Heterogeneity Breaks the Game: Evaluating Cooperation-Competition with Multisets of Agents

Yue Zhao[1], José Hernández-Orallo✉[2,3]

[1] School of Computer Science and Engineering, Northwestern Polytechnical University
zhaoyueplc@163.com
[2] Valencian Research Institute for Artifcial Intelligence (VRAIN), Universitat Politècnica de València
[3] Leverhulme Centre for the Future of Intelligence, University of Cambridge
jorallo@upv.es

Abstract. The value of an agent for a team can vary significantly depending on the heterogeneity of the team and the kind of game: cooperative, competitive, or both. Several evaluation approaches have been introduced in some of these scenarios, from homogeneous competitive multi-agent systems, using a simple average or sophisticated ranking protocols, to completely heterogeneous cooperative scenarios, using the Shapley value. However, we lack a general evaluation metric to address situations with both cooperation and (asymmetric) competition, and varying degrees of heterogeneity (from completely homogeneous teams to completely heterogeneous teams with no repeated agents) to better understand whether multi-agent learning agents can adapt to this diversity. In this paper, we extend the Shapley value to incorporate both repeated players and competition. Because of the combinatorial explosion of team multisets and opponents, we analyse several sampling strategies, which we evaluate empirically. We illustrate the new metric in a predator and prey game, where we show that the gain of some multi-agent reinforcement learning agents for homogeneous situations is lost when operating in heterogeneous teams.

Keywords: Multi-agent reinforcement learning · Cooperation-Competition game · Evaluation.

1 Introduction

The evaluation of how much a member contributes to a team is a key question in many disciplines, from economics to biology, and has been an important element of study in artificial intelligence, mostly in the area of multi-agent systems (MAS). When a homogeneous multi-agent system has to achieve a collaborative goal, evaluation can be based on measuring overall performance under several agent configurations. However, a more general and realistic version of the problem is when teams are heterogeneous, with players behaving differently and

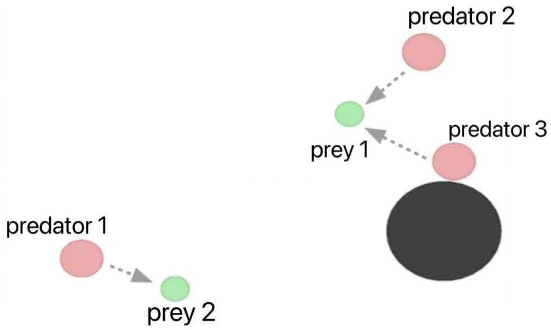

Fig. 1: Predator-prey game using the Multi-agent Particle Environment (MPE) [3,4] where we see 3 predators (in red), 2 preys (in green), and landmarks (in black). With $m = 5$ agents playing in total, and $l = 3$ different kinds of agents to choose from (MADDPG, DDPG and random), the combinations with repetitions of the team sizes configurations $(l_{pred}, l_{prey}) = (4, 1), (3, 2), (2, 3), (1, 4)$ make a total of 45+60+60+45= 210 experiments, and a larger number if we also consider experiments with $m < 5$. Determining which agent has the most contributions to the team considering all roles, and estimating this number with a small number of experiments is the goal of this paper.

reacting in various ways depending on their teammates. The Shapley value [1] is a well-known metric of the contribution of a player to a heterogeneous team taking into account different coalition formations.

Things become more sophisticated in situations where the players are learning agents [2]. Even if some of these agents use the same algorithm, they may end up having different behaviour after training, with important variations when the same episode is re-run. Despite this variability, they still should be considered as 'repeated' players, something that the original Shapley value does not account for well. Finally, and yet more generally, teams may compete against other teams in asymmetric games, and the contribution of each player will depend on the composition of its team but also on the composition of the opponent team, with the same algorithm possibly appearing once or more on one team or both. This is the general situation we address in this paper.

This situation suffers from poor stability in the payoffs when teams are composed of several learning agents: the same algorithm will lead to very different payoffs depending on the configuration of teams [5]. This requires many iterations in the evaluation protocols, which makes each value for a team configuration expensive to calculate. Consequently, it is even more difficult than in other uses of the Shapley value to collect all possible team configurations. As a result, approximations based on sampling become necessary to deal with the huge number of combinations [6].

Motivated by these issues, we present the following contributions. First, we extend the Shapley value to incorporate repeated players and opposing teams:

more technically, the new Shapley value can be applied to cooperative-competitive scenarios when asymmetric teams are multisets of players. Second, we analyse several sampling strategies to approximate this new Shapley value, which we evaluate empirically. Third, we apply this extended Shapley value estimation to a popular asymmetric multi-team multi-agent reinforcement learning (MARL) scenario: predator and prey teams composed of three different kinds of algorithms, which accounts for the heterogeneity of the team. An example of the scenarios we want to evaluate is presented in Fig. 1. We show that some MARL algorithms that work well in homogeneous situations, such as MADDPG, degrade significantly in heterogeneous situations.

The rest of the paper is organised as follows. The following section overviews related work on the evaluation of multi-agent systems, and multi-agent reinforcement learning in particular. Section 3 builds on the original definition of the Shapley value to the extension for multisets and opposing teams (cooperative-competitive), also showing what original properties are preserved. Some sampling methods for approximating this extension are explained in section 4. Section 5 discusses the MARL and single-agent reinforcement learning algorithms and defines the experimental setting. Section 6 covers the experimental results and section 7 closes the paper.

2 Background

The evaluation of competitive and cooperative games is at the heart of game theory, pervading many other disciplines. Let us start the analysis with competitive (non-cooperative) games, for two-player games or in multi-agent games. Nash equilibrium [7] is the most common way to define the solution of a non-cooperative game and is invariant to redundant tasks and games, but discovering the Nash equilibrium is not always easy or possible in a multi-agent system [8]. Some new methods are based on the idea of playing a meta-game, that is, a pair-wise win-rate matrix between N agents, as in [9] and the recently proposed α-Rank method [10], which was shown to apply to general games. These methods are also inspired by early ranking systems used in (symmetric) competitive games, like the Elo score in chess [11], which estimates the strength of a player, based on the player's performance against *some* of the other opponents. With sparse match results and strongly non-transitive and stochastic players, the predictive power of Elo may be compromised, and this gets even worse in multi-agent games with more than two players per game. As a result, other rating systems such as Glicko [12], TrueSkill [13] and Harkness [14] have been proposed. However, these extensions still show problems of consistency [15,16], very sensitive to non-transitivity and high variability of results between matches.

On the other hand, in purely cooperative games, players are organised into a coalition, a group of players that need to cooperate for the same goal. When the team is homogeneous, the evaluation is easy, as n equal copies of an algorithm or policy are evaluated each time. The best policy or algorithm can be selected just by averaging results. However, in heterogeneous teams, we need to determine the contribution of each specific player in a wide range of situations with complex interactions –the attribution problem. The Shapley value [1] has emerged as a key concept in multi-agent systems to determine each agent's contribution. Given all the coalitions and their payoffs, the Shapley value determines the final contribution of each player. Because of the combinatorial explosion in team formations, approximations are required, both to reduce the computational cost [17] but more importantly to reduce the number of experiments to be run or actual games to be played. Still, in cooperative game theory, the Shapley value provides a key tool for analysing situations with strong interdependence between players [18,19].

The general situation when both competition and cooperation need to be evaluated has been present in many disciplines for centuries, from economics to biology, from sports to sociology. It is also increasingly more prevalent in artificial intelligence, with areas such as reinforcement learning introducing better algorithms for cooperative games but also for cooperative-competitive environments. For instance, DDPG is a deep reinforcement learning agent based on the actor-critic framework, with each agent learning the policy independently without considering the influence of other agents. MADDPG [3] is also based on an actor-critic algorithm, but extends DDPG into a multi-agent policy gradient algorithm where each agent learns a centralised critic based on the observations and actions of all agents. This and other methods (e.g., [20]) are illustrated on some testbed tasks showing that they outperform the baseline algorithms. However, this comparison assumes a homogeneous situation (all the agents in the team use the same algorithm). It is unclear whether these algorithms can still operate in heterogeneous situations. In some cases, the algorithms do not work well when the exchange of information only happens for a subset of agents in the coalition, but in many other cases it is simply that the only available metric is an average reward and the problem of attribution reemerges [21].

Finally, things become really intricate when we consider both competition and cooperation, and we assume that teams can be heterogeneous. But this scenario is becoming increasingly more common as more algorithms could potentially be evaluated in mixed settings (cooperation and competition) [3,22,23,24]. It is generally believed that more collaboration always leads to better system performance, but usually because systems are evaluated in the homogeneous case. Are these 'better' agents robust when used in a mixed environment, when they can take different roles (in either team in a competitive game) and have to collaborate with different agents? This is fundamental for understanding how well AI systems perform in more realistic situations where agents have to collaborate with other different agents (including humans). This question remains unanswered because of several challenges: (1) No formalism exists to determine the contribution of each agent —its value— in these (possibly asymmetric) competition-cooperation situations with repeated agents (2) Heterogeneous situations are avoided because any robust estimation requires a combinatorially high number of experiments to evaluate all possible formations. These two challenges are what we address in this paper. We start by extending the Shapley value for competitive games and repeated agents next.

3 Extending the Shapley Value

s

A cooperative *game*(N, v) is defined from a set of $n = |N|$ players, and a characteristic function $v : 2^N \to \mathbb{R}$. If $S \in 2^N$ is a coalition of players (a team), then $v(S)$ is the worth of coalition S, usually quantifying the benefits the members of S can get from the cooperation. The Shapley value of player i reflects its contribution to the overall goal by distributing benefits fairly among players, defined as follows:

$$\varphi_i(v) = \frac{1}{n} \sum_{S \subseteq N \setminus \{i\}} \binom{n-1}{|S|}^{-1} [v(S \cup \{i\}) - v(S)] \tag{1}$$

where $(v(S \cup \{i\}) - v(S))$ is the marginal contribution of i to the coalition S, and $N \setminus \{i\}$ is the set of players excluding i. The combinatorial normalisation term divides by the number of coalitions of size $|S|$ excluding i.

Note that the above expression assumes that the size of the largest team, let us denote it by m, is equal to the number of players we have, n. However, in general, these two values may be different, with $m \leq n$, and a generalised version of the Shapley value is expressed as:

$$\varphi_i(v) = \frac{1}{m} \sum_{j=0}^{m-1} \binom{n-1}{j}^{-1} \sum_{S \subseteq N \setminus \{i\} : |S| = j} [v(S \cup \{i\}) - v(S)] \tag{2}$$

It is now explicit that the marginal contributions are grouped by the size of S, i.e., $|S| = j$. Also, we see that the number of 'marginal contributions' to compute for each φ_i is $r_i = \sum_{j=0}^{m-1} \binom{n-1}{j}$, and for all φ_i in total this is $r = \sum_{j=0}^{m} \binom{n}{j}$. This counts the sets with $\leq m$ elements including $\emptyset$, even if we assume $v(\emptyset) = 0$). For the special case of $n = m$ we have $r = |2^N| = 2^n$, i.e., we have to calculate as many experiments as the power set of N.

3.1 Multisets of Agents

One first limitation of the Shapley value is that coalitions are sets of players. If we have n agents then the coalitions will have sizes up to n. However, a common situation in artificial intelligence is that we can replicate some agents as many times as we want. This decouples the number of agents from the size of the coalitions. For instance, with agents $\{a, b, c\}$ and coalitions up to $m = 4$ agents, we could have coalitions as multisets such as $S_1 = \{a, a, b\}$ or $S_2 = \{b, b, b, d\}$. A straightforward way of extending the Shapley value with multisets is to consider that, if there are l different players and the coalitions are of size m, we can define m 'copies' of each of the l different agents into a new set N=$\{a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4, c_1, c_2, c_3, c_4\}$. With this we end up having $|N| = n = l \cdot m$ agents and no multisets. In the examples above, we would have m=4 and l=3, n=12, with S_1=$\{a_1, a_2, b_1\}$ or S_2=$\{b_1, b_2, b_3, d_1\}$ (actually there are several possible equivalent variants of each of them).

We can now use Eq. 2, but many results should be equivalent, e.g., $v(\{a_1, b_3\})$ =$v(\{a_2, b_3\})$ with all possible variants. Suppose R is the subset of 2^N, where all redundant coalitions have been removed and only a canonical one has been kept. Then Eq. 2 can be simplified into:

$$\varphi_i(v) = \frac{1}{m} \sum_{j=0}^{m-1} \left(\!\!\binom{l}{j}\!\!\right)^{-1} \sum_{S \in R: |S|=j} [v(S \cup \{i\}) - v(S)] \tag{3}$$

where $\left(\!\!\binom{x}{y}\!\!\right)$ denotes the combinations of size y of x elements with repetitions. The derivation simply replaces the combinations of j elements taken from n by the combinations of j elements taken from l=$\frac{n}{m}$ with repetitions. Note that S can now contain i, and we have situations where the marginal contribution is calculated over a coalition S that already has one or more instances of i compared to S with an extra instance of i. Now, the number of required values (or experiments) for each $\varphi_i(v)$ is $r_i = \sum_{j=0}^{m-1} \left(\!\!\binom{l}{j}\!\!\right)$, with a total of

$$r = |R| = \sum_{j=0}^{m} \left(\!\!\binom{l}{j}\!\!\right) \tag{4}$$

For instance, with l=m=3, we have n=9 and we have r=1+3+6+10=20 possible sets. With l=3 and m=4, this would be r=1+3+6+10+15=35. With l=m=4, this would be r=1+4+10+20+35=70. With l=m=5, this would be r=1+5+15+35+70 +126=252.

3.2 Cooperation-Competition Games

The Shapley value was designed for cooperation, so there is only one team, with the same goal and share of the payoff for each agent. However, in situations where there are more than one team competing against each other, several instances of the same type of agents can be part of one or more teams. An agent cooperates with the members of the same team, while different teams compete against them. We extend the Shapley value for this situation. We will work with two opposing teams, but this can be extended to any number of teams.

Consider the two team roles $\{A, B\}$ in a competitive game, e.g., A could be predators and B could be preys. When considering the role A we define the $game^A(v^A, N^A)$, where B is the opponent. Similarly, for $game^B(v^B, N^B)$ the role is B and A is the opponent. Role A can have teams up to m^A players, and role B can have teams up to m^B agents. N^A is the set of the l^A different agents of role A, with this we end up having $n^A = l^A \cdot m^A$ agents, and similarly for B.

The possible teams for role A, namely $R^A = \{T_1^A, T_2^A, ...\}$, are the same as we did for R for cooperative games avoiding repetitions. Similarly, $R^B = \{T_1^B, T_2^B, ...\}$ for B. Then we now extend v for competition by defining $v^A(T^A, T^B)$, as the value of team $T^A \in R^A$ in role A against $T^B \in R^B$ as opponent in role B. Note that if we fix the opponent, e.g., T^B, from the point of the role A, we have its Shapley value from Eq. 3:

$$\varphi_i^A(v^A, T^B) = \frac{1}{m^A} \sum_{j=0}^{m^A-1} \left\{ \left(\!\!\binom{n^A/m^A}{j}\!\!\right)^{-1} \sum_{S \in R^A : |S| = j} dv^A(S, i, T^B) \right\} \tag{5}$$

where $dv^A(S, i, T^B) = \left[v^A(S \cup \{i\}, T^B) - v^A(S, T^B)\right]$ is the marginal contribution of agent i to coalition S when the opponent team is T^B. Then, if we have all possible teams for role B, then we can define $\varphi_i^A(v^A)$:

$$\begin{aligned} \varphi_i^A\left(v^A\right) = & \frac{1}{m^A\left|R^B\right|} \sum_{j=0}^{m^A-1} \left\{ \left(\!\!\binom{n^A/m^A}{j}\!\!\right)^{-1} \right. \\ & \left. \cdot \sum_{S \in R^A : |S| = j} \; \sum_{T^B \in R^B} dv^A\left(S, i, T^B\right) \right\} \end{aligned} \tag{6}$$

The amount that agent i gets given a team $game_T^B(v^B, N^B)$ when playing against T^A in role A is $\varphi_i^B(v^B, T^A)$. And the Shapley value $\varphi_i^B(v^B)$ is defined symmetrically to Eq. 6.

The value of agent i for all its possible participations in any team of any role is finally given by:

$$\varphi_i(v^A, v^B) = \frac{1}{2}\left[\varphi_i^A(v^A) + \varphi_i^B(v^B)\right] \tag{7}$$

The above equation makes sense when v^A and v^B have commensurate values (e.g., through normalisation), otherwise one role will dominate over the other. A particular case where this equation is especially meaningful is for symmetric team games, where both roles have the same scoring system. Finally, the total required values (experiments) for all φ_i^A is:

$$r^A = |R^A| \cdot |R^B| = \left[\sum_{j=0}^{m^A} \left(\!\!\binom{n^A/m^A}{j}\!\!\right)\right] \cdot \left[\sum_{j=0}^{m^B} \left(\!\!\binom{n^B/m^B}{j}\!\!\right)\right] \tag{8}$$

For instance, for $l^A = l^B = 3$ and $m^A = m^B = 4$, we have $r = 35^2 = 1225$ experiments (note that they are the same experiments for φ_i^B, so we do not have to double this). The huge numbers that derive from the above expression, also illustrated in Fig. 1 for a small example, means that calculating this extension of the Shapley value with repetitions and opposing teams exacerbates the combinatorial problem of computing the value of a huge number of coalitions. Consequently, we need to find ways of approximating the value, through sampling, as we see next.

3.3 Properties

In this work, we propose extending the Shapley value to calculate the benefits of each agent in the case of mixed settings (cooperation and competition games). The original Shapley value is characterised by the well-known properties of efficiency, symmetry, linearity, and null player. Let us analyse these properties for Eq. 6. We will see here that if $game^A(N^A, v^A)$ is defined from a set of $n^A = |N^A|$ players, we find that a special case of efficiency holds, the symmetry and the linearity property are met completely, while the null player property does not make sense in our case.

1. Efficiency. Efficiency in $game^A(v^A, N^A)$ requires that the sum of all the Shapley values of all agents is equal to the worth of grand coalition:

   $$\sum_{i \in N^A} \varphi_i^A(v^A) = v^A(N^A)$$

   For $m^A < n^A$, we cannot define the grand coalition if the maximum number of team members (in the lineup) is lower than the total number of agents. This is similar to many games such as football or basketball, where only a subset of players (11 and 5 respectively) can play at the same time. Accordingly, it is impossible to have a coalition with n^A agents. For the very special case where $m^A = n^A$, then we have the simple case of only one kind of agent $l_A = 1$ and the property is not insightful any more.

2. Symmetry. Now we see that the symmetry property holds in full:

   Proposition 1. *If for a pair $i, k \in N^A$, we have that $v^A(S \cup \{i\}, T^B) = v^A(S \cup \{k\}, T^B)$ for all the sets S that contain neither i nor k, then $\varphi_i^A(v^A) = \varphi_k^A(v^A)$.*

3. Linearity. The easiest one is linearlity as we only have composition of linear functions.

   Proposition 2. *If $v^A(S, T^B)$ and $w^A(S, T^B)$ are the value functions describing the worth of coalition S, then the Shapley value should be represented by the sum of Shapley values of the player derived from v^A and v^B: $\varphi_i^A\left(v^A + w^A\right) = \varphi_i\left(v^A\right) + \varphi_i\left(w^A\right)$. And for a, we have $\varphi_i^A\left(av^A\right) = a\varphi_i\left(v^A\right)$.*

4. Null player. A null player refers to a player who does not contribute to the coalition regardless of whether the player is in the coalition or not. For many team formations having both cooperation and competition, e.g., predator and prey game, even if a player is completely motionless, the other team members and the opponent team's members are affected by this agent, and it cannot have null effect. For instance, in the prey and predator game, if there is a collision, it will produce a reward to this player, which is not in line with the understanding of a null player. This property is not really important in our setting, as many factors affect the result to look for a normalised case where an absolute zero value is meaningful.

4 Approximating the Shapley Value

Applying the Shapley value requires the calculation of many v^A and v^B as per Eq. 6. In deep reinforcement learning, for instance, calculating v^A for a pair of teams in simple games such as predator-prey with a reasonable number of steps and episodes may require enormous resources. We explore what kind of sampling is most appropriate for the new extension of the Shapley value taking into account the trade-off between the number of experiments to be run (e.g., number of different v^A and v^B that are calculated) while keeping a good approximations to the actual Shapley values (note that we need a value for each i of the l different players).

4.1 Algorithms for Sampling

Monte-Carlo is the common and practical approach approximating the Shapley value [25,26]. Castro et al. [27] propose a sampling method to approximate the Shapley value by using a polynomial method. The stratified sampling method was first applied by Maleki [28]. These methods and many extensions have been successfully applied to approximate the Shapley value [29,30,31].

In what follows we present three methods for our setting. We have to sample from R^A and R^B, but we will only discuss sampling for one role to simply notation.

Simple random sampling This method simply chooses k elements $\mathcal{S} = \{S_1, S_2, ..., S_k\}$ from R with a uniform distribution and without replacement. Then, for each agent i, where $i = 1..l$, we compose all $S \cup \{i\}$ for each $S \in \mathcal{S}$, and check whether the new composed set is already in the sample. Then we calculate the approximation φ_i^*. Note that we sample on the population of experiments (the sets S in $\mathcal{S}$) and not on the population of marginal contribution pairs $\{v(S \cup \{i\}), v(S)\}$. If we fix s, the number of sampled experiments, and then try to find or generate the case when i is added, then the exact number of complete pairs will depend on the number of overlaps. If we want to get a particular value of pairs, we can sample elements from R incrementally until we reach the desired value.

Stratified random sampling The way the Shapley value is calculated by groups of coalitions of the same size (with j going from 0 to $m{-}1$) suggests a better way of sampling that ensures a minimum of coalitions to calculate at least some marginal contribution pairs for each value of j. Stratified sampling divides R into strata, which each stratum containing all the sets S such that $|S|{=}j$. If the size of a stratum $\mathit{\Gamma}_j$ is lower than or equal to a specific value $\mathit{\Gamma}_{min}$, we will sample all the elements from the stratum. For all the other strata, we will pick the same number for each. For instance, with $l{=}m{=}4$ and $\mathit{\Gamma}_{min}{=}5$, and $s{=}14$, we would do $s{=}1{+}4{+}3{+}3{+}3$ from the total of $r{=}1{+}4{+}10{+}20{+}35{=}70$. If $s{=}23$ we would do $s{=}1{+}4{+}6{+}6{+}6$. Once $\mathcal{S}$ is done, we proceed as in the simple random sampling when $\mathit{\Gamma}_j > 5$: for each $i = 1..l$, we compose all $S \cup \{i\}$ for each $S \in \mathcal{R}$, and check whether the new composed set is already in the sample. Then we calculate the approximation φ_i^*.

Information-driven sampling While stratified sampling tries to get information from all sizes, when samples are small, we may end having very similar coalitions, e.g., $\{a, a, b\}$ and $\{a, b, b\}$. Information-driven sampling usually aims for a more diversified sampling procedure. In our case, we use the Levenshtein distance as a metric of similarity between the different samples (assuming the multisets are ordered). Our version of information-driven sampling is actually based on the stratified random sampling presented above, where similarity is used intra-stratum and the coalitions are ordered with the largest average Levenshtein distance from the previous ones in the stratum.

4.2 Analysis of Sampling Methods

To evaluate which sampling method is best, we need to be able to calculate the actual Shapley values for several values of l and m. Doing this in a real scenario would be unfeasible, so we use synthetically generated data and explore different degrees of sampling for each method, to determine the method with best tradeoff between the approximation of the Shapley value and the number of experiments required.

The synthetic data is generated as follows. First, the worth v of each player (singleton sets) is generated from a uniform distribution $v \sim U(0, 1)$. Second, the contribution of a coalition is the sum of separate player contributions, i.e., $v(\{a, b\}) = v(\{a\})+v(\{b\})$. Third, we corrupt a number ν of these v for multisets, also using a uniform distribution $\sim U(0, 1)$. With this procedure, we have created six datasets. Synthetic data 1 is a game with m=l=4. There are three variants with ν= 1, 5 and 10 corrupted data, named 'test1', 'test2' and 'test3' respectively. Synthetic data 2 is a game with m=l=5. There are also three variants with ν = 1, 10 and 30 corrupted data. We used $\mathit{\Gamma}_{min} = 5$.

With these six synthetic datasets, we now evaluate the three methods and compare the approximate Shapley value with the true Shapley value using all coalitions. The total number of different coalitions (range of the x-axis) for synthetic data 1, with m=l=4 (n=16) is 70, and synthetic data 2, with m=l=5 (n=25) is exactly 252, coming from Eq. 4. To achieve a stable and robust evaluation, we repeat sampling 50 times before corruption and create 50 repetitions in each case for the corruptions. Then, we have 50×50 repetitions in total.

We computed the Spearman correlation and Mean Square Error (MSE) between the true Shapley and the approximation value. Fig. 2 shows the three sampling methods for increasing sampling size. The stratified and information-driven sampling methods only need a few coalitions (around 20 for m=l=4, and around 40 for m=l=5) to reach high Spearman correlation (0.98) and very low MSE. Since we do not see a clear difference between the stratified and information-driven methods, we will use the former in what follows.

5 Experimental Setting

Now we can explore how the new extensions are useful to determine the value of different algorithms in heterogeneous multi-agent systems with both competition

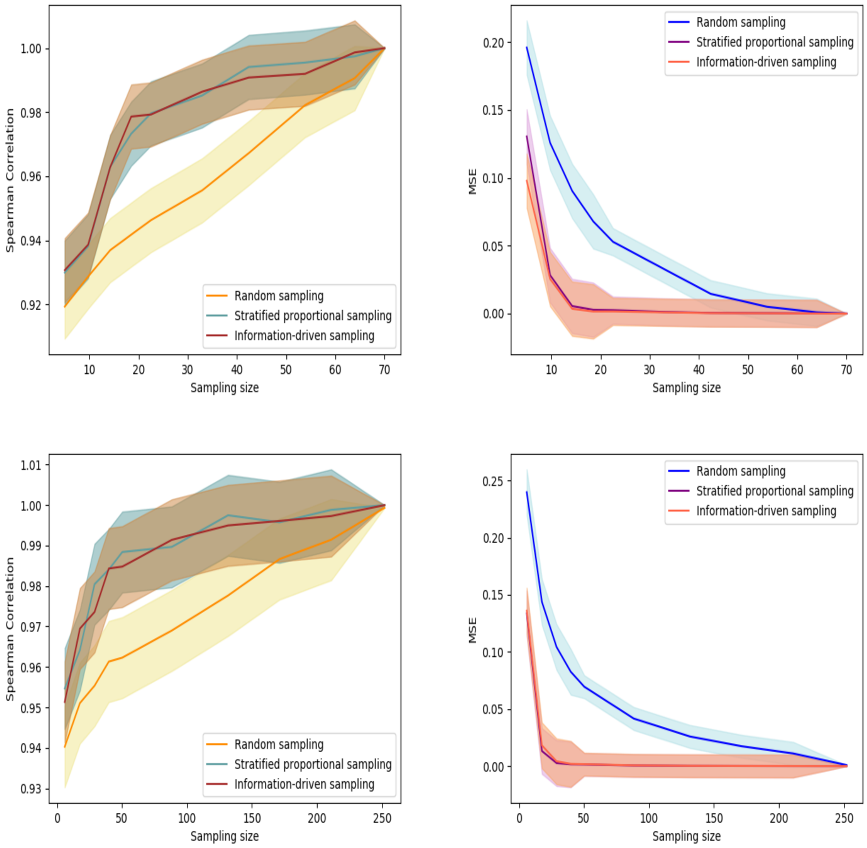

Fig. 2: Evolution of sampling methods (simple random, stratified proportional and information-driven) for $m = l = 4$ (top) and $m = l = 5$ (bottom). Left: Spearman correlation between the true Shapley values and the approximate values. Right: MSE.

and cooperation. In order to do this, we choose MPE (multi-agent particle environments), a simple multi-agent particle world [3,4] that integrates the flexibility of considering several game configurations with different kinds of learning agents. In particular, MPE comes with a single-agent actor-critic algorithm, Deep Deterministic Policy Gradient (DDPG), in which the agent will learn directly from the observation spaces through the policy gradient method, and a multi-agent variation, Multi-Agent DDPG (MADDPG), where decentralised agents learn a centralised critic based on the observations and actions of all agents.

We will explore their behaviour in the predator-prey game, a common cooperative and competitive game, where several predators (A) have to coordinate to capture one or more preys (B). Preys can also coordinate as well to avoid being caught. In the MPE standard implementation of this game, preys are faster than predators. The arena is a rectangular space with continuous coordinates. Apart from the agents themselves, there are also some static obstacles, which agents must learn to avoid or take advantage of. Agents and obstacles are circles of different size, as represented in Fig. 1. The observation information for each agent combines data from the physical velocity, physical position, positions of all landmarks in the agent's reference frame, all the other agents' position, and all the other agents' velocity. The prey will increase the reward for increased distance from the adversary. If collision, the reward will be –10. Contrarily, the adversary will decrease the reward for increased distance from the prey. If collision, the reward will be +10. In addition, prey agents will be penalised for exiting the screen.

Several questions arise when trying to understand how MADDPG and DDPG perform in heterogeneous situations. In particular, (1) Is MADDPG robust when it has to cooperate with different agents? (2) Is this the case when non-cooperative agents, such as a random agent is included? (3) Are the results similar for the predator role as for the prey role? To answer these questions, we will explore a diversity of situations (roles as prey or predator) and three types of agents (in both teams, so l^A=l^B=3). These are MADDPG, DDPG, and a random walk agent, represented by M, D and R respectively in the team. The total number of training episodes in the experiments is 60,000. We variate the number of agents in our experiments with a maximum of m^A=m^B=4. The number of combinations is 35×35=1225, according to Eq. 8. We do stratified sampling with sizes ranging from 37 to 199, using Γ_{min}=3. We use the same sampling for prey.

6 Results

We report here a summary of results. Further results with all the code and data readily available at a git repository[4].

One of the main motivations for MADDPG was showing that when several agents of this kind cooperate they can achieve better results than their single-agent version, DDPG. In this homogeneous setting, [3] show that "MADDPG

[4] https://github.com/EvaluationResearch/ShapleyCompetitive.

predators are far more successful at chasing DDPG prey (16.1 collisions/episode) than the converse (10.3 collisions/episode)”. We analysed the same situation with homogeneous teams of predators of 2, 3 and 4 MADDPG agents against 13 variations of prey teams of size 1, 2, 3 and 4. We do the same experiments with DDPG predators with exactly the same preys. In Table 1 (first two rows) we show the average rewards of the 39 games each. As expected, the predator M teams scored better than those with only D agents. The values are consistent with the apparent superiority of MADDPG over DDPG.

Table 1: Average reward for 39 homogeneous predator teams composed of two to four agents (first row with Ms only and second row with Ds only) against a diversity of 13 prey coalitions of size 1 to 4 (the same in both cases). Average rewards for 22 heterogeneous predator teams of sizes between 2 and 4 (all including at least one random agent R) against a diversity of 11 prey coalitions of size one to four (the teams in the third row contain an agent M that is systematically replaced by an agent D in the fourth row)

Teams	Predator	Prey
M (hom.: MM, MMM, MMMM)	3788K	–3324K
D (hom.: DD, DDD, DDDD)	3517K	–3375K
M (het.: MR, MRX, MR$XY$)	3121K	–3437K
D (het.: DR, DRX, DR$XY$)	3184K	–3380K

We can tentatively explore whether this advantage is preserved in heterogeneous teams. If we now build predator teams where apart from M or D agents we include other agents (and always a non-cooperative random agent R), we now get worse results (Table 1, last two rows) as expected, but interestingly we see that the average reward of the M teams is now worse than the D teams.

Because of the careful pairing of the experiments M vs D in Table 1, the average return are meaningful to illustrate the difference, but they do not really clarify whether the contributions of M and D are positive or negative (the average for predator will typically be positive as most results are positive, and the opposite for prey). This phenomenon is replicated when we calculate the average for all the experiments. In this predator-prey game, adding more preys (even if they are good) usually leads to more negative rewards, and hence the averages are negative. But could we still have a good agent, whose contribution is positive for the prey team? This is possible for the Shapley value, as the difference between two negative values can be positive, and does happen for some examples. Consequently, the Shapley values in Fig. 3 show a clearer picture of the actual contributions of each agent to the team (for either roles). While there are some fluctuations, the trends seem to stabilise around a sample size of 130, showing that the sampling method is effective beyond this level.

Looking at the sample size 199, as predators, the MADDPG agent has a value of 1387K while DDPG is at 1413K. The value of the random agent plummets to

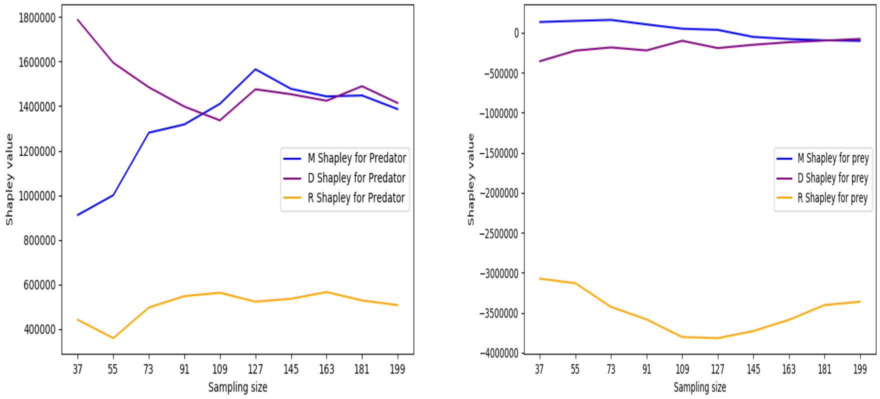

Fig. 3: Approximating Shapley for predator-prey environment with increasing sample size. Left: Predator. Right: Prey.

507K, which makes sense. As preys, the DDPG agent is also the most valuable, with a Shapley value of −79K while MADDPG goes down to −102K. The random agent is further down, at −3361K. Comparing with the results of Table 1, the approximation of the Shapley value integrates both homogeneous and heterogeneous cases, and shows that the gains of M in the homogeneous situations are counteracted by the poorer performance in the heterogeneous situations. Overall, for both predator and prey, the results for M and D are very close. The take-away message in this particular game is that D or M should be chosen depending on the proportion of heterogeneous coalitions that are expected or desirable.

7 Conclusions

The Shapley value provides a direct way of calculating the value of an agent in a coalition, originally introduced in cooperative scenarios with no repeated agents (completely heterogeneous). For the first time, we introduce an extension that covers both cooperative and competitive scenarios and a range of situations from complete heterogeneity (all agents being different) to complete homogeneity (all agents in a team equal). These multisets, and the existence of two or more teams competing, increase the combinatorial explosion. To address this, we have analysed several sampling methods, with stratified sampling finding good approximations with a relatively small number of experiments. We have applied these approximations to a prey-predator game, showing that the benefits of a centralised RL agent (MADDPG) in the homogeneous case are counteracted by the loss of value in the heterogeneous case, being comparable overall to DDPG.

There are a few limitations of this extension. First, as we have seen in the asymmetric game of predator-prey, the Shapley value as predator is not commensurate with the Shapley value as prey, and these values should be normalised before being integrated into a single value for all roles in the game. Second, the extension does not take into account the diversity of the team, something that might be positive in some games (or some roles of a game). Third, the Shapley value does not consider that some coalitions may be more likely than others, something that could be addressed by including weights or probabilities over agents or teams in the formulation, or in the sampling method. These are all directions for future work.

Acknowledgements

This work has been partially supported by the EU (FEDER) and Spanish MINECO grant RTI2018-094403-B-C32 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe", Generalitat Valenciana under grant PROMETEO/2019/098, EU's Horizon 2020 research and innovation programme under grant agreement No. 952215 (TAILOR), the EU (FEDER) and Spanish grant AEI/PID2021-122830OB-C42 (Sfera) and China Scholarship Council (CSC) scholarship (No. 202006290201). We thank the anonymous reviewers for their comments and interaction during the discussion process. All authors declare no competing interests.

References

1. Alvin E Roth. The Shapley value: essays in honor of Lloyd S. Shapley[M]// Cambridge Univ. Press, 1988.
2. Shihui Li, Yi Wu, Xinyue Cui, Honghua Dong, Fei Fang, and Stuart Russell. Robust multi-agent re- inforcement learning via minimax deep deterministic pol- icy gradient[C]//AAAI,2019,33:4213–4220
3. Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environ- ments. arXiv preprint arXiv:1706.02275, 2017
4. Mordatch I, Abbeel P. Emergence of grounded compositional language in multi-agent populations[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018, 32(1).
5. Pier Giuseppe Sessa, Ilija Bogunovic, Maryam Kamgarpour, and Andreas Krause. No-regret learning in unknown games with correlated payoffs[J]// NeurIPS, 2019.
6. Kjersti Aas, Martin Jullum, and Anders Løland. Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. arXiv preprint arXiv:1903.10464, 2019
7. Nash Jr J F. Equilibrium points in n-person games[J]// Proceedings of the national academy of sciences, 1950, 36(1): 48-49.
8. Aleksandrov M, Walsh T. Pure Nash Equilibria in Online Fair Division[C]//IJCAI. 2017: 42-48.
9. Balduzzi D, Tuyls K, Perolat J, et al. Re-evaluating evaluation[J]. Advances in Neural Information Processing Systems, 2018, 31.
10. Omidshafiei S, Papadimitriou C, Piliouras G, et al. α-rank: Multi-agent evaluation by evolution[J]. Scientific reports, 2019, 9(1): 1-29.

11. rpad E. Elo. The rating of chess players, past and present[M]. Acta Paediatrica, 32(3-4):201–217, 1978.
12. Glickman M E, Jones A C. Rating the chess rating system[J]. CHANCE-BERLIN THEN NEW YORK-, 1999, 12: 21-28.
13. Minka T, Cleven R, Zaykov Y. Trueskill 2: An improved bayesian skill rating system[J]. Technical Report, 2018.
14. Kenneth Harkness. Official chess hand- book[M]. D. McKay Company, 1967.
15. Kiourt C, Kalles D, Pavlidis G. Rating the skill of synthetic agents in competitive multi-agent environments[J]. Knowledge and Information Systems, 2019, 58(1): 35-58.
16. Kiourt C, Kalles D, Pavlidis G. Rating the skill of synthetic agents in competitive multi-agent environments[J]. Knowledge and Information Systems, 2019, 58(1): 35-58.
17. Fatima S S, Wooldridge M, Jennings N R. A linear approximation method for the Shapley value[J]. Artificial Intelligence, 2008, 172(14): 1673-1699.
18. Kotthoff L, Fréchette A, Michalak T P, et al. Quantifying Algorithmic Improvements over Time[C]//IJCAI. 2018: 5165-5171.
19. Li J, Kuang K, Wang B, et al. Shapley Counterfactual Credits for Multi-Agent Reinforcement Learning[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 934-942.
20. Yu C, Velu A, Vinitsky E, et al. The Surprising Effectiveness of PPO in Cooperative, Multi-Agent Games[J]. arXiv preprint arXiv:2103.01955, 2021.
21. Omidshafiei S, Pazis J, Amato C, et al. Deep decentralized multi-task multi-agent reinforcement learning under partial observability[C]//International Conference on Machine Learning. PMLR, 2017: 2681-2690.
22. Bowyer C, Greene D, Ward T, et al. Reinforcement learning for mixed cooperative/competitive dynamic spectrum access[C]//2019 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN). IEEE, 2019: 1-6.
23. Iqbal S, Sha F. Actor-attention-critic for multi-agent reinforcement learning[C]//International Conference on Machine Learning. PMLR, 2019: 2961-2970.
24. Ma J, Lu H, Xiao J, et al. Multi-robot target encirclement control with collision avoidance via deep reinforcement learning[J]. Journal of Intelligent & Robotic Systems, 2020, 99(2): 371-386.
25. Touati S, Radjef M S, Lakhdar S. A Bayesian Monte Carlo method for computing the Shapley value: Application to weighted voting and bin packing games[J]. Computers & Operations Research, 2021, 125: 105094.
26. Ando K, Takase K. Monte Carlo algorithm for calculating the Shapley values of minimum cost spanning tree games[J]. Journal of the Operations Research Society of Japan, 2020, 63(1): 31-40.
27. Castro J, Gómez D, Tejada J. Polynomial calculation of the Shapley value based on sampling[J]. Computers & Operations Research, 2009, 36(5): 1726-1730.
28. Maleki S. Addressing the computational issues of the Shapley value with applications in the smart grid[D]. University of Southampton, 2015.
29. Burgess M A, Chapman A C. Approximating the Shapley Value Using Stratified Empirical Bernstein Sampling[C]. International Joint Conferences on Artificial Intelligence Organization, 2021.
30. Gnecco G, Hadas Y, Sanguineti M. Public transport transfers assessment via transferable utility games and Shapley value approximation[J]. Transportmetrica A: Transport Science, 2021, 17(4): 540-565.
31. Illés F, Kerényi P. Estimation of the Shapley value by ergodic sampling[J]. arXiv preprint arXiv:1906.05224, 2019.