Emergent Cooperation and Deception in Public Good Games

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

Erman Acar
University of Amsterdam
e.acar@uva.nl

Davide Grossi
University of Groningen
University of Amsterdam
d.grossi@rug.nl

Roxana Rădulescu
Vrije Universiteit Brussel
roxana.radulescu@vub.be

ABSTRACT

Communication is a widely used mechanism to promote cooperation in multi-agent systems. In the field of emergent communication agents are usually trained on a particular type of environment: cooperative, competitive, or mixed-motive. Motivated by the idea that real-world settings are characterised by incomplete information and that humans face daily interactions under a wide spectrum of incentives, we hypothesise that emergent communication could be simultaneously exploited in the totality of these scenarios. In this work we pursue this line of research by focusing on social dilemmas, and develop an extended version of the Public Goods Game which allows us to train independent reinforcement learning agents simultaneously on different scenarios where incentives are aligned (or misaligned) to various extents. Additionally, we introduce uncertainty regarding the alignment of incentives, and we equip agents with the ability to learn a communication policy, to study the potential of emergent communication for overcoming uncertainty. We show that in settings where all agents have the same level of uncertainty, communication can help improve the cooperation level of the system, while, when uncertainty is asymmetric, certain agents learn to use communication to deceive and exploit their uncertain peers.

KEYWORDS

Emergent Communication, Social Dilemmas, Multi-Agent Reinforcement Learning, Public Goods Game

1 INTRODUCTION

Cooperation is a fundamental feature of human societies. Its emergence among self-interested agents has traditionally been considered a challenge by disciplines concerned with human interaction, such as biology or sociology [40]. More recently, cooperation has been gaining a central stage also in artificial intelligence research: the ability for cooperation is now considered an essential feature for artificial agents to be able to operate meaningfully within human societies [1, 7]

Substantial literature exists already on the emergence of cooperation among artificial agents [8, 37, 38]. The bulk of it, however, concentrates on the ability of agents to learn to cooperate in cooperative environments, that is, environments where clear incentives for cooperation exist. More recently, a handful of papers have also started to address the challenge of the emergence of cooperation in environments where the incentives for cooperation are weaker [5, 35], which are called mixed-motives environments [41]. Typical instances of such environments are those involving interactions known in economics and sociology as social dilemmas [24]. The present paper focuses on one such social dilemma, known as the public good game [2], taking a reinforcement learning (RL) [44] perspective.

Within the multi-agent reinforcement learning (MARL) framework, agents are usually trained to act optimally in either cooperative, competitive or mixed-motives environments. However, humans and animals learn to operate on all of those environments at the same time. Therefore we believe that is important to study the learning outcome of RL agents when trained on a spectrum of environments, in which their incentives are (mis)aligned to various degrees. This study is arguably essential for the development of MARL based applications that are able to deal with complex real-world scenarios.

In this work, we analyze the learning process of agents trained on a spectrum of environments eliciting different levels of cooperation. Crucially, we aim to understand what are the effects of uncertainty regarding the degree of incentives' alignment on the level of cooperation that agents are able to achieve, and whether cheap-talk [14] (i.e., non-binding and costless signals) emergent communication can help improve it.

We present the following key *contributions*:

- We develop a multi-agent environment based on the Public Goods Game, which we refer to as the Extended Public Good Game (EPGG). This environment allows to train agents on a spectrum of games ranging from fully cooperative, mixed-motives, to fully competitive;
- Within the EPGG, we analyze the impact of uncertainty on the level of cooperation achieved by independent RL agents. We show that the introduction of uncertainty in a mixed cooperative-competitive scenario shifts the outcome towards a less cooperative behavior.
- We explore the role of emergent communication in the EPGG. We empirically show that when all the agents are uncertain, introducing a communication channel between agents can help to overcome uncertainty, specifically moving the outcome of the learning process towards cooperation. Moreover, when uncertainty is asymmetric, communication can be used by the certain agents to deceive the others.
- We also show that adding communication to the case with uncertainty allows to improve cooperation over the case

Proc. of the Adaptive and Learning Agents Workshop (ALA 2023), Cruz, Hayes, Wang, Yates (eds.), May 9-10, 2023, Online, https://ala2023.github.io/. 2023.

with full observability, opening up a novel interesting route to support cooperation when incentives are not fully aligned.

2 RELATED WORK

Our work builds on previous research in the fields of emergent communication and social dilemmas within multi-agent reinforcement learning.

Emergent Communication. Seminal works in the emergent communication field were published by Foerster et al [16, 17], which studied a sequential decision-making problem in a partially observable multi-agent system. Later, [43] used a continuous communication protocol to solve cooperative tasks, and [42] was the first to work on mixed and competitive tasks. In subsequent years the field of emergent communication took off — with an overview presented in [23, 26] — and many papers focused on solving referential games, where a pair of sender and receiver agents have to learn a communication protocol to solve cooperative tasks [13, 18, 27, 28].

Most of the work concerning the emergence of communication focused on cooperative scenarios. Nonetheless, the study of mixed scenarios remains highly relevant, since operating in real-world settings often includes contexts that are not completely cooperative. One of the papers focusing on this setting is the one by Cao et al. [5], studying a bargaining negotiating scenario in which agents use emergent communication to split a common pool of items. This work explores how communication aids prosocial agents, but neglects the self-interested case. A handful of papers focus on emergent communication among competing teams of agents. An example is the work of Liang et al. [32], where authors analyze emergent communication in the game "Task, Talk & compete". Similar work has been developed by Vanneste et al. [46], where the focus is the prey-predator game, with the addition of an (emergent) communication channel among predators; as well as by Brandizzi et al. [4], where authors explore the role of emergent communication in the social-deduction Werewolf game, as a tool to improve the performance of the playing teams. [35] tackled a mixed cooperative-competitive signaling game. Their setting defines a non-situated game with variable amount of cooperation. These characteristics make this setting similar to ours; however the authors study the results of training the agents separately in different environments with different incentive alignments, while we focus on learning concurrently on a set of environments that present these differences. Moreover, referential tasks are usually restricted to two agents, while social dilemmas are scalable to any number agents.

Social Dilemmas. As Kollocks defines them in 1998, "Social Dilemmas are situations where individual rationality leads to collective irrationality" [24], and are characterized by the presence of suboptimal equilibria. Social dilemmas have been studied in game theory, psychology and economics. Researchers have focused on how to solve social dilemmas avoiding suboptimal competitive equilibria and increase the cooperativeness of the outcome. A common outcome of these studies is the positive effect of communication on the cooperation level of the agents [9, 24, 25]. Communication among players has also been studied from a game theoretic perspective: in [6] the authors focused on information exchange in a sender-receiver signaling game, and conclude that perfect communication occurs only when agents' goals are perfectly aligned, and the more the incentives of the agents are misaligned, the more it declines.

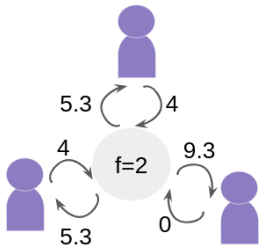


Figure 1: Representation of the Public Goods Game with three players and multiplication factor $f = 2$.

The effect of uncertainty with respect to the payoffs in the public goods game has been studied extensively in literature [3, 11, 15, 30]: different works show that, in this case, the contributions to the public good are significantly lowered. Social dilemmas have been studied also from the reinforcement learning perspective. Much literature focuses in sequential social dilemmas, which are temporally-extended games with game theoretic payoff matrices. [10, 29] try to learn cooperation in multi-agent social dilemmas, without the use of communication. O'Callaghan et al. [36] use multi-objective reinforcement learning to tune the cooperative-competitive behavior of agents. In [20] authors implement fairness norms to solve dilemmas. In [21] authors use influence rewards to enhance coordination and communication among agents. The public good game has been previously studied only in [34], where authors implemented a simulator of reinforcement learning agents playing the game.

3 PRELIMINARIES

In this section we define the Public Goods Game and discuss its equilibria, and introduce our proposed extension.

3.1 The Public Goods Game

A public good is an asset that can benefit all individuals in a social group, both the ones that participated in its production and the ones who did not [24]. Given this definition, the strategy a rational agent should adopt (i.e., an agent whose only goal is to maximize its own earning) is profiting from the public good without investing on it — also called free-riding. This strategy is driven by the goal of maximizing the utility of the individual, and also reinforced by the fear that not enough of the other participants will cooperate to the public good, in which case the individual would end up losing all or part of its endowment [24]. However, if every individual follows the rational strategy the public resource would not be created, so no-one can benefit of it. In this game the incentives of the players are neither perfectly aligned nor misaligned. Using the terminology of [41], we refer to this type of interaction where the utilities of the players are affected by the presence of partial conflict and interdependence as "mixed-motive".

Definition 3.1 (Public Goods Game). A Public Goods Game (PGG) is a tuple $\langle N, \boldsymbol{c}, A, f, \boldsymbol{u} \rangle$, where N denotes the set of players, and

$|N| = n \in \mathbb{N}$ is the number of players; $\boldsymbol{c} = (c_1, \ldots, c_n)$ with $c_i \in \mathbb{R}$, for agents $i \in 1, \ldots, n$ is the profile of endowments, that is the amount of coins that each agent possesses; $A = \{C, D\}$ is the set of actions, that is, every player can either cooperate (investing the full endowment in the public good) or defect (no investment); we define $\boldsymbol{a} = (a_1, \ldots, a_n) \in A$ as the action profile representing actions chosen by each agent. $f \in \mathbb{R}^{\geq 0}$ is the multiplication factor, that is, the factor by which the collective investment is multiplied to generate the public good. $\boldsymbol{u}$ is the vector of utilities the agent receive after acting, where the utility for agent i is a function $u_i : A^n \times \mathbb{R}^{\geq 0} \times \mathbb{R}^n \to \mathbb{R}$, defined as:

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

where a_j denotes the j-th entry of $\boldsymbol{a}$ and $I(a_j)$ is the indicator function, equal to 1 if the action of the agent j is cooperative, and 0 otherwise.

We define the strategy s_i of player i as a probability distribution over her set of actions A: $s_i \in S$, where $S = \Delta(A)$ is the set of all possible strategies, and $\Delta(A)$ is the set of probability distributions over the set of actions. A strategy profile $\boldsymbol{s}$ is then a tuple containing the chosen strategies of all the agents $\boldsymbol{s} = (s_1, \ldots, s_n) \in S^n$. A strategy is called pure when the player has probability 1 of taking a specific action, and is called mixed otherwise. We refer to a profile $\boldsymbol{s}$ where each agent selects C (respectively, D) with probability 1 as the *fully cooperative* (respectively, *fully competitive*) profile. We can then define the expected utility on the profile vector s as:

$$u_i(\boldsymbol{s}, f, \boldsymbol{c}) = \sum_{a \in A} u_i(a, f, \boldsymbol{c}) \prod_{j=1}^{n} s_j(a_j) \tag{2}$$

Hereafter, where the values f and $\boldsymbol{c}$ are fixed, we call $u(\boldsymbol{s}) = u(\boldsymbol{s}, f, \boldsymbol{c})$.

Definition 3.2 (Domination). Given two strategies $s_i, {s_i}' \in S$ for player i, and the set of strategy profiles for all the other players S_{-i}, with $s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$, we can say that:

- s_i dominates s_i' if $u((s_i, s_{-i})) \geq u((s_i', s_{-i}))$, and $\exists\ s_{-i}^{\star} \in S_{-i}$ for which $u((s_i, s_{-i}^{\star})) > u((s_i', s_{-i}^{\star}))$
- s_i weakly dominates s_i' if $u((s_i, s_{-i})) \geq u((s_i', s_{-i}))\ \forall s_{-i} \in S_{-i}$

A strategy (weakly) dominates all the other strategies for an agent if it is (weakly) dominant [31].

Definition 3.3 (Pareto Optimality). Strategy profile $\boldsymbol{s}$ dominates profile $\boldsymbol{s}'$ if, $\forall i \in N$, $u_i(\boldsymbol{s}) \geq u_i(\boldsymbol{s}')$ and there exists $i \in N$ such that $u_i(\boldsymbol{s}) > u_i(\boldsymbol{s}')$. A strategy profile $\boldsymbol{s}^*$ is Pareto optimal if there exists no strategy profile $\boldsymbol{s}$ such that $\boldsymbol{s}$ dominates $\boldsymbol{s}^*$.

Definition 3.4 (Dominant Strategy Equilibria). A strategy profile $\boldsymbol{s}$ is a (weakly) *dominant strategy equilibrium* if $\forall i \in N$, s_i is a (weakly) dominant strategy for i, that is: $u_i(\boldsymbol{s}) \geq u_i((s_i, \boldsymbol{s}_{-i}))\ \forall s_i \in S$.

We can characterize equilibria in PGG based on the value of the multiplication factor f, as follows:

Proposition 1. *For any PGG $\langle N, \boldsymbol{c}, A, f, \boldsymbol{u} \rangle$:*

- *If $f > n$, then the fully cooperative profile is the only dominant strategy equilibrium. Such profile is, furthermore, Pareto optimal;*
- *If $1 < f \leq n$, then the only Pareto optimal profile is the fully cooperative one. The fully competitive profile is one of the weakly dominant strategy equilibria (when considering only two players, then both the fully cooperative and the fully competitive profiles are weakly dominant strategy equilibria);*
- *if $0 \leq f \leq 1$, the fully competitive profile is the only dominant strategy equilibrium. Such equilibrium is, furthermore, Pareto optimal.*

The proposition follows from the following simple observations. The best-case scenario for agent i are profiles where all other players invest, generating a public good of value at least $f \cdot c \cdot (n-1)$. In such case, if $f > n$, then the dominant strategy for i is to contribute by Equation (1). If $f \leq n$, then it is (weakly) better for i to defect, and this remains to be the case through to the worst-case scenario in which every other player defects. However, public goods whose value exceed their costs can be produced whenever $f > 1$. In such cases, Pareto optimality demands full cooperation. An example of Proposition 1 is given in Figure 2, which illustrates the normal form representation of the game for two players using the multiplication factors $f = \{0.5, 1.5, 2, 3.5\}$.

3.2 The Extended Public Goods Game

The classic Public Goods Game studied in game theory and behavioral economics focuses on the interval in which the multiplication factor is bigger than 1 and smaller than the number of players $1 < f < N$. However, in this work we want our agents to be able to cope with a spectrum of environments that goes from cooperative, to competitive, and the mixed. Therefore we extended the interval of allowed multiplication factors to $0 < f < R_+$, where $R_+ > N$ is an arbitrary value, to enable the agents to play in cooperative settings too. The range $f < 1$ allows us to face competitive scenarios as well. We call this setting as the Extended Public Goods Game (EPGG). Varying the multiplication factor, we are defining a set of environments on which we train our RL agents. For every interaction, a different environment is sampled from the set of available ones. In these environments, the agents have access to the information regarding the amount of coins they are endowed with and the value of the multiplication factor. With these observations, we expect agents to learn to cooperate when the multiplication factor is bigger than the number of players, and to learn to defect with a certain probability when the value is smaller. In particular, agents should learn to defect with probability 1 whenever $f = 0$.

The main objective of our work is to study the effect of emergent communication on uncertain observations when agents are trained on a set of cooperative, competitive and mixed environments. In particular, we want to observe the effects of these ingredients when the uncertainty is imposed on the degree of alignment of incentives. For this purpose, we introduce the possibility of observing the multiplication factor with uncertainty.

4 METHODS

In this section we describe the framework we adopt to tackle the learning process in the EPGG setting, i.e. Multi-Agent Reinforcement Learning, and the architectures we use, dwelling on the implementation of how agents deal with uncertainty and the communication protocol.

$f = 0.5$

Player Y \ Player X	C	D
C	2, 2	1, 5
D	5, 1	4, 4

$f = 1.5$

Player Y \ Player X	C	D
C	6, 6	3, 7
D	7, 3	4, 4

$f = 2$

Player Y \ Player X	C	D
C	8, 8	4, 8
D	8, 4	4, 4

$f = 3.5$

Player Y \ Player X	C	D
C	14, 14	7, 11
D	11, 7	4, 4

Figure 2: Normal form games for two players with 4 coins each.

4.1 Multi-Agent Reinforcement Learning

When mapping the EPGG to the MARL framework, we structure the interactions in the following manner. At the beginning of each episode t (i.e., interaction), a multiplication factor f_t is sampled from a given set $F = \{f_0, \ldots, f_K\}$ of factors, where K is the total number of available values (for additional information see Section 5). The state of agent i (with $i = 0, \ldots, n$), is defined by the sampled multiplication factor and the endowment: $s_{i,t} = (f_t, c_{i,t})$.[1] Under no uncertainty, the agents receive the precise value of the multiplication factor, otherwise they observe f_t with added noise (Section 4.2) i.e. $\hat{f}_t$. We study distinctly two scenarios: the one without communication among agents, and the one with. We implemented both the communication policy π_C and the action policy π_A of the agents as multi-layered perceptrons with one or two hidden layers, and *tanh* nonlinearities. Agents are trained independently, for 500 epochs, using the REINFORCE algorithm with baseline [47]. The environment is implemented using the PettingZoo library [45], and is available, together with the experiments.[2]

No communication scenario. In this setting, after receiving the observation, all the agents act simultaneously. The input of the action policy is only defined by the observation the agent gets: $\pi_{A_i} : O_i \times A \to [0, 1]$, where O_i is the set of possible observations for agent i. After acting, every agent receives a reward $r_{i,t}$ from the environment, which is equal to the utility function of the EPGG presented in Equation 1, and therefore depends on the current value of the multiplication factor f_t, the endowment $\boldsymbol{c}_t$ and the current joint action of the agents $\boldsymbol{a}_t$: $r_{i,t} = u_{i,t}(\boldsymbol{a}_t, f_t, \boldsymbol{c}_t)$. Since different scenarios bring different rewards, we normalized the rewards between 0 and 1. The normalization of the reward for the agents playing in a specific scenario is computed by dividing the current reward received by the maximum possible reward the agent could have received in that scenario.

Communicative scenario. In this setting, before acting, a subset Z of agents (hence, $0 < |Z| \leq n$) communicate by sending a message sampled from the output layer of their communication network. The observations received from the environment define the input of the communication policy $\pi_{C_i} : O_i \times M \to [0, 1]$ where M is the set of possible messages, whose size is a training hyperparameter. Basing ourselves on previous work in emergent communication, we choose to work with a discrete message set. This choice is based on the idea that a discrete policy could be naturally interfaced with natural language [19, 26]. After receiving the observation from the environment, the communicating agent i send a message $m_{i,t}$.

[1]We note that in our experiments the agents' endowments are all fixed to the same constant value $c_{i,t} = c$, and thus have no effect on the equilibria of the game.
[2]See https://github.com/nicoleorzan/marl-emecom.

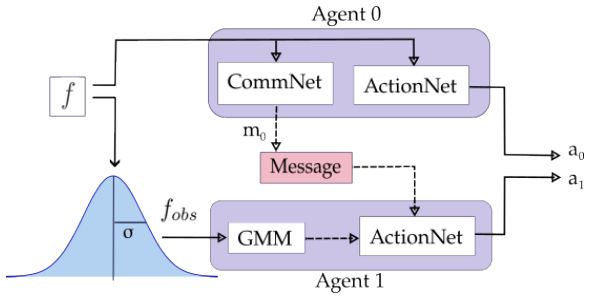


Figure 3: Representation of the setup with two learning agents and communication, where Agent 0 has complete observation of the multiplication factor f, and sends messages to Agent 1 who observes the factor with uncertainty.

The messages of all the communicating agents are concatenated, and represent, together with the observations, the input to the action policy $\pi_{A_i} : O_i \times M^{|Z|} \times A \to [0, 1]$. From this point onwards, the episode proceeds as in the non-communicative scenario. Algorithm 1 presents the learning algorithm employed in the case with communication.

Algorithm 1: Training with communication policies

define: N set of agents; Z set of communicative agents; M message set for every agent; A action set for each agent; O_i observations set for agent i; f sampled multiplication factor; $\boldsymbol{\sigma}$ vector of uncertainties;
$\pi_{C_i} : O_i \times M \to [0, 1]$ communication policy of agent i;
$\pi_{A_i} : O_i \times M^{|Z|} \times A \to [0, 1]$ action policy of agent i;
for $e = 0; e < Epochs; e++$ **do**
  for $b = 0; b < B; b++$ **do** `// batch size`
    $f \leftarrow$ env.sample();
    $\hat{f}_i \leftarrow G(f, \sigma_i)$;
    for $i = 1; i < |N|; i++$ **do**
      if i *in* Z **then**
        $\mathrm{m}_i \leftarrow \pi_{C_i}(\hat{f}_i)$;
    $\boldsymbol{m} \leftarrow$ `concat` (m_i);
    $a_i \leftarrow \pi_{A_i}(\hat{f}, \boldsymbol{m})$;
  update π_{C_i} and π_{A_i} for every agent

4.2 Uncertainty

In the EPGG, uncertainty is introduced as Gaussian noise over the observation of the multiplication factor: the observed $\hat{f}_{i,t}$ is

randomly sampled from the distribution $\hat{f}_{i,t} \sim N(f_t, \sigma_i)$, where f_t is the multiplication factor of the environment at training step t, and σ_i is the uncertainty of agent i.

When uncertainty is present, we allow agents to handle it with or without using a model. Without a model, the noisy observation is directly provided as part of the input of the agents, and we let the neural networks to inherently model the uncertainty. Otherwise, agents keep a probabilistic model of the observed factors via a Gaussian Mixture Model (GMM)[3]: we keep a history of all the multiplicative factors observed during the training, and we use those to fit the GMM. Afterwards, the vector of predicted probabilities for each component of the mixture is provided as part of the input to the agent's network.

Figure 3 depicts the adopted setup in the case of two learning agents with communication, and uncertainty for agent 1.

4.3 Communication

In order to facilitate the agents to use the communication channel we bias the loss function of the communication policy as suggested by Eccles et al. in [12]. To this end, we add a loss term which is minimized when the average entropy of the message policy is high $H(\overline{\pi_{C_i}})$, and the entropy of the message policy conditioned on the input $H(\pi_{C_i}|o_i)$ reaches a target value H_{target} (a hyperparameter):

$$L_{ps}(\pi_{C_i}, o_i) = -\mathbb{E}(\lambda H(\overline{\pi_{C_i}}) - (H(\pi_{C_i}|o_i) - H_{target})^2), \quad (3)$$

where $H(\overline{\pi^i_C})$ is estimated over batches of messages during training. To bias for positive listening we add a term to the loss of the action policy, which is maximized when the actions of an agent are highly influenced by the messages they receive. We do this by computing the divergence between the agent's policy conditioned on the received messages m_t, and the unconditioned one, where the message is replaced by the empty vector $\mathbf{0}$:

$$L_{pl}(\pi_{A_i}, o_i, m_i) = -\sum_{a \in A^i} |\pi^i_A(a|o_i, m_i) - \pi^i_A(a|o_i, \mathbf{0})| \quad (4)$$

In order to quantify the information exchanged by the agents, we implement measures to detect signaling and listening behaviors, following the methods proposed by [33]. In particular we measure the mutual information, which quantifies the correlation between messages and actions. Given a message policy and an action policy, the mutual information between messages and actions is:

$$MI = \sum_{a \in A} \sum_{m \in M} p(a, m) \log \frac{p(a, m)}{p(a)p(m)}, \quad (5)$$

where A is the set of actions available to the agents, and M is the set of messages. Probabilities are computed empirically during training, averaging over the messages and actions of the epoch. If messages and actions come from the same agent, this measure takes the name of *speaker consistency*, and allows us to determine how much the actions and the messages of the agent are aligned. If instead they come from different agents, we are observing how much the messages send by one agent influence the actions taken by another. In this case, the metric takes the name of *instantaneous coordination*, and it measures positive listening for the agent that takes the actions [33].

[3]We implement the GMM module using Scikit-Learn's Gaussian Mixture class [39].

5 RESULTS

We want to observe the behavior of the system when agents are trained on a pre-specified set of multiplication factors that contains values defining cooperative, competitive and mixed situations. The values have been chosen so that the expected utility is equal under either the cooperative or competitive actions, given that the underlying game is not known to the agents during their interactions. Therefore the dominant strategy ex-ante is to select among the cooperative or defective actions uniformly at random.

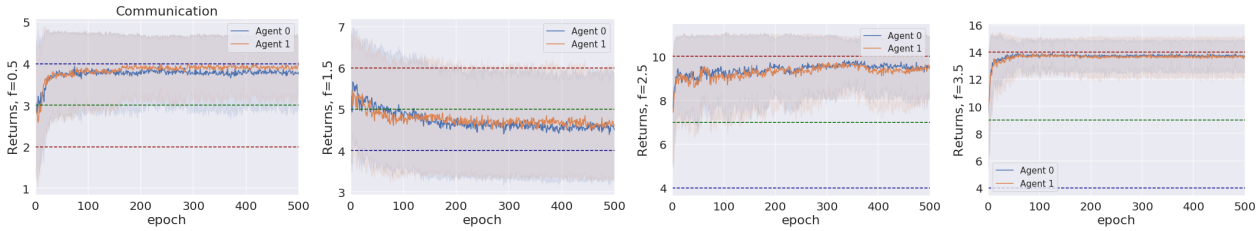
The value of the endowment is fixed to 4 coins for all the agents. In every scenario involving uncertainty, experiments are run both with and without the GMM module for the uncertain agents. The results shown come from averages of 100 experiments on each considered scenario.

5.1 Two-Player Games

In the scenario consisting of two-players, we define the following set of multiplication factors: $F = \{0.5, 1.5, 2.5, 3.5\}$. In this game, for rational agents it is dominant to defect when $f \in \{0.5, 1.5\}$, and to cooperate when $f \in \{2.5, 3.5\}$. Hereafter we refer to $f \in \{2.5, 3.5\}$ as cooperative games, to $f = 0.5$ as the competitive games and to $f = 1.5$ as the mixed-motive one. We perform experiments for three different training scenarios: 1) both agents have full observability of the environment; 2) one agent receives uncertain observations of the multiplication factor; 3) both agents receive uncertain observations on the multiplication factor. Table 1 summarizes the probability of cooperation resulting from the different experiments.Below we discuss the main findings, and we show the plots representing the returns for the agents in the different scenarios. In these plots we also add three baselines, displayed as horizontal lines: the returns the agents would obtain if they always cooperate (dashed red line), always defect (dashed blue line), cooperate or defect with probability 0.5 (dashed green line).

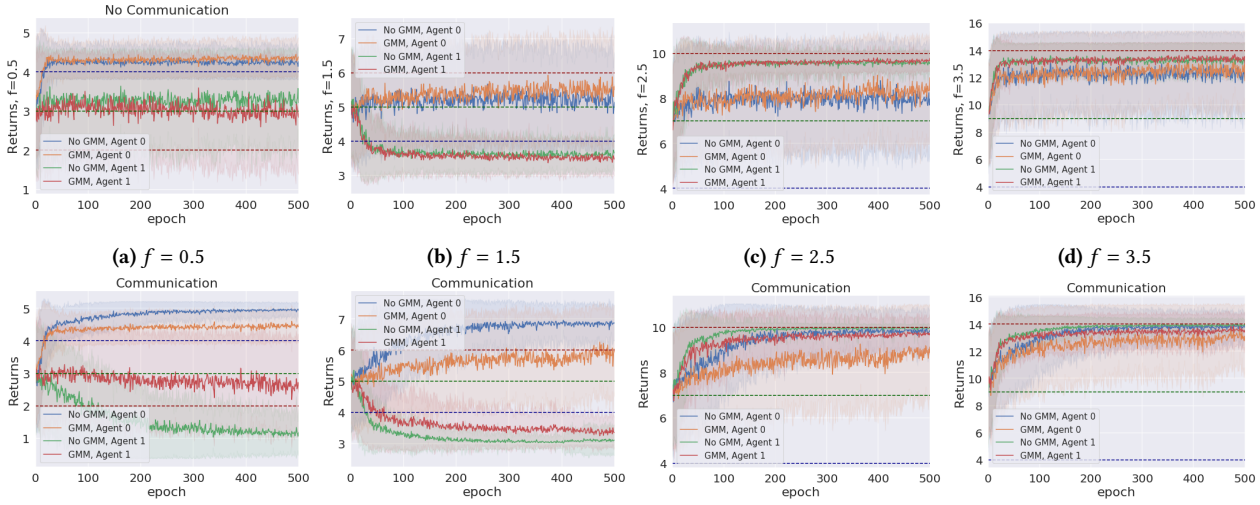
Figure 4 shows the returns during training in the scenario with no uncertainty and with communication, grouped per multiplication factor. In the communication scenario, we allow both agents to send and receive messages. The results for the setting with and without communication are the consistent for the cooperative and competitive games: in the cooperative games the agents converge to cooperation, which is the strictly dominant strategy equilibria and Pareto optimal profile, and allows to maximize the benefit of the whole group. The same happens when $f = 0.5$, where as expected the agents converge to defection. When faced with the mixed-motive game, the non-communicating agents converge to the defection as well, while in the communication scenario the probability to cooperate reaches 0.32 (i.e., they avoid full defection, but fail to reach the Pareto optimal behavior): here the social welfare of the group (i.e. the sum of returns over all the agents) increases over the dominant strategy profile.

Figure 5 depicts the returns during training in the scenario where one agent is uncertain (agent 1), and no communication is allowed. The uncertainty value is fixed to $\sigma = 2$. As expected, adding uncertainty over the observations of one agent worsens its performance:



(a) $f = 0.5$ **(b)** $f = 1.5$ **(c)** $f = 2.5$ **(d)** $f = 3.5$

Figure 4: The returns during training for two agents and no uncertainty, in the communication setting.



(a) $f = 0.5$ **(b)** $f = 1.5$ **(c)** $f = 2.5$ **(d)** $f = 3.5$

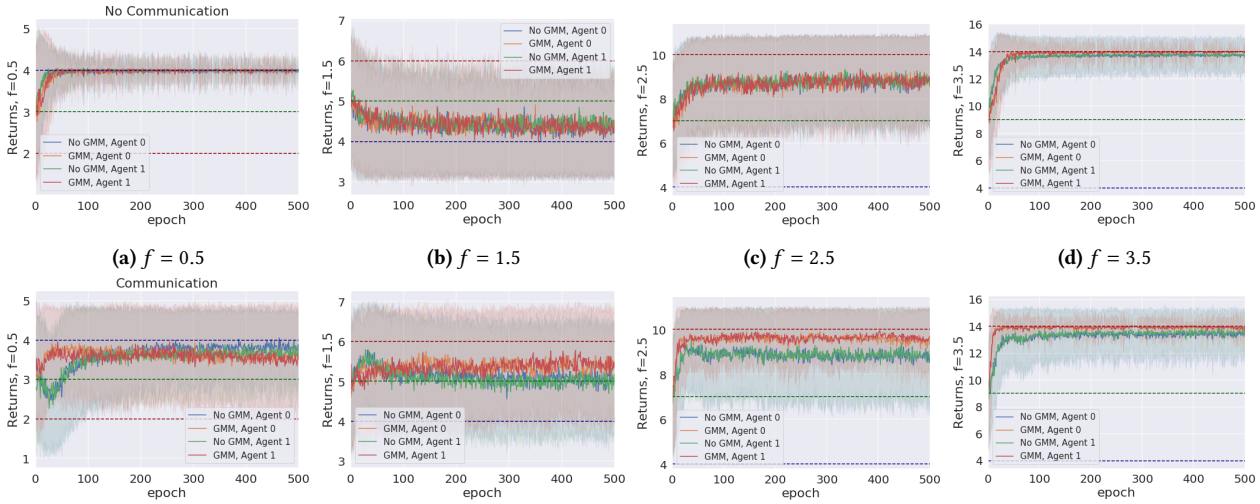
(e) $f = 0.5$ **(f)** $f = 1.5$ **(g)** $f = 2.5$ **(h)** $f = 3.5$

Figure 5: The return during training, in the scenario with one uncertain agent (agent 1, with $\sigma = 2$), in the non-communication case (upper row) vs communication case (lower row).

the agent cannot disentangle the situations in which it is preferred to cooperate from the ones in which it is preferred to defect. Modeling uncertainty in this case does not change the outcome. The certain agent (agent 0), as expected, keeps converging to the rational behaviors. We introduce a unidirectional communication step in this scenario, where the certain agent is allowed to send messages to the uncertain one. We notice the following effects: 1) in the cooperative games, communication helps the uncertain agent to recover the optimal strategy, and act cooperatively with probability 0.97; in this scenario modeling uncertainty does not provide any additional benefit; 2) in the competitive and mixed games, communication allows the certain agent to exploit the uncertain agent, mainly in the case in which no GMM module is used. We argue that explicitly modelling the uncertainty of the environment can provide the agent with additional information, and it is therefore less likely to be deceived.

This second claim is supported as well by Figure 7, that shows the trend of the instantaneous coordination and speaker consistency (Equation 5) during training (in the three scenarios). We observe that the instantaneous coordination is higher when one agent is uncertain and no modeling over uncertainty is allowed (Fig. 7b): here, the only way to recover information is to listen to the received messages and act accordingly. However, the uncertain agent gets easily deceived in this way, as shown by the speaker consistency metric (i.e., agent 0 is not acting in line to its own messages).

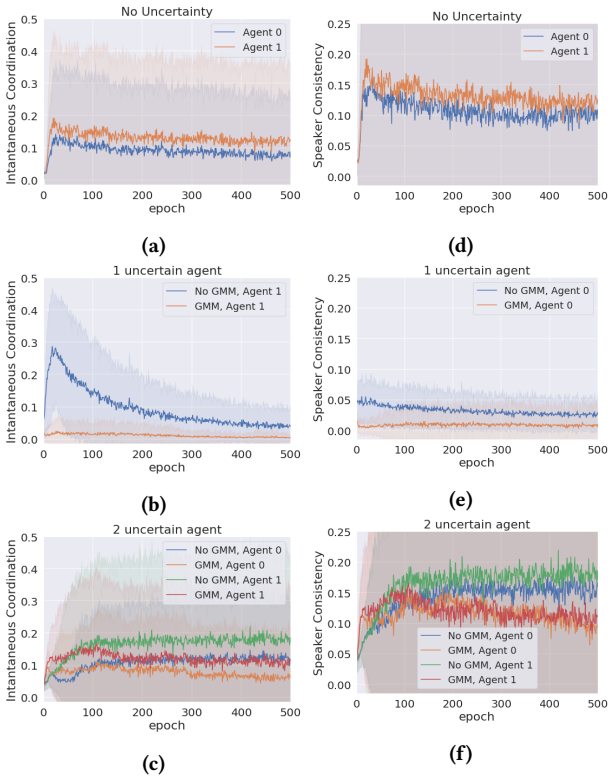
The returns during training for the scenario with two uncertain agents (and $\sigma = 0.5$) are shown in Figure 6. Again, adding uncertainty lowers the ability of the two agents to distinguish the situations in which it is rational to cooperate or defect. This happens especially when $f \in \{1.5, 2.5\}$ since here the introduction of uncertainty makes it more likely that the agents believe they are playing a different game that the true one. We introduce a bidirectional communications step in this scenario: here both agents send and receive messages. We observe that the main effect of communication is to improve the cooperation in the two aforementioned intermediate environments, while it has almost no impact on the other two. However, when $f \in \{1.5, 2.5\}$, it is the combined effect of communication and uncertainty modeling that has the highest impact in aiding cooperation. In particular, in the mixed motive scenario, those move the equilibria of the game from complete defection, to cooperation with average probability 0.65, and to 0.92 when $f = 2.5$ (see Table 1). As Figure 7 shows, speaker consistency



(a) $f = 0.5$ **(b)** $f = 1.5$ **(c)** $f = 2.5$ **(d)** $f = 3.5$

(e) $f = 0.5$ **(f)** $f = 1.5$ **(g)** $f = 2.5$ **(h)** $f = 3.5$

Figure 6: The returns during training, in the scenario with two uncertain agents ($\sigma = 0.5$), in the non communication case (upper row) vs communication case (lower row).



(a) **(d)**

(b) **(e)**

(c) **(f)**

Figure 7: Instantaneous coordination for the cases: full communication and no uncertainty (top row), one uncertain agent (central row), two uncertain agents (bottom row).

signals that in this setting both agents send reliable messages (Fig. 7f). The instantaneous coordination shows a non-zero amount of information transfer as well (Fig. 7c). We also observe how the agents cooperate with higher probability in the mixed motive scenario, when uncertainty and communication are present (the average cooperation probability is 0.65), with respect to the scenario with full observability (average cooperation probability 0.32). This effect is worth investigating further, and will be part of our future work.

5.2 Three-Players Games

We run experiments for the three-players scenario, for which we define the following set of multiplication factors: $F = \{0.5, 1.5, 2.5, 3.5, 4.5, 5.5\}$. This values are chosen following the criterion defined at the beginning of Section 5. In this setting, the dominant strategy for rational agents is to defect when $f \in \{0.5, 1.5, 2.5\}$, and cooperate when $f \in \{3.5, 4.5, 5.5\}$. Here, we refer to $f \in \{3.5, 4.5, 5.5\}$ as cooperative games, to $f \in \{1.5, 2.5\}$ as mixed-motive ones and to $f = 0.5$ as the competitive one. We perform experiments for three different scenarios: 1) agents have no uncertainty on f; 2) two agent receives uncertain observations of f; 3) three agents receive uncertain observations of f.

The experiments on these three setting are consistent with the two players scenarios. The experiments on the first setting result in the exact same behavior that we find for two-agents and no uncertainty: agents converge to cooperation in the cooperative games and to defection in the competitive one, while in the mixed-motive ones cooperation is almost zero when no communication is included, and this increases when agents communicate (see Table 2). When two agents are uncertain ($\sigma = 2$), again they struggle to distinguish when it is convenient to cooperate from when it is convenient to defect. This uncertainty slightly affects the certain agent as well, which can no longer perfectly converge to the dominant strategies. Adding the communication step (from the certain agent to the uncertain ones) allows the certain agent to deceive the others when

Cooperation probabilities for the two-agents scenarios: non-communication (NC) and communication (C), with uncertainty modeling (UM) and without, averaged over 80 runs. When more agents have the same characteristics, the average measure is reported. The bold values highlight the principal outcomes.

Table 1: Two agents

Agents' Cooperation Probabilities	Multiplicative Factor							
	f=0.5		f=1.5		f=2.5		f=3.5	
	NC	C	NC	C	NC	C	NC	C
Both Certain (avg)	0.00	0.08	0.02	**0.32**	1.00	0.93	1.00	0.95
Certain Agent	0.00	0.09	0.01	0.17	1.00	0.89	1.00	0.90
Uncertain Agent ($\sigma = 2$)	0.17	**0.95**	0.36	**0.96**	0.61	0.97	0.80	0.97
Certain Agent	0.00	0.00	0.00	0.03	1.00	0.97	1.00	0.98
Uncertain Agent, UM ($\sigma = 2$)	0.36	**0.47**	0.54	**0.64**	0.68	0.77	0.84	0.85
Both Uncertain (avg) ($\sigma = 0.5$)	0.00	0.10	0.20	0.54	0.76	0.78	0.97	0.93
Both Uncertain, UM (avg) ($\sigma = 0.5$)	0.00	0.18	0.16	**0.65**	0.76	0.92	0.98	0.99

Table 2: Three agents

Agents' Cooperation Probabilities	Multiplicative Factor											
	f=0.5		f=1.5		f=2.5		f=3.5		f=4.5		f=5.5	
	NC	C	NC	C	NC	C	NC	C	NC	C	NC	C
All Certain (avg)	0.00	0.10	0.00	**0.16**	0.11	**0.21**	0.95	0.33	1.00	0.68	1.00	0.73
Certain Agent	0.11	0.00	0.12	0.02	0.24	0.14	0.60	0.78	0.66	0.97	0.67	0.97
Uncertain Agents (avg) ($\sigma = 2$)	0.32	**0.54**	0.39	**0.58**	0.39	**0.60**	0.47	0.61	0.48	0.66	0.55	0.68
Certain Agent	0.15	0.10	0.15	0.11	0.25	0.19	0.62	0.56	0.65	0.65	0.65	0.65
Uncertain Agents, UM (avg) ($\sigma = 2$)	0.33	0.77	0.40	0.81	0.44	0.82	0.53	0.86	0.58	0.86	0.69	0.86
All Uncertain (avg) ($\sigma = 1$)	0.10	0.36	0.15	0.39	0.32	0.46	0.55	0.57	0.62	0.67	0.68	0.71
All Uncertain, UM (avg) ($\sigma = 1$)	0.01	0.73	0.07	**0.75**	0.37	**0.78**	0.63	0.80	0.88	0.81	0.97	0.82

$f \in \{0.5, 1.5, 2.5\}$: in these three scenarios, while the certain agent's strategy converges to defection, the others cooperate with higher probabilities in comparison to the cases where no communication is involved. In this case, explicitly modelling uncertainty using the GMM module improves cooperation in all the games.

When all the three agents are uncertain ($\sigma = 1$) the observed effect is the same as in the two uncertain agents scenario. Uncertainty decreases the ability of agents to distinguish the situations in which it is rational to cooperate or defect. This effect is evident especially when $f \in \{1.5, 2.5, 3.5\}$. The combined effect of communication and the GMM model helps to improve cooperation considerably in these three games, as well as in the competitive one. Moreover, we observe again that the agents cooperate with higher probability in the mixed motive scenarios, when uncertainty and communication are present, with respect to the scenario with full observability.

6 CONCLUSION

In this paper we investigated the effects of emergent communication on independent learning agents trained on a spectrum of environments with different incentive alignments, and in the presence of uncertainty. We observed how learning agents with the same amount of uncertainty can use the combined effect of communication and uncertainty modeling to improve the cooperation of the whole group, overcoming the defection equilibria of mixed-motive scenarios. This resulted in the improvement of the social welfare of the group. Moreover, we showed that in asymmetric uncertainty cases communication can be employed by the certain agents to deceive the others. We believe that the employment of emergent communication in mixed cooperative-competitive multi-agent systems provides a good boilerplate for the development of reliable artificial agents, and therefore reliable human-AI communication.

As future work, we plan to investigate the learning dynamics in larger settings, investigate the effect of social structures (e.g., reputation and norms [24]) and different communication frameworks (e.g., graph neural networks [22]). Our long term goal is to work towards the development of hybrid scenarios (i.e., involving both human and artificial agents).

REFERENCES

[1] Zeynep Akata, Dan Balliet, Maarten De Rijke, Frank Dignum, Virginia Dignum, Guszti Eiben, Antske Fokkens, Davide Grossi, Koen Hindriks, Holger Hoos, et al. 2020. A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer* 53, 08 (2020), 18–28.

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

[3] Anders Biel and Tommy Gärling. 1995. The role of uncertainty in resource dilemmas. *Journal of Environmental Psychology* 15, 3 (1995), 221–233. https://doi.org/10.1016/0272-4944(95)90005-5 Green Psychology.

[4] Nicolo' Brandizzi, Davide Grossi, and Luca Iocchi. 2021. RLupus. *Intelligenza Artificiale* 15, 2 (2021), 55–70.

[5] Kris Cao, Angeliki Lazaridou, Marc Lanctot, Joel Z. Leibo, Karl Tuyls, and Stephen Clark. 2018. Emergent Communication through Negotiation. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net.

[6] Vincent P Crawford and Joel Sobel. 1982. Strategic information transmission. *Econometrica: Journal of the Econometric Society* (1982), 1431–1451.

[7] Allan Dafoe, Yoram Bachrach, Gillian Hadfield, Eric Horvitz, Kate Larson, and Thore Graepel. 2021. Cooperative AI: machines must learn to find common ground.

[8] Allan Dafoe, Edward Hughes, Yoram Bachrach, Tantum Collins, Kevin R. McKee, Joel Z. Leibo, Kate Larson, and Thore Graepel. 2020. Open Problems in Cooperative AI. https://doi.org/10.48550/ARXIV.2012.08630

[9] Robyn M Dawes, Jeanne McTavish, and Harriet Shaklee. 1977. Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of personality and social psychology* 35, 1 (1977), 1.

[10] Enrique Munoz de Cote, Alessandro Lazaric, and Marcello Restelli. 2006. Learning to cooperate in multi-agent social dilemmas. In *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*. 783–785.

[11] David L Dickinson. 1998. The voluntary contributions mechanism with uncertain group payoffs. *Journal of economic behavior & organization* 35, 4 (1998), 517–533.

[12] Tom Eccles, Yoram Bachrach, Guy Lever, Angeliki Lazaridou, and Thore Graepel. 2019. Biases for emergent communication in multi-agent reinforcement learning. *Advances in neural information processing systems* 32 (2019).

[13] Katrina Evtimova, Andrew Drozdov, Douwe Kiela, and Kyunghyun Cho. 2018. Emergent communication in a multi-modal, multi-step referential game.

[14] Joseph Farrell and Matthew Rabin. 1996. Cheap Talk. *Journal of Economic Perspectives* 10, 3 (September 1996), 103–118. https://doi.org/10.1257/jep.10.3.103

[15] Urs Fischbacher, Simeon Schudy, and Sabrina Teyssier. 2014. Heterogeneous reactions to heterogeneity in returns from public goods. *Social Choice and Welfare* 43, 1 (2014), 195–217.

[16] Jakob Foerster, Ioannis Alexandros Assael, Nando De Freitas, and Shimon Whiteson. 2016. Learning to communicate with deep multi-agent reinforcement learning. *Advances in neural information processing systems* 29 (2016).

[17] Jakob N Foerster, Yannis M Assael, Nando de Freitas, and Shimon Whiteson. 2016. Learning to Communicate to Solve Riddles with Deep Distributed Recurrent Q-Networks. (2016).

[18] Serhii Havrylov and Ivan Titov. 2017. Emergence of language with multi-agent games: Learning to communicate with sequences of symbols. *Advances in neural information processing systems* 30 (2017).

[19] Charles F Hockett and Charles D Hockett. 1960. The origin of speech. *Scientific American* 203, 3 (1960), 88–97.

[20] Edward Hughes, Joel Z Leibo, Matthew Phillips, Karl Tuyls, Edgar Dueñez-Guzman, Antonio García Castañeda, Iain Dunning, Tina Zhu, Kevin McKee, Raphael Koster, et al. 2018. Inequity aversion improves cooperation in intertemporal social dilemmas. *Advances in neural information processing systems* 31 (2018).

[21] Natasha Jaques, Angeliki Lazaridou, Edward Hughes, Caglar Gulcehre, Pedro Ortega, Dj Strouse, Joel Z. Leibo, and Nando De Freitas. 2019. Social Influence as Intrinsic Motivation for Multi-Agent Deep Reinforcement Learning. In *Proceedings of the 36th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 97)*, Kamalika Chaudhuri and Ruslan Salakhutdinov (Eds.). PMLR, 3040–3049. https://proceedings.mlr.press/v97/jaques19a.html

[22] Jiechuan Jiang, Chen Dun, Tiejun Huang, and Zongqing Lu. 2020. Graph Convolutional Reinforcement Learning. In *International Conference on Learning Representations (ICLR 2020)*.

[23] Tatsuya Kasai, Hiroshi Tenmoto, and Akimoto Kamiya. 2008. Learning of communication codes in multi-agent reinforcement learning problem. In *2008 IEEE Conference on Soft Computing in Industrial Applications*. 1–6. https://doi.org/10.1109/SMCIA.2008.5045926

[24] Peter Kollock. 1998. Social Dilemmas: The Anatomy of Cooperation. *Annual Review of Sociology* 24 (1998), 183–214. http://www.jstor.org/stable/223479

[25] Sandeep Krishnamurthy. 2001. Communication effects in public good games with and without provision points. In *Research in Experimental Economicss*. Emerald Group Publishing Limited.

[26] Angeliki Lazaridou and Marco Baroni. 2020. Emergent Multi-Agent Communication in the Deep Learning Era. *CoRR* abs/2006.02419 (2020). arXiv:2006.02419 https://arxiv.org/abs/2006.02419

[27] Angeliki Lazaridou, Karl Moritz Hermann, Karl Tuyls, and Stephen Clark. 2018. Emergence of Linguistic Communication from Referential Games with Symbolic and Pixel Input. In *International Conference on Learning Representations*. https://openreview.net/forum?id=HJGv1Z-AW

[28] Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. 2017. Multi-Agent Cooperation and the Emergence of (Natural) Language. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net. https://openreview.net/forum?id=Hk8N3Sclg

[29] Joel Z Leibo, Vinicius Zambaldi, Marc Lanctot, Janusz Marecki, and Thore Graepel. 2017. Multi-agent Reinforcement Learning in Sequential Social Dilemmas. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. 464–473.

[30] M Vittoria Levati, Andrea Morone, and Annamaria Fiore. 2009. Voluntary contributions with imperfect information: An experimental study. *Public Choice* 138, 1 (2009), 199–216.

[31] Kevin Leyton-Brown and Yoav Shoham. 2008. Essentials of game theory: A concise multidisciplinary introduction. *Synthesis lectures on artificial intelligence and machine learning* 2, 1 (2008), 1–88.

[32] Paul Pu Liang, Jeffrey Chen, Ruslan Salakhutdinov, Louis-Philippe Morency, and Satwik Kottur. 2020. On Emergent Communication in Competitive Multi-Agent Teams. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*. 735–743.

[33] Ryan Lowe, Jakob Foerster, Y-Lan Boureau, Joelle Pineau, and Yann Dauphin. 2019. On the Pitfalls of Measuring Emergent Communication. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. 693–701.

[34] U ManChon and Zhen Li. 2010. Public goods game simulator with reinforcement learning agents. In *2010 Ninth International Conference on Machine Learning and Applications*. IEEE, 43–49.

[35] Michael Noukhovitch, Travis LaCroix, Angeliki Lazaridou, and Aaron Courville. 2021. Emergent Communication under Competition. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*. 974–982.

[36] David O'Callaghan and Patrick Mannion. 2021. Tunable behaviours in sequential social dilemmas using multi-objective reinforcement learning. In *Proceedings of the 20th international conference on autonomous agents and multiagent systems*. 1610–1612.

[37] Afshin Oroojlooy and Davood Hajinezhad. 2022. A review of cooperative multi-agent deep reinforcement learning. *Applied Intelligence* (2022), 1–46.

[38] Liviu Panait and Sean Luke. 2005. Cooperative multi-agent learning: The state of the art. *Autonomous agents and multi-agent systems* 11 (2005), 387–434.

[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[40] Elizabeth Pennisi. 2009. On the origin of cooperation.

[41] Thomas C Schelling. 1958. The strategy of conflict. Prospectus for a reorientation of game theory. *Journal of Conflict Resolution* 2, 3 (1958), 203–264.

[42] Amanpreet Singh, Tushar Jain, and Sainbayar Sukhbaatar. 2019. Learning when to Communicate at Scale in Multiagent Cooperative and Competitive Tasks. In *International Conference on Learning Representations*.

[43] Sainbayar Sukhbaatar, Rob Fergus, et al. 2016. Learning multiagent communication with backpropagation. *Advances in neural information processing systems* 29 (2016).

[44] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.

[45] J Terry, Benjamin Black, Nathaniel Grammel, Mario Jayakumar, Ananth Hari, Ryan Sullivan, Luis S Santos, Clemens Dieffendahl, Caroline Horsch, Rodrigo Perez-Vicente, et al. 2021. Pettingzoo: Gym for multi-agent reinforcement learning. *Advances in Neural Information Processing Systems* 34 (2021), 15032–15043.

[46] Astrid Vanneste, Wesley Van Wijnsberghe, Simon Vanneste, Kevin Mets, Siegfried Mercelis, Steven Latré , and Peter Hellinckx. 2021. Mixed Cooperative-Competitive Communication Using Multi-agent Reinforcement Learning. In *Advances on P2P, Parallel, Grid, Cloud and Internet Computing*. Springer International Publishing, 197–206. https://doi.org/10.1007/978-3-030-89899-1_20

[47] Ronald J Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning* 8, 3 (1992), 229–256.