0 \leq f < 1$, the EPGG models competitive scenarios, in which the incentives of the participants are fully misaligned. Here the defection profile is a Pareto optimal dominant strategy (and therefore Nash) equilibrium; When instead $n \leq f \leq R_+$, the EPGG models cooperative scenarios, in which the incentives of the participants are instead fully aligned. Here, the cooperation profile is a Pareto optimal dominant strategy (and therefore Nash) equilibrium.

3.2 Multi-Objective Stochastic Games

We model the multi-objective multi-agent interactions using the *multi-objective stochastic game* framework, defined as the tuple $M = (S, \mathcal{A}, T, \mathcal{R})$, with $n \geq 2$ agents and $d \geq 2$ objectives, where:

- S is the state space
- $\mathcal{A} = A_1 \times \cdots \times A_n$ is the set of joint actions, with A_i being the action set of agent i
- $T : S \times \mathcal{A} \times S \to [0, 1]$ is the probabilistic transition function
- $\mathcal{R} = \mathbf{R}_1 \times \cdots \times \mathbf{R}_n$ are the reward functions, where $\mathbf{R}_i : S \times \mathcal{A} \times S \to \mathbb{R}^d$ is the vectorial reward function of agent i for each of the d objectives.[1]

[1]We note that in this article the terms *reward* and *payoff* are synonyms. For the sake of clarity and consistency, we stick to the former term which aligns with the reinforcement learning terminology.

We take a utility-based perspective [45] for multi-objective decision making, assuming that each agent i has a utility function $u_i : \mathbb{R}^d \to \mathbb{R}$ that maps the received reward vector to a scalar value, determining the desired trade-off between the objectives.

3.3 Optimization Criteria

In MORL, the goal of each agent is to find a policy π that maximizes the scalarized return V_u^π under their preferred optimisation criteria. Depending on how agents derive their utility, two optimization criteria can be employed in the scalarisation process, when maximising the expected discounted long-term reward vector:

- The Scalarised Expected Return (SER) criterion:

$$\boldsymbol{V}_u^\pi = u\left(\mathbb{E}_\pi\left[\sum_{t=0}^{\infty} \gamma^t \boldsymbol{r}_t \mid \mu_0\right]\right), \tag{2}$$

  where μ_0 is the distribution over initial states, γ is the discount factor, and π is the agent's policy. $\boldsymbol{r}_t = \mathbf{R}(s_t, \mathbf{a}_t, s_{t+1})$ is the vectorial reward obtained by an agent at timestep t.
- The Expected Scalarised Return (ESR) criterion:

$$\boldsymbol{V}_u^\pi = \mathbb{E}_\pi\left[u\left(\sum_{t=0}^{\infty} \gamma^t \boldsymbol{r}_t\right) \mid \mu_0\right] \tag{3}$$

Which criteria to choose depends on the problem at hand [42]. If we care about the goodness of a single policy execution, ESR is the correct criterion. If instead we are interested in the quality of average policy executions, we should use SER. While under linear utility function the optimization criteria are equivalent, under non-linear utility functions they may output different solutions. In this work, we opt for the SER criterion, modelling agents that are interested in optimising their behaviour in repeated interaction settings.

4 MULTI-OBJECTIVE EPGG

We formulate a multi-objective version of the EPGG, called Multi-Objective Extended Public Goods Game (MO-EPGG), by employing the framework of Multi-Objective Stochastic Games, outlined in Section 3.2. In our framework, the state space consists of the value of the multiplication factor f of the game currently being played, and the action space coincides with that of the single-objective EPGG, outlined in Section 3.1. The transition function is simply a random sampling from the set of possible multiplication factors at the beginning of each episode, and deterministically returns that same value of f at all the subsequent steps of the episode.

To complete our multi-objective formulation of the EPGG, we need to vectorize the scalar reward signal obtained by agents in the EPGG. This process is called *multi-objectivization* of single-objective problems [27, 33]. By observing the form of the reward function in Equation 1, we can easily distinguish between the part that defines the collective (r^C) and the individual payoff (r^I):

$$r_i^C(\boldsymbol{a}, f, \boldsymbol{c}) = \frac{1}{n} \sum_{j=1}^{n} c_j I(a_j) \cdot f \tag{4}$$

$$r_i^I(\boldsymbol{a}, \boldsymbol{c}) = c_i(1 - I(a_i)). \tag{5}$$

Then, in the proposed MO-EPGG, the vectorial reward received by agent i, given action profile $\boldsymbol{a}$, current multiplication factor f, and tuple of endowments $\boldsymbol{c}$, is:

$$\boldsymbol{r}_i(\boldsymbol{a}, f, \boldsymbol{c}) = \left(r_i^C(\boldsymbol{a}, f, \boldsymbol{c}), r_i^I(\boldsymbol{a}, \boldsymbol{c})\right). \tag{6}$$

This completes our description of the MO-EPGG as a MOSG. In Figure 1 we display an example for the vectorial rewards received by $N = 2$ agents playing the MO-EPGG for three different values of the multiplication factor f.

In order to perform multi-objective optimization, we need to define a utility function to optimize the agents' policies. To do this, we follow the approach outlined in [35], by applying a non-linear transformation over the collective component of the reward vector. However, we adjust their approach to our setting, using the non-linear function to model agent's preferences over the collective outcome as risk seeking or risk avoiding behaviors. In particular, in our model the benefit obtained from the collective reward behaves non-linearly by means of an exponential function. Therefore, we define the following non-linear utility function that specifies the final scalarised utility for the MO-EPGG, for each agent i:

$$u_i(\boldsymbol{\nu}_i) = w_i^C \left(\nu_i^C\right)^{\beta_i} + w_i^I \, \nu_i^I, \tag{7}$$

with hyperparameters w_i^C, w_i^I, and β_i. Here, ν^C and ν^I are the expected discounted sums of rewards: $\nu^k = \sum_t \gamma^t r_t^k$, for $k \in \{C, I\}$. We note that we are employing expected returns rather than rewards, since we are working under the SER criterion. w_i^C and w_i^I are weights applied to the cooperative and competitive components of the vectorial expected return $\boldsymbol{\nu}_i$, for each agent i, and take values in $\mathbb{R}^{\geq 0}$. They can also be interpreted as individual preferences. In this equation, the parameter β_i governs the risk seeking/averse behavior of agent i towards the collective expected return ν_i^C: $\beta = 1$ returns a linear utility function, $\beta < 1$ generates a concave function, modeling a risk avoiding agent, while $\beta > 1$ generates a convex function, modeling a risk seeking agent [35].

In Equation 7, the exponent is only applied over the collective reward component. This choice is motivated by our conceptualization of the collective reward as the result of a (possibly) risky investment. And the result depends on 1) the value of the multiplication factor f, which might not be known with certainty by the agents, and 2) the actions of the other players. Moreover, one can observe that the preference between the cooperative or defective behavior depends on the relationship between three values, namely, f, c and β. In particular, assuming an equal value of β among the whole population of agents ($\beta_i = \{\beta\}_{i \in N}$), the collective cooperative action ($\boldsymbol{a}_C = \{C\}_{i \in N}$) is preferred over the collective defective action ($\boldsymbol{a}_D = \{D\}_{i \in N}$) by all the agents when $r^C(\boldsymbol{a}_C, f, \boldsymbol{c}) > r_i^I(\boldsymbol{a}_D, \boldsymbol{c})$, which is the case when $(cf)^\beta > c$. This relationship between the variables induces a shared preference over collective cooperative behavior in otherwise defective scenarios (the case when $f < 1$). In general, collective cooperation is preferred over collective defection whenever either of the following conditions holds:

$$\beta < \frac{\log(c)}{\log(cf)} \quad \text{if} \quad 0 < cf < 1 \tag{8}$$

$$\beta > \frac{\log(c)}{\log(cf)} \quad \text{if} \quad cf > 1. \tag{9}$$

$f = 0.5$

Player 0 \ Player 1	C	D
C	[2, 0], [2, 0]	[1, 0], [1, 4]
D	[1, 4], [1, 0]	[0, 4], [0, 4]

$f = 1.5$

Player 0 \ Player 1	C	D
C	[6, 0], [6, 0]	[3, 0], [3, 4]
D	[3, 4], [3, 0]	[0, 4], [0, 4]

$f = 2.5$

Player 0 \ Player 1	C	D
C	[10, 0], [10, 0]	[5, 0], [5, 4]
D	[5, 4], [5, 0]	[0, 4], [0, 4]

Figure 1: Multi-objective payoff matrices received by 2 players with 4 coins each, playing the MO-EPGG with multiplication factors of 0.5, 1.5 and 2.5, when taking the cooperative (C) or defective (D) actions.

In the same way, collective defection is preferred over collective cooperation whenever $(cf)^\beta < c$. [2] We remark that this result does not mean that $\boldsymbol{a}_C$ is a Nash equilibrium of the game, since other mixed-strategy equilibria could be present.

5 EXPERIMENTAL SETUP

5.1 Algorithms

In this work we train independent RL agents by utilising the multi-objective version of the Deep Q-network (DQN) algorithm [34] described in [48], that allows us to optimize policies under the SER criterion. In their work, the DQN is trained to predict a Q-function for every objective. Therefore, the output of the DQN has dimensionality $|A| \times d$ where d represents the number of objectives. To adapt their DQN modification to our setting, we adjust their algorithm to work with our scalarization function (Equation 7). The loss function for the DQN can be expressed as follows:

$$L(\theta) = \mathbb{E}_{s,a,s',r \sim D}\left[\left(\boldsymbol{r} + \gamma \, \hat{\boldsymbol{Q}}_{\theta'}(s', a^*) - \hat{\boldsymbol{Q}}_{\theta}(s, a) \right)^2 \right], \tag{10}$$

where θ and θ' represent the weights of the DQN at two different timesteps of the training. D represents the buffer of stored transitions, and $\boldsymbol{r}$ is the vector reward. We find the best action a^* by applying the SER optimization criterion:

$$a^* = \text{argmax}_{a \in A} u\left(\mathbb{E}[\boldsymbol{r} + \gamma \, \hat{\boldsymbol{Q}}_{\theta'}(s', a')] \right), \tag{11}$$

[2] Setting the value $c = 4$ for every agent, cooperation becomes rational in defective environments whenever $\beta < \log(4)/\log(4f)$ if $f < 0.25$, and whenever $\beta > \log(4)/\log(4f)$ if $f > 0.25$.

namely, by applying our custom scalarization function u to update of the DQN function described in [48][3].

5.2 Experiments

The preliminary experiments are run over a pool of $N = 20$ agents. At each iteration t of the learning, a multiplication factor f_t is uniformly sampled from the interval $[f_{\min}, f_{\max}]$, chosen such as to include cooperative, competitive, and mixed-motive games. Afterwards, M agents, sampled from the pool, participate in the game for 10 rounds. We fixed $M = 4$ and picked $f_{\min} = 0.5$ and $f_{\max} = 6.5$. This choice enables the sampling from a set that contains competitive ($f < 1$), mixed motive ($1 < f < M$), and cooperative games ($f > M$). Each agent receives as observation the current value of the multiplication factor — which can be observed with uncertainty — together with the previous actions taken by all opponents at the previous time step: $\boldsymbol{o}_t^i = (f_{obs}^i, \boldsymbol{a}_{t-1}^{-i})$, where $\boldsymbol{a}^{-i} = (a^j)_{j \in M} : j \neq i$. Therefore, each agent will learn a policy $\pi_A^i : O_i \times A \rightarrow [0, 1]$ which is optimized under the selected utility function, using the SER optimization criteria (see Section 3.3). We model uncertainty over the observation of the multiplication factor as Gaussian noise over the value of f received from the environment: $f_{obs}^i = f + \mathcal{N}(0, \sigma_i^2)$, where σ_i is the uncertainty experienced by agent i. To maintain consistency with the allowed values in the EPGG, negative sampled values are rounded up to 0.

All the experiments are run for 20000 epochs, and results are averaged over 20 runs for every condition. The learning rate is set to $\lambda = 0.001$, and $\gamma = 0.99$. The DQN has one hidden layer with size 4, and the action selection mechanism is ϵ-greedy, with $\epsilon = 0.01$. The values of the weights are fixed as $w^C = w^I = 1$ for all the agents. The plots show the values of the average cooperation of the active agents at every evaluation step of the learning process.

6 RESULTS

6.1 Learning with homogeneous preferences

We first explore the impact of different values of β on the scenarios with and without uncertainty on the observations. We performed experiments for three values of β, that define a linear ($\beta = 1$), a convex ($\beta = 2$) and a concave ($\beta = 0.5$) utility function. In each of these three experiments, the β values are identical for every agent. The results for these experiments are depicted in Figure 2. The experiments with $\beta = 0.5$ symbolize a system of risk avoiding agents playing the MO-EPGG, that strongly push the system behavior toward competition across all games. The result stays consistent across the scenarios with and without uncertainty. In Section 4 we observed that collective cooperation is preferred over collective defection every time that $(cf)^\beta > c$. However, here, with $\beta = 0.5$, collective cooperation should be preferred over collective defection every time $f > 4$, but this was not observed in the experiments. This can be due to the presence of other mixed-motive equilibria, or the effect of the concurrent learning among the set of games. We plan to investigate these possibilities as part of our future work.

The experiments with $\beta = 1$ represent the baseline in which the agents are playing the linear version of the EPGG. Therefore, in the games without uncertainty, we observe as expected convergence to cooperation whenever $f > M$, convergence to defection whenever $f < 1$, and a very slight percentage of cooperation when $1 < f < M$. When uncertainty is introduced, cooperation is increased in all the competitive and mixed-motive scenarios. This results from the concurrent learning from a set of games with different levels of incentive alignment, as previously observed in [39].

The experiments with $\beta = 2$ symbolize a system of risk seeking agents playing in the MO-EPGG. Here, we observe that the cooperation of the system is drastically increased in all games. Again, we note that for $\beta = 2$ and $f = 0.5$ we expect mutual cooperation to be preferred over mutual defection. However, this was not the case, and the most likely explanation is the presence of additional mixed-strategy equilibria. Interestingly, in the scenarios with uncertainty, cooperation emerges as the only learned equilibrium even in competitive games, overcoming such effect.

6.2 Learning with heterogeneous preferences

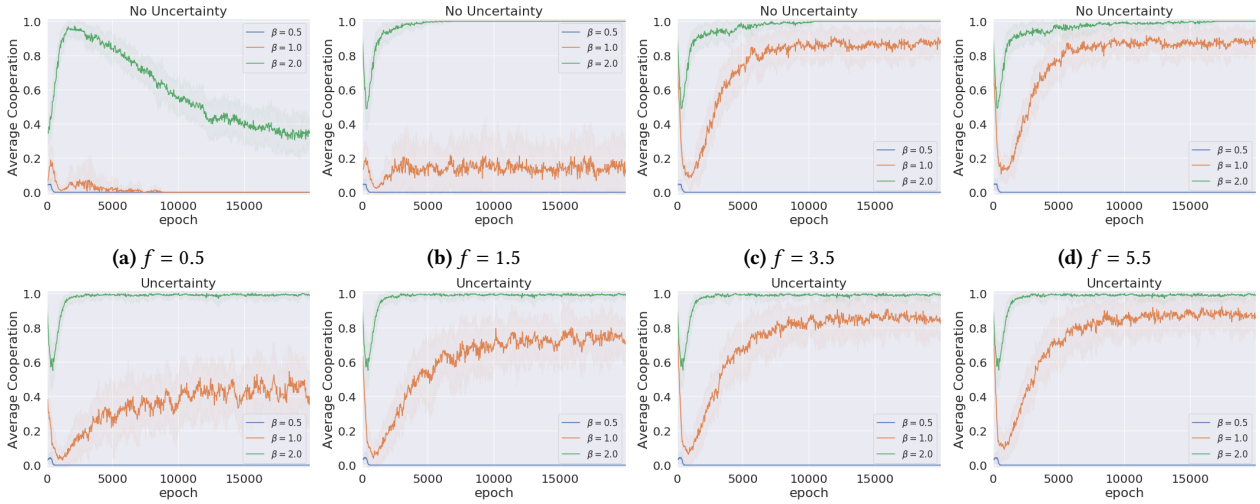
Secondly, we investigate the impact of learning in the MO-EPGG when the agents' preferences β_i are heterogeneous. Specifically, we observe the case in which the value of β_i for every agent i is sampled form a normal distribution centered in 1: $\beta_i \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2)\ \forall\ i \in N$, with $\mu_\beta = 1$ and $\sigma_\beta = 0.5$. The resulting system represents an heterogeneous population where not every individual has the same risk preference, and is centered on risk-neutrality ($\beta = 1$). Figure 3 reports the results of the experiments for the scenarios without (top row) and with (bottom row) uncertainty on the observations. We can observe that the effect of heterogeneity, when uncertainty is not introduced, is to drastically reduce cooperation in the cooperative scenarios ($f > M$), while keeping competition as an equilibrium in the competitive and mixed-motive games. When uncertainty is introduced, cooperation is increased in the competitive and mixed games with respect to the cases without uncertainty. This result is consistent with previous findings over the presence of uncertainty in non-cooperative environments [38, 39]. Interestingly, only the non-cooperative games are affected by the presence of uncertainty: in the cooperative games, the average cooperation of the system is equal to the one observed in the scenario without uncertainty. We plan to delve in the reason for this outcome in our future work.

7 CONCLUSIONS AND FUTURE WORK

In this work, we investigate the role of misalignment of incentives, uncertainty and individual preferences on agents' cooperation in the scenario of a novel multi-objective extended public goods game, employing the tool of multi-objective reinforcement learning. In particular, we observed how risk averse attitudes can increase defection in cooperative environments, and, inversely, risk seeking ones can grow cooperation in competitive and mixed games, especially when uncertainty is introduced. Moreover, we observed how a population with heterogeneous risk attitudes, centered in risk neutrality, tends to not reach cooperation in cooperative games. When also uncertainty is introduced, the observed behavior is to act cooperatively 50% probability.

As future work, we aim to analyze the proposed scenario in two orthogonal directions. First we plan to investigate the effect of a population with heterogeneous preferences (w_i^C, w_i^I and β) on

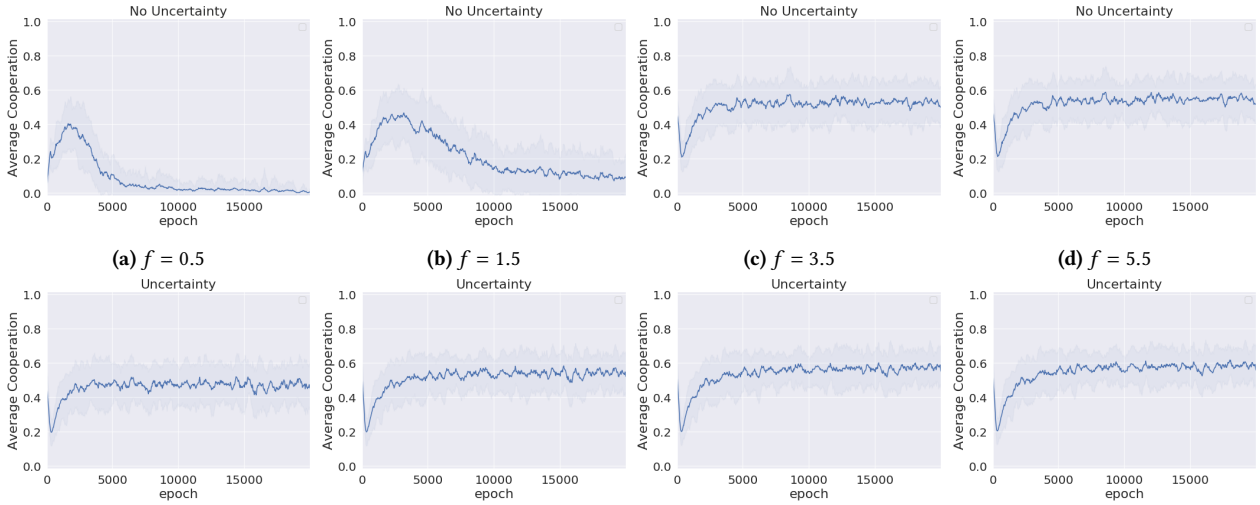
[3] As pointed out also in [48], the scalarization of the expectation is hard to compute, therefore we actually compute the expectation of the scalarization, which is its lower bound.



(a) $f = 0.5$ **(b)** $f = 1.5$ **(c)** $f = 3.5$ **(d)** $f = 5.5$

(e) $f = 0.5$ **(f)** $f = 1.5$ **(g)** $f = 3.5$ **(h)** $f = 5.5$

Figure 2: Average cooperation values for the active DQN agents trained across environments with different multiplication factors, without uncertainty (top row) and with uncertainty on the observations of the multiplication factor $\sigma_i = 2 \; \forall \; i \in N$ (bottom row). Both experiments have been run with three different values of β, identical for every agent $\beta_i = \beta \; \forall \; i \in N$. Shaded regions represent standard deviations.



(a) $f = 0.5$ **(b)** $f = 1.5$ **(c)** $f = 3.5$ **(d)** $f = 5.5$

(e) $f = 0.5$ **(f)** $f = 1.5$ **(g)** $f = 3.5$ **(h)** $f = 5.5$

Figure 3: Average cooperation values for the active DQN agents trained across environments with different multiplication factors, without uncertainty (top row) and with uncertainty on the observations of the multiplication factor $\sigma_i = 2 \; \forall \; i \in N$ (bottom row). Both experiments have been ran with values of β randomly sampled from a normal distribution $\beta_i \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2) \; \forall \; i \in N$, with $\mu_\beta = 1$ and $\sigma_\beta = 0.5$. Shaded regions represent standard deviations.

the learning dynamics and cooperation levels in the MO-EPGG. Second, we will investigate the role of uncertainty over the observation of the multiplication factor, extending the work conducted by Orzan et al. [38]. We thus plan to explore the interplay between heterogeneous preferences and uncertainty. We also intend to explore the outcomes of learning with non-linear utility functions for the PGG when reputation mechanisms and social norms are present. Moreover, we plan to take into account other non-linear risk averse/seeking functions, and compare results with different state-of-the-art independent RL algorithms, such as independent proximal policy optimization (IPPO) [15].

REFERENCES

[1] Mridul Agarwal, Vaneet Aggarwal, and Tian Lan. 2022. Multi-objective reinforcement learning with non-linear scalarization. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*. 9–17.

[2] Simon P Anderson, Jacob K Goeree, and Charles A Holt. 1998. A theoretical analysis of altruism and decision error in public goods games. *Journal of Public economics* 70, 2 (1998), 297–323.

[3] Peter Andras, John Lazarus, Gilbert Roberts, and Steven J Lynden. 2006. Uncertainty and cooperation: Analytical results and a simulated agent society. *Journal of Artificial Societies and Social Simulation* (2006).

[4] James Andreoni. 1988. Why free ride?: Strategies and learning in public goods experiments. *Journal of public Economics* 37, 3 (1988), 291–304.

[5] James Andreoni, Paul M Brown, and Lise Vesterlund. 2002. What makes an allocation fair? Some experimental evidence. *Games and Economic Behavior* 40, 1 (2002), 1–24.

[6] Marco Archetti and István Scheuring. 2011. Coexistence of cooperation and defection in public goods games. *Evolution* 65, 4 (2011), 1140–1148.

[7] Marco Archetti and István Scheuring. 2016. Evolution of optimal Hill coefficients in nonlinear public goods games. *Journal of Theoretical Biology* 406 (2016), 73–82.

[8] Robert Axelrod. 1981. The emergence of cooperation among egoists. *American political science review* 75, 2 (1981), 306–318.

[9] Robert Axelrod and William D Hamilton. 1981. The evolution of cooperation. *science* 211, 4489 (1981), 1390–1396.

[10] Jonathan Bendor. 1993. Uncertainty and the evolution of cooperation. *Journal of Conflict resolution* 37, 4 (1993), 709–734.

[11] Damien Challet and Y-C Zhang. 1997. Emergence of cooperation and organization in an evolutionary game. *Physica A: Statistical Mechanics and its Applications* 246, 3-4 (1997), 407–418.

[12] John S Chuang, Olivier Rivoire, and Stanislas Leibler. 2010. Cooperation and Hamilton's rule in a simple synthetic microbial system. *Molecular systems biology* 6, 1 (2010), 398.

[13] Robyn M Dawes. 1980. Social dilemmas. *Annual review of psychology* 31, 1 (1980), 169–193.

[14] Kris De Jaegher. 2020. High thresholds encouraging the evolution of cooperation in threshold public-good games. *Scientific Reports* 10, 1 (2020), 5863.

[15] Christian Schroeder De Witt, Tarun Gupta, Denys Makoviichuk, Viktor Makoviychuk, Philip HS Torr, Mingfei Sun, and Shimon Whiteson. 2020. Is independent learning all you need in the starcraft multi-agent challenge? *arXiv preprint arXiv:2011.09533* (2020).

[16] Kuiying Deng and Tianguang Chu. 2011. Adaptive evolution of cooperation through Darwinian dynamics in public goods games. *PLoS One* 6, 10 (2011), e25496.

[17] Paul Deutchman, Dorsa Amir, Matthew R Jordan, and Katherine McAuliffe. 2022. Common knowledge promotes cooperation in the threshold public goods game by reducing uncertainty. *Evolution and Human Behavior* 43, 2 (2022), 155–167.

[18] Andreas Diekmann. 1985. Volunteer's dilemma. *Journal of conflict resolution* 29, 4 (1985), 605–610.

[19] James Dow and Sérgio Ribeiro da Costa Werlang. 1992. Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica: Journal of the Econometric Society* (1992), 197–204.

[20] Zimeng Fan, Nianli Peng, Muhang Tian, and Brandon Fain. 2022. Welfare and Fairness in Multi-objective Reinforcement Learning. *arXiv preprint arXiv:2212.01382* (2022).

[21] Ernst Fehr and Urs Fischbacher. 2002. Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. *The economic journal* 112, 478 (2002), C1–C33.

[22] K. R. Foster. 2004. Diminishing returns in social evolution: the not-so-tragic commons. *Journal of Evolutionary Biology* 17, 5 (09 2004), 1058–1072. https://doi.org/10.1111/j.1420-9101.2004.00747.x arXiv:https://academic.oup.com/jeb/article-pdf/17/5/1058/54434641/j.1420-9101.2004.00747.x.pdf

[23] Steven A Frank. 1995. Mutual policing and repression of competition in the evolution of cooperative groups. *Nature* 377, 6549 (1995), 520–522.

[24] Niko A Grupen, Bart Selman, and Daniel D Lee. 2022. Cooperative multi-agent fairness and equivariant policies. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 9350–9359.

[25] Conor F Hayes, Roxana Rădulescu, Eugenio Bargiacchi, Johan Källström, Matthew Macfarlane, Mathieu Reymond, Timothy Verstraeten, Luisa M Zintgraf, Richard Dazeley, Fredrik Heintz, et al. 2022. A practical guide to multi-objective reinforcement learning and planning. *Autonomous Agents and Multi-Agent Systems* 36, 1 (2022), 26.

[26] Joseph G Johnson and Jerome R Busemeyer. 2010. Decision making under risk and uncertainty. *Wiley Interdisciplinary Reviews: Cognitive Science* 1, 5 (2010), 736–749.

[27] Joshua D Knowles, Richard A Watson, and David W Corne. 2001. Reducing local optima in single-objective problems by multi-objectivization. In *International conference on evolutionary multi-criterion optimization*. Springer, 269–283.

[28] Peter Kollock. 1998. Social dilemmas: The anatomy of cooperation. *Annual review of sociology* 24, 1 (1998), 183–214.

[29] Shirli Kopelman, J Mark Weber, and David M Messick. 2002. Factors influencing cooperation in commons dilemmas: A review of experimental psychological research. *The drama of the commons* (2002), 113–156.

[30] Egbert G Leigh Jr. 1977. How does selection reconcile individual advantage with the good of the group? *Proceedings of the National Academy of Sciences* 74, 10 (1977), 4542–4546.

[31] Haim Levy and Moshe Levy. 2002. Arrow-Pratt risk aversion, risk premium and decision weights. *Journal of Risk and Uncertainty* 25 (2002), 265–290.

[32] Chunming Liu, Xin Xu, and Dewen Hu. 2014. Multiobjective reinforcement learning: A comprehensive overview. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 45, 3 (2014), 385–398.

[33] Xiaoliang Ma, Zhitao Huang, Xiaodong Li, Yutao Qi, Lei Wang, and Zexuan Zhu. 2021. Multiobjectivization of single-objective optimization in evolutionary computation: a survey. *IEEE Transactions on Cybernetics* (2021).

[34] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-level control through deep reinforcement learning. *nature* 518, 7540 (2015), 529–533.

[35] Timothy L Mullett, Rebecca L McDonald, and Gordon DA Brown. 2020. Cooperation in Public Goods Games Predicts Behavior in Incentive-Matched Binary Dilemmas: Evidence for Stable Prosociality. *Economic Inquiry* 58, 1 (2020), 67–85.

[36] John F Nash Jr. 1950. Equilibrium points in n-person games. *Proceedings of the national academy of sciences* 36, 1 (1950), 48–49.

[37] Martin A Nowak, Akira Sasaki, Christine Taylor, and Drew Fudenberg. 2004. Emergence of cooperation and evolutionary stability in finite populations. *Nature* 428, 6983 (2004), 646–650.

[38] Nicole Orzan, Erman Acar, Davide Grossi, and Roxana Radulescu. 2023. Emergent Cooperation and Deception in Public Good Games. https://alaworkshop2023.github.io 2023 Adaptive and Learning Agents Workshop at AAMAS, ALA 2023 ; Conference date: 29-05-2023 Through 30-05-2023.

[39] Nicole Orzan, Erman Acar, Davide Grossi, and Roxana Rădulescu. 2024. Emergent Cooperation under Uncertain Incentive Alignment. In *Proceedings of the 2024 International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024)*.

[40] Craig Packer and Lore Ruttan. 1988. The evolution of cooperative hunting. *The American Naturalist* 132, 2 (1988), 159–198.

[41] Arunava Patra, Vikash Kumar Dubey, and Sagar Chakraborty. 2022. Coexistence of coordination and anticoordination in nonlinear public goods game. *Journal of Physics: Complexity* 3, 4 (2022), 045006.

[42] Roxana Rădulescu, Patrick Mannion, Diederik M Roijers, and Ann Nowé. 2020. Multi-objective multi-agent decision making: a utility-based analysis and survey. *Autonomous Agents and Multi-Agent Systems* 34, 1 (2020), 10.

[43] David G Rand and Martin A Nowak. 2013. Human cooperation. *Trends in cognitive sciences* 17, 8 (2013), 413–425.

[44] Mathieu Reymond, Conor F Hayes, Denis Steckelmacher, Diederik M Roijers, and Ann Nowé. 2023. Actor-critic multi-objective reinforcement learning for non-linear utility functions. *Autonomous Agents and Multi-Agent Systems* 37, 2 (2023), 23.

[45] Diederik M Roijers, Peter Vamplew, Shimon Whiteson, and Richard Dazeley. 2013. A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research* 48 (2013), 67–113.

[46] Francisco C Santos, Marta D Santos, and Jorge M Pacheco. 2008. Social diversity promotes the emergence of cooperation in public goods games. *Nature* 454, 7201 (2008), 213–216.

[47] Leonard J Savage. 1972. *The foundations of statistics*. Courier Corporation.

[48] Umer Siddique, Paul Weng, and Matthieu Zimmer. 2020. Learning fair policies in multi-objective (deep) reinforcement learning with average and discounted rewards. In *International Conference on Machine Learning*. PMLR, 8905–8915.

[49] John Von Neumann and Oskar Morgenstern. 1947. Theory of games and economic behavior, 2nd rev. (1947).

[50] Arjaan Wit and Henk Wilke. 1998. Public good provision under environmental and social uncertainty. *European journal of social psychology* 28, 2 (1998), 249–256.

[51] Yanling Zhang, Feng Fu, Te Wu, Guangming Xie, and Long Wang. 2013. A tale of two contribution mechanisms for nonlinear public goods. *Scientific reports* 3, 1 (2013), 2021.