On the Role of Mobility and Interaction Topologies in Social Dilemmas

Joe Collenette[1], Katie Atkinson[1], Daan Bloembergen[2] and Karl Tuyls[1]

[1]Department of Computer Science, University of Liverpool, UK
[2]Intelligent and Autonomous Systems, Centrum Wiskunde & Informatica, Amsterdam, NL
j.m.collenette@liverpool.ac.uk

Abstract

Numerous studies have developed and analysed strategies for maximising utility in social dilemmas from both an individual agent's perspective and more generally from the viewpoint of a society. In this paper we bring this body of work together by investigating the success of a wide range of strategies in environments with varying characteristics, comparing their success. In particular we study within agent-based simulations, different interaction topologies, agents with and without mobility, and strategies with and without adaptation in the form of reinforcement learning, in both competitive and cooperative settings represented by the Prisoner's Dilemma and the Stag Hunt, respectively. The results of our experiments show that allowing agents mobility decreases the level of cooperation in the society of agents, due to singular interactions with individual opponents that limit the possibility for direct reciprocity. Unstructured environments similarly support a greater number of singular interactions and thus higher levels of defection in the Prisoner's Dilemma. In the Stag Hunt, strategies that prioritise risk taking show a greater level of success regardless of environment topology. Our range of experiments yield new insights into the role that mobility and interaction topologies play in the study of cooperation in agent societies.

Introduction

The extensive work on social dilemmas with self-interested agents has focused mainly on agents whose opponents are in an unchanging set, often represented as nodes in a network. When mobility is introduced to these agents, their opponents are no longer consistent since an interaction between any two agents is not guaranteed nor are any subsequent interactions. This paper explores what effects the move to agents who have differing sets of opponents has on a strategy. We are also interested to see if the outcomes of strategies change when mobility is present in a social dilemma setting. In addition we explore how the environment topology affects these strategies, and how these effects are interlinked. Our study is concerned with observing these effects in terms of both an individual agent's performance and the performance of the society as a whole. The performance of an agent or a society will be measured by the level of cooperation and the overall payoff achieved.

Throughout this paper we refer to static and mobile agents. A static agent refers to an agent which has a fixed number of opponents and does not move throughout the environment, which is modelled as a (static) network. A mobile agent refers to an agent that moves throughout the environment, modelled as an arena, and whose opponents will change over time. Previous work has documented what effects can arise when introducing mobility to agents (Ranjbar-Sahraei et al., 2014). This work has shown with mobile pure defectors and cooperators that when defectors were in the minority they were successful in a small world environment, but not in a regular environment. Collenette et al. (2016b, 2017a) expanded on this to include random and fully connected networks, along with random and empty arenas. That work was concerned with the mobility effects on the strategies implemented. Their results showed that the performance of the strategies depends on the environment topology in which they were used. Collenette et al. (2016b, 2017a) further showed that different densities of mobile agents in an arena can change the observed effects.

The work described above has focused on observing a single strategy in the Prisoner's Dilemma, limiting the ability to effectively analyse the effects mobile environment topologies have on social dilemmas and the society in general. We wish to investigate the different levels of cooperation that different self-interested strategies yield in a variety of environments. By expanding the range of agents and topologies, we will gain a greater insight into the external factors that will affect all strategies. Using these insights we gain understanding of an agent's behaviour when moving from a more theoretical exercise to the real world.

To isolate these external factors we will be running a number of different experiments where we simulate a society of self-interest agents. The agents will use a mix of different strategies, including fixed strategies that do not change over time, such as Tit-for-Tat, and adaptive strategies which use reinforcement learning, including SARSA and a Moody model. The arena shape for the mobile agents will be designed to be comparable to the network representation used for the static agents. We will then perform an analysis on the

	Coop	*Defect*
Coop	3, *3*	0, *5*
Defect	5, *0*	1, *1*

	Coop	*Defect*
Coop	3, *3*	0, *2*
Defect	2, *0*	1, *1*

Table 1: Payoff matrix of the Prisoner's Dilemma (Left) and the Stag Hunt (Right).

level of cooperation the society has achieved along with the average payoff of both the individual agent and for the whole society. This analysis allows us to determine the effects of mobility and the effects of different environments allowing us to further expand on what needs to be taken into account when designing agents that will be placed in an arena.

We find that mobile agents show a lower level of cooperation when compared to static agents. This is due to mobile agents facing a larger range of opponents and interacting with these opponents fewer times than static agents. Mobile agents in an open arena will have a higher level of defection than in arenas with more obstacles, however this only results in an increase in the society's payoff in Prisoner's Dilemma and not the Stag Hunt in our scenarios. This difference is due to temptation payoff being higher than mutual cooperation in the Prisoner's Dilemma and lower in the Stag Hunt.

Background and Related Work

Social Dilemmas

Our study uses both the Prisoner's Dilemma and the Stag Hunt games. In both these social dilemmas two players have the choice of cooperation or defection. The choice is made simultaneously and without prior communication. The payoffs for both dilemmas are shown in Table 1.

In both dilemmas, both players choosing to cooperate will give the highest payoff for the group as a whole. In the Prisoner's Dilemma there is a strong incentive for a player to defect, which leads to a Nash Equilibrium of mutual defection, giving the worst outcome for the society. The Stag Hunt (Skyrms, 2004) reduces the incentive to defect below the payoff that an agent would get for mutual cooperation, leading to two pure strategy Nash Equilibria. These are mutual cooperation and mutual defection. Each equilibrium has its own benefit and cost: if an agent chooses to cooperate there is the risk of losing all the payoff if the opponent chooses to defect. When the agent chooses to defect there is no chance of the agent losing its payoff, however it will have given up the chance for the highest payoff, making this the risk-dominant strategy.

Exploring how cooperation can evolve between groups of self-interested agents has been an active topic of research (Axelrod and Hamilton, 1981; Santos et al., 2008; Bloembergen et al., 2014; Skyrms, 2004; Bolton et al., 2016). We adopt this model of interaction so we can expand on the body of knowledge in regards to how the nature of this interaction changes when deployed in different environments.

Networks

Throughout this work we will be conducting experiments with different interaction topologies, modelled as networks. Each node in our networks will represent an agent with the edges of the network allowing interactions to take place between the two connected agent nodes. In order to make a meaningful analysis between mobile arenas and networked interactions, we will be using the effective degree of the mobile arena to construct our networks, where the degree is the number of unique agents they faced in an arena.

We use networks with different structural properties in our work, namely: **Small world** networks (Watts and Strogatz, 1998), which are networks with high clustering and small characteristic path length, intuitively this is a network where the nodes have very few neighbours but the distance between any two given nodes is also small; the **fully connected** network, also known as a complete network, in which every node is connected to every other node; **random** networks, where the edges of the network are randomly distributed; and **regular** networks, in which all nodes have the same degree. In this work we will focus on random regular networks, where the connections to specific nodes may differ, but the degree of each node will be the same, allowing us to make a meaningful comparison with the regular arena. We acknowledge that scale-free networks are an important part of exploring network interactions (Barabási, 2009), however we wish to expand upon the current knowledge, by allowing direct comparisons with previous work.

Collenette et al. (2016b); Ranjbar-Sahraei et al. (2014) focused on analysing a small subset of possible strategies, within regular and small world environments. Collenette et al. (2017a) provide an initial analysis on random and empty environments. The aim of this work is to generalise these effects for social dilemma scenarios. Ranjbar-Sahraei et al. (2014) showed that there is a difference between regular and small world arenas for the Prisoner's Dilemma; exploring this difference allows us to better consider the success of the strategies utilised. We can then use the found differences to better evaluate these strategies when moving from the more theoretical background of static networks to the "agents in the field" type of arenas. Previous work has shown that decreasing the available area to move around in arenas with the same topology reduces the variance in payoff between agents Collenette et al. (2017a, 2016b).

When considering the Prisoner's Dilemma in networks, previous work has shown that as the network connectivity increases in regular and small-world, the level of cooperation decreases, matching the Nash Equilibrium (Santos and Pacheco, 2005; Lieberman et al., 2005; Vukov et al., 2006; Barrat and Weigt, 2000). In random networks the level of cooperation in networks with low connectivity depends on the initial conditions of the agents. When the connectivity

is high, the level of cooperation is independent of the initial conditions (Durán and Mulet, 2005). For the Stag Hunt we see similar results where the level of cooperation is dependant on the network structure, with more connected networks showing lower levels of cooperation (Szolnoki and Perc, 2009; Starnini et al., 2011).

Strategies

In our experiments we will be including a number of different strategies, which we have split into two different types. The first type we define is the fixed strategies, which have a deterministic outcome. Secondly, we define the adaptive strategies which use reinforcement learning as part of their strategy. The fixed strategies include:

Tit-For-Tat (TFT) Initially cooperates, then copies the opponent's last action (Axelrod and Hamilton, 1981);

Win-Stay Lose-Shift (WSLS) Initially cooperates, repeating the current action as long as it receives the highest payoff possible (Nowak et al., 1993);

Random Cooperates or defects with equal chance;

Always Cooperate (ALL COOP) Always cooperates;

Always Defect (ALL DEFECT) Always defects;

Emotional - Active Shown to be the most effective emotional strategy in an arena (Collenette et al., 2016a,b). This strategy will switch to cooperation when the opponent has cooperated twice in a row, and switch to defection when the opponent defects, named E2 in previous works (Lloyd-Kelly et al., 2012b,a);

Emotional - Trustful Identified as the most effective emotional strategy in a network (Lloyd-Kelly et al., 2012a,b), named E7 in that work. This strategy will switch to cooperation when the opponent cooperates, and switch to defection when the opponent defects three times in a row.

The adaptive strategies use reinforcement learning as their strategy. Reinforcement learning uses trial and error to learn about its environment and how to optimise agents' actions based on the current state of the environment, in order to yield the greatest payoff for themselves. Actions that return the highest payoff are reinforced, whereas actions which return a low payoff or punishments are reduced. For our experiments the states will be the opponents. The two adaptive strategies are:

SARSA This is an on-policy reinforcement learning algorithm (Sutton and Barto, 1998). Definition 1 shows the Q-Value update algorithm.

Definition 1. Let S be the set of states with $s \in S$. Let A be the set of actions with $a \in A$. Let t be the time, r represent the reward, α the learning step size and γ the discount factor of future rewards. Then, SARSA updates $Q(s_t, a_t)$ using the following equation:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] \quad (1)$$

Moody This is also a reinforcement learning algorithm that uses a model of mood in both its action selection and estimation of future rewards (Collenette et al., 2017b). Definition 2 gives the Q-Value update algorithm

Definition 2. Let Mem_i^a be the set of rewards obtained by agent i when using action a where $|Mem_i^a|$ is at maximum 20, and $Mem_i^a(0)$ returns the most recent reward. Let m_i return the mood of agent i.

$$\begin{aligned} Q(s_t, a_t) &\leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma\Psi - Q(s_t, a_t)] \\ \Psi &= \left(n \sum_{0}^{\beta} Mem_i^a(n) \right) / \beta \\ \beta &= ceil(|Mem_i^a| / \alpha_i) \\ \alpha_i &= (100 - m_i)/100 \end{aligned} \quad (2)$$

Equation 2 shows how the moody agent evaluates its expected payoff by taking the average of the previous interactions for that agent and action, where Mem_i^a is an array of payoffs from each agent and action pair. How far back the agent looks is adjusted by the mood level, with higher moods looking at fewer previous interactions. For example if the mood (m_i) is 25 then β is 75% of the number times the agent has faced an opponent. Ψ is the mean payoff the agent received against that opponent and action for the last 75% interactions with that opponent.

For the emotional strategies the initial action will be split equally between cooperation and defection. For the adaptive strategies the initial action will have an equal chance of cooperation or defection. The adaptive strategies' initial Q value for an action will be set to the first payoff they receive for that state action pair. We do this as Shteingart et al. (2013) shows that this best reflects how people learn which action to take when given a choice between a risky or safe option. Collenette et al. (2017b) states that reflecting psychology is a requirement for the moody strategy and applying this to SARSA allows a meaningful comparison between them.

In our adaptive agent the action selection strategy will use the ϵ-greedy method, which selects a random action with probability ϵ, and the action with the highest Q-Value with probability $1 - \epsilon$. We set $\epsilon = 0.1$ for both adaptive agents. The moody strategy can adapt the value of ϵ depending on the value of the mood: when the agent is in a neutral mood, no change will be made. ϵ increases in line with how strongly the mood is felt, and gives a change of changing an action that is in conflict with the current mood

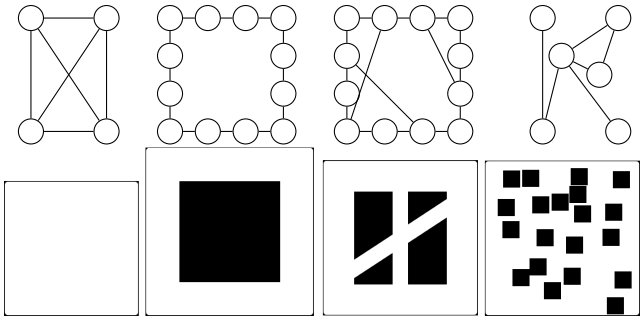

Figure 1: Fully connected, Regular, Small World, and Random environments (left to right). Example networks (top) and corresponding arenas (bottom)

of an agent. When the mood is high and the agent is defecting, ϵ increases the chance of cooperation. When the mood is low and the agent is cooperating, ϵ increases the change of defection. Collenette et al. (2017b) gives the formal definitions of the algorithm. We have altered one step of the mood update algorithm so that the mood reduces quicker when a poor outcome is obtained at high mood levels. This is achieved by altering the mood by the difference between the reward and the perceived average. Previously the reward was altered by the difference between the perceived average and the actual average. This is shown in Definition 3, which shows how the mood level goes up by the size of the payoff, including any adjustments that the Homo Egualis equation (Fehr and Schmidt, 1999) makes to the agent's perception of the reward. The Homo Egualis equation is an inequity aversion model, which models what the value received would be perceived as, when taking into context what other people have received.

Definition 3. Let I be the set of agents where $i, j \in I$ where j is the opponent, m_i be the mood of agent i, t denotes time. r_i^t denotes the payoff of agent i at time t. $\Omega_{i,j}^t$ denotes the Homo Egualis equation (Fehr and Schmidt, 1999) for agents' i and j at time t.

$$m_i^t = m_i^{t-1} + (r_i^{t-1} - \Omega_{i,j}^{t-1}) \tag{3}$$

The agents will be able to differentiate between opponents, each strategy will be applied to each agent individually and for the adaptive agents the opponent will represent the state. We set the learning rate as $\alpha = 0.1$ and the discount rate to $\gamma = 0.95$.

Experiment Setup

As previously noted, the environments we will be conducting our experiments in are using the fully connected, regular, small world, and random networks. Examples of the networks are shown in Figure 1. A fully connected network is equivalent to an empty environment (no blocks), as the agents have an equal chance of meeting any other agent.

In the random environment 20 blocks are placed randomly around the environment, leaving 36% of the environment available for movement. For the static equivalent we will use the Erdős-Rényi method (Erdős and Rényi, 1959) to generate a random network with a component of one, where each edge has a 36% chance of being generated.

To calculate the degree needed to generate the random regular network, we obtain the average number of unique opponents faced by an agent in the mobile arena. To ensure that this number is an accurate representation, we exclude any opponents who are faced once only. We use the algorithm described by Steger and Wormald (1999) which generates a random regular network in a relatively quick time (Kim and Vu, 2003), also we ensure the graph has a component of one. Using the method for obtaining the degree as before, we construct the random small world network using the Watts-Strogatz method (Watts and Strogatz, 1998). This is constructed as a ring network where the edges in the graph have a rewiring probability of 40%. That is, each connection has a 40% chance of changing to a different agent. Examples of each of the static networks are given in Figure 1. Where a 0% rewiring probability is a ring network and a 100% rewiring probability is a random network, with any number between producing a small world network.

Our experiments use 108 agents each using a single strategy. Each strategy will be represented equally. The agents' initial positions are randomised for each run, in both the network and the mobile arenas. In the arenas, shown in Figure 1, the agents move randomly around the environment. The agents have basic obstacle avoidance and generate a random heading between -45○ and 45○ each second to allow a random walk. An interaction occurs when two agents are facing each other and are closer than 20cm. Interactions may happen after each second, after which they are given two seconds where they may not interact. This prevents agents having more than one interaction while they are passing each other. The arenas will be simulated using e-pucks (Mondada et al., 2009) in Stage (Vaughan, 2008). The mobile agents will interact for 20 minutes, then the positions will be re-initialised while the agents will retain any knowledge they have accumulated, to allow agents sufficient chance to meet a range of opponents. We stop the simulation once the agents have converged. To calculate if convergence has occurred we take the proportions of mutual cooperation, mutual defection, and non-mutual outcomes of the 30 most recent 20-minute runs. We then compare this to the 10 most recent 20-minute runs and we say that convergence has occurred if the difference between the proportions calculated is within 0.005 of each other. We repeat the simulation 50 times in order to generate an accurate result on what proportions of actions the agents converge on.

For the networked experiments, the agents' positions are randomised in the network. As time has no true meaning in the network, an agent will interact with every neighbour

the average number of times the mobile equivalent interacts with a specific agent. For example, if in the arena an agent will average 3 interactions with a specific opponent, then in the networked environment the agent will interact with each neighbour 3 times. This counts as a run, we then use the same convergence properties as before.

Hypotheses

We predict that networks will be more successful in supporting cooperation when compared to the arenas; this is hypothesis 1 (H1). This prediction is due to agents being able to retaliate against exploitative agents in a reliable manner, ensuring that cooperative agents will both receive a high average payoff as a pair. In arenas there is no guarantee that agents will meet the same agent more than once, allowing exploitative agents to be successful, as the opponent is unable to retaliate. We predict that this will lead to successful arena strategies being those which take advantage of agents met rarely while cooperating with agents met frequently. The TFT and Trustful strategies are effective in maintaining cooperation over time, and depending on their initial action can be effective in taking advantage in a one-shot interaction.

We expect there to be differences between the environments in terms of which environments support the highest levels of payoff. Collenette et al. (2016a, 2017a) have shown high levels of mobility are a factor in supporting higher levels of average payoff in arenas with mobile agents. That work leads us to expect that the empty environment will support the highest levels of average payoff; this is hypothesis 2 (H2). While the empty arena is represented by the fully connected network, we expect this network to achieve the lowest amount of cooperation and therefore payoff, as Lieberman et al. (2005) show on networked interactions; this is hypothesis 3 (H3).

Results and Analysis

We examine the level of cooperation achieved by the society in each environment. Figure 2 show the results for the Prisoner's Dilemma (Top) and the Stag Hunt (Bottom). The level of cooperation is given as a percentage where 100% represents that every outcome was mutual cooperation for that particular run. These figures support our hypothesis H3, as in both the Prisoner's Dilemma and Stag Hunt the network with the least amount of cooperation was the Fully Connected network. The level of cooperation increases with the level of randomness in the network construction, further supporting our hypothesis H3, and Lieberman et al. (2005).

In addition we note that arenas have a consistently lower level of cooperation and take longer to converge when compared to networked interactions (H1). To explore why this is the case, we look into the main difference between the two environments, namely the number of unique opponents the agents will face. Figure 3 shows a histogram of how many unique opponents an agent faced, for each arena. Overlaying

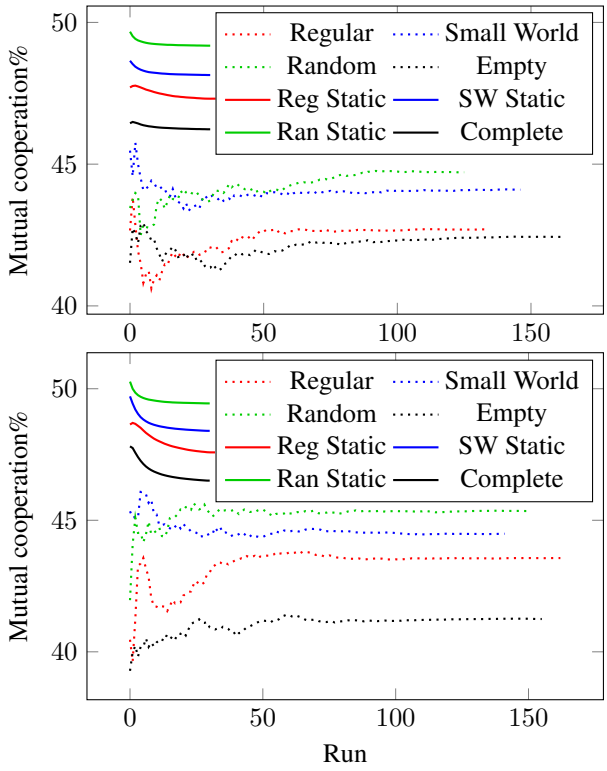

Figure 2: Proportion of mutual cooperation for each run, in each of the environments, for the agents playing the Prisoner's Dilemma (Top) and the Stag Hunt (Bottom).

is a second histogram which excludes any opponent they interacted with exactly once. When an agent interacts with an opponent exactly once we term this a *singular interaction*.

When we include the singular interactions, the range of agents faced does not line up with the expected distributions for the different arena structures. Across the networks we would expect a singular peak in the histogram which would lean further right as the arena becomes more open. Excluding the singular interactions these figures show that the shape of the environment affects the number of opponents as would be expected, with an agent in the empty environment interacting with the largest range of agents and the random environment facing the smallest range, in-line with expectations as this mimics the static networks. A significant factor can be seen in how the number of opponents decreases greatly when we exclude singular interactions.

What the above means for the agents is that when mobility is introduced there are effectively both the iterative social dilemmas and one-shot social dilemmas being played at same time, with no knowledge on which is being played. Agents are able to take advantage of some opponents without retaliation, allowing defectors to go unpunished. Mutual cooperation is harder to sustain as cooperating agents can effectively be split up, as they may not interact with each

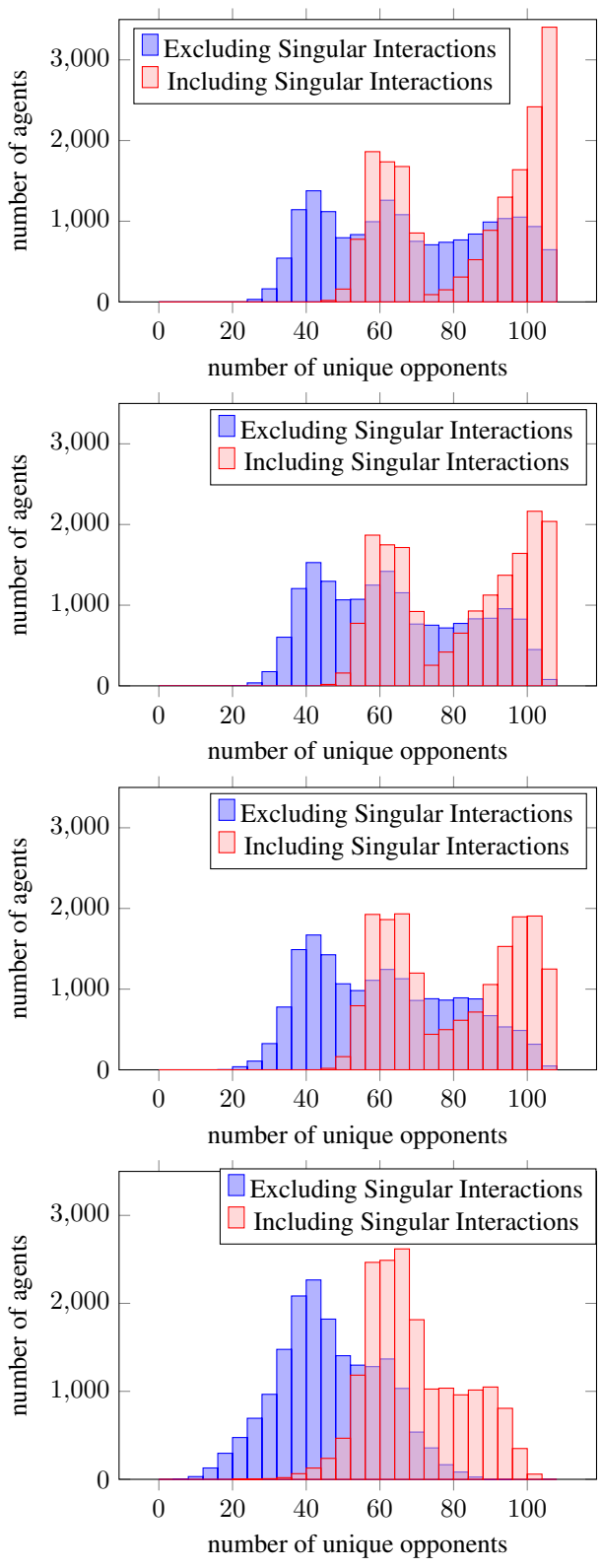

Figure 3: Histogram of unique opponents faced in the empty, regular, small world, and random environments (top to bottom), including and excluding singular interactions

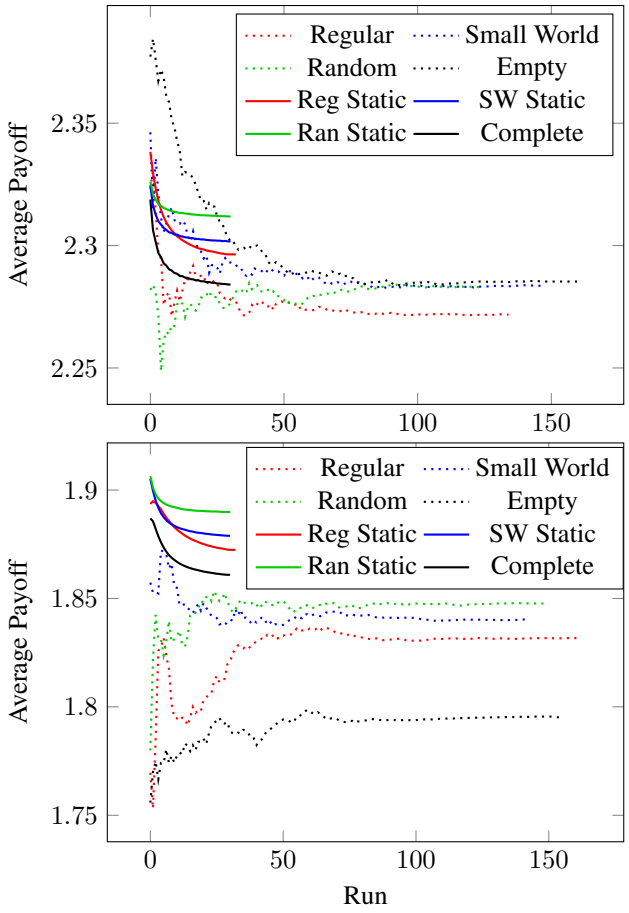

Figure 4: Average Scores for each environment in the Prisoner's Dilemma (Top) and the Stag Hunt (Bottom)

other again. The average range of opponents is 89 (14.65 standard difference) in the mobile arenas, with an average of 76.5 (32 standard difference) in the static environments. The larger range of opponents in the mobile environment extends the convergence time, along with the irregular number of interactions with a specific opponent.

Payoffs

To analyse hypothesis H2 and H3 we look at Figure 4 which shows the average score of the society of agents in the Prisoner's Dilemma (left) and the Stag Hunt (right), for each environment. The results in these figures allow us to reach an initial conclusion on hypothesises H2 and H3.

For hypothesis H3 we see that the fully connected network performed the worst with more randomised networks achieving higher scores in both the Prisoner's Dilemma and the Stag Hunt, validating this hypothesis. Hypothesis H2, where we expected the empty environment to perform best, is supported for the Prisoner's Dilemma but not for the Stag Hunt. This leads us to the conclusion that there does not seem to be a rule to directly relate the environment structure

and the average payoff. Rather than focusing on whether there exists an environment which can affect the payoff directly, we shall analyse why two different social dilemmas yield different orderings (in terms of society payoff) for the different arenas, by looking deeper into the form of the games, and the effects on individual strategies.

We show the average scores of each strategy in each environment along with what percentage of their chosen actions was to cooperate, for the Prisoner's Dilemma and the Stag Hunt, in Table 2. This allows us to see whether a majority of cooperation or a majority of defection is the most successful in our environments.

We see in the Prisoner's Dilemma that fully cooperative agents perform the worst. The empty arena supports high levels of average payoff for strategies that choose defection more often. This defection is not supported in the Stag Hunt, which supports cooperative strategies. This is due to the temptation payoff being higher than the mutual cooperation payoff for an individual agent in the Prisoner's Dilemma, since the empty arena supports a large range of opponents, with a significant number of them being one-shot interactions. The agent that defects does not receive any retaliation allowing it to achieve high levels of payoff. The more cooperative agents do not receive a reduction in payoff as large as the increase in payoff that less cooperative agents receive in the other arenas, allowing the empty arena to yield the highest average payoff.

The above effect is not seen in the Stag Hunt as the temptation payoff is less than the mutual cooperation payoff. Since defecting still receives less retaliation in the empty arena this does not result in the increased payoff that would be expected in the Prisoner's Dilemma. In general we see that the environment has less of an effect on the strategies in the Stag Hunt. This highlights how the payoff matrix of a social dilemma and the environment are interlinked. We note that TFT was the most successful as per Axelrod and Hamilton (1981), and that Trustful was also a successful strategy across both social dilemmas, supporting H1 further.

Conclusion

We have conducted experiments with a number of different strategies, evaluating them in the Prisoner's Dilemma and the Stag Hunt games. The agents have played the games in a number of different network topologies and their equivalent arenas with mobile agents. Through these experiments we have shown how arenas lower the average payoff, the level of cooperation, and increase convergence times when compared to the equivalent static network. We can attribute this to an inherent property of mobility in an arena, namely the range of opponents that will be faced and how a number of them will only be faced once, effectively making them one-shot interactions and thus limiting an agent's ability to retaliate against defection.

We have shown that the more open an arena, the larger the range of opponents and number of singular interactions. For an agent to take advantage of the lack of retaliation, the payoff matrix needs to also support defection over cooperation for an individual agent. The Prisoner's Dilemma has this support so defecting agents achieve high levels of payoff in the empty arena. Conversely the Stag Hunt does not have this support as the temptation payoff is lower than the mutual outcome payoff for an individual agent.

We believe that our results are relevant for agent designers transferring theoretical network interactions into real world practical environments. Future work will consider a wider range of environment topologies to support this aim by further generalising the work.

References

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

Barabási, A.-L. (2009). Scale-free networks: A decade and beyond. *Science*, 325(5939):412–413.

Barrat, A. and Weigt, M. (2000). On the properties of small-world network models. *The European Physical Journal B-Condensed Matter and Complex Systems*, 13(3):547–560.

Bloembergen, D., Ranjbar-Sahraei, B., Bou Ammar, H., Tuyls, K., and Weiss, G. (2014). Influencing social networks: An optimal control study. In *Proc of ECAI'14*, pages 105–110.

Bolton, G. E., Feldhaus, C., and Ockenfels, A. (2016). Social interaction promotes risk taking in a stag hunt game. *German Economic Review*, 17(3):409–423.

Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. (2016a). The effect of mobility and emotion on interactions in multi-agent systems. In Pearce, D. and Pinto, H. S., editors, *Proc of STAIRS'16*, pages 39–50. IOS Press.

Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. (2016b). Mobility effects on the evolution of co-operation in emotional robotic agents. In *Proc of ALA'16*.

Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. (2017a). Environmental effects on simulated emotional and moody agents. *The Knowledge Engineering Review*, 32.

Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. (2017b). Mood modelling within reinforcement learning. In *Proc of ECAL'17*, pages 106–113. MIT Press.

Durán, O. and Mulet, R. (2005). Evolutionary prisoner's dilemma in random graphs. *Physica D: Nonlinear Phenomena*, 208(3):257–265.

Erdős, P. and Rényi, A. (1959). On random graphs i. *Publ. Math. Debrecen*, 6:290–297.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly journal of Economics*, 114:817–868.

Kim, J. H. and Vu, V. H. (2003). Generating random regular graphs. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 213–222. ACM.

	Environment							
Strategy	Empty	Small World	Regular	Random	Complete	SW Static	Reg Static	Ran Static
WSLS	2.28 (80.4%)	2.33 (77.7%)	2.29 (79%)	2.37 (76.2%)	2.37 (74.8%)	2.37 (74.5%)	2.36 (75.3%)	2.38 (74.5%)
Random	2.32 (50%)	2.29 (50%)	2.30 (50%)	2.22 (50%)	2.23 (50%)	2.23 (50%)	2.24 (50%)	2.23 (50%)
All Coop	2.10 (100%)	2.14 (100%)	2.14 (100%)	2.12 (100%)	2.16 (100%)	2.16 (100%)	2.15 (100%)	2.16 (100%)
All Defect	2.34 (0%)	2.27 (0%)	2.27 (0%)	2.25 (0%)	2.18 (0%)	2.17 (0%)	2.19 (0%)	2.17 (0%)
TFT	2.35 (70.1%)	**2.35 (68.4%)**	**2.31 (67.5%)**	**2.36 (67.2%)**	**2.39 (66.5%)**	**2.39 (66.2%)**	**2.39 (66.6%)**	**2.39 (66.3%)**
SARSA	**2.35 (39.4%)**	2.26 (39.1%)	2.27 (39.3%)	2.27 (37.7%)	2.22 (37.3%)	2.21 (37.4%)	2.23 (37.5%)	2.20 (37%)
Active	2.26 (44.1%)	2.23 (44.8%)	2.23 (43.5%)	2.20 (43.9%)	2.18 (45%)	2.18 (44.8%)	2.19 (45%)	2.18 (45.3%)
Trustful	2.27 (71.6%)	2.26 (71.2%)	2.24 (70.5%)	2.28 (71.7%)	2.28 (72.6%)	2.27 (72.5%)	2.27 (72.6%)	2.27 (72.3%)
Moody	2.29 (50.5%)	2.26 (50.4%)	2.30 (50.5%)	2.26 (50.2%)	2.23 (50%)	2.22 (50.1%)	2.23 (50.3%)	2.22 (50.1%)
WSLS	2.05 (80.7%)	2.11 (79%)	2.09 (79.3%)	2.06 (75.9%)	2.10 (74.9%)	2.10 (74.8%)	2.09 (75.3%)	2.10 (74.5%)
Random	1.54 (50%)	1.53 (50.1%)	1.55 (50%)	1.50 (50%)	1.50 (50%)	1.49 (50%)	1.50 (50%)	1.49 (50%)
All Coop	2.06 (100%)	2.14 (100%)	2.12 (100%)	2.13 (100%)	2.16 (100%)	2.15 (100%)	2.16 (100%)	2.16 (100%)
All Defect	1.34 (0%)	1.33 (0%)	1.33 (0%)	1.32 (0%)	1.29 (0%)	1.29 (0%)	1.30 (0%)	1.29 (0%)
TFT	**2.13 (68.6%)**	**2.15 (68.3%)**	**2.17 (69.6%)**	**2.17 (67.4%)**	**2.20 (66.6%)**	**2.20 (66.5%)**	**2.20 (66.8%)**	**2.20 (66.7%)**
SARSA	1.61 (37.7%)	1.63 (39.8%)	1.62 (38.6%)	1.64 (39.3%)	1.61 (38.2%)	1.61 (38.1%)	1.62 (38.1%)	1.60 (38%)
Active	1.81 (43.5%)	1.84 (45%)	1.83 (44.3%)	1.85 (45.5%)	1.85 (45.1%)	1.85 (45.1%)	1.85 (45.1%)	1.84 (45%)
Trustful	2.05 (70.4%)	2.07 (70.6%)	2.10 (72%)	2.15 (72.4%)	2.17 (72.7%)	2.18 (72.8%)	2.17 (72.8%)	2.17 (72.6%)
Moody	1.59 (51.3%)	1.55 (50.6%)	1.56 (51.1%)	1.50 (50.4%)	1.50 (50.6%)	1.50 (50.4%)	1.51 (50.9%)	1.49 (50.2%)

Table 2: Average score (Percentage cooperation chosen) of each strategy in each of the environments in the Prisoner's Dilemma (Top) and the Stag Hunt (Bottom). The most successful strategy is shown in bold.

Lieberman, E., Hauert, C., and Nowak, M. A. (2005). Evolutionary dynamics on graphs. *Nature*, 433(7023):312.

Lloyd-Kelly, M., Atkinson, K., and Bench-Capon, T. (2012a). Developing co-operation through simulated emotional behaviour. In *13th International Workshop on Multi-Agent Based Simulation*.

Lloyd-Kelly, M., Atkinson, K., and Bench-Capon, T. (2012b). Emotion as an enabler of co-operation. In *ICAART (2)*, pages 164–169.

Mondada, F., Bonani, M., Raemy, X., et al. (2009). The e-puck, a robot designed for education in engineering. In *Proc of ICARSC'09*, pages 59–65.

Nowak, M., Sigmund, K., et al. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, 364(6432):56–58.

Ranjbar-Sahraei, B., Groothuis, I. M., Tuyls, K., and Weiss, G. (2014). Valuation of cooperation and defection in small-world networks: A behavioral robotic approach. In *Proc of BNAIC 2014*, pages 103–110.

Santos, F. and Pacheco, J. (2005). Scale-Free Networks Provide a Unifying Framework for the Emergence of Cooperation. *Physical Review Letters*, 95(9):1–4.

Santos, F. C., Santos, M. D., and Pacheco, J. M. (2008). Social diversity promotes the emergence of cooperation in public goods games. *Nature*, 454(7201):213–216.

Shteingart, H., Neiman, T., and Loewenstein, Y. (2013). The role of first impression in operant learning. *Journal of Experimental Psychology: General*, 142(2):476.

Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge University Press.

Starnini, M., Sánchez, A., Poncela, J., and Moreno, Y. (2011). Coordination and growth: the stag hunt game on evolutionary networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(05):P05008.

Steger, A. and Wormald, N. C. (1999). Generating random regular graphs quickly. *Comb. Probab. Comput.*, 8(4):377–396.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT press Cambridge.

Szolnoki, A. and Perc, M. (2009). Resolving social dilemmas on evolving random networks. *EPL (Europhysics Letters)*, 86(3):30007.

Vaughan, R. (2008). Massively multiple robot simulations in stage. *Swarm Intelligence*, 2(2-4):189–208.

Vukov, J., Szabó, G., and Szolnoki, A. (2006). Cooperation in the noisy case: Prisoner's dilemma game on two types of regular random graphs. *Physical Review E*, 73(6):067103.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442.