Spatial configurations and learning rules impact on the evolution of cooperative behaviour in the *n*-person iterative prisoner's dilemma

Department of Electrical Engineering, Mathematics and Computer Science, Technische Universiteit
Delft, Netherlands
Roberta Gismondi

Abstract

Decision-making dynamics and their impact of human behaviour have raised a large number of questions throughout the years. Traits like competition and collaboration amongst agents are often studied, in the context of Game Theory, by the medium of games such as the Iterative Prisoners' Dilemma. Furthermore, many realistic scenarios and possible real world applications of the Iterative Prisoners' Dilemma (e.g. socio-geographic and economic ones) can only be modeled by a more genral instance of the game that allows for multiple numbers of players such as the ***n-person*** IPD.
Work has been done to analyse the effect of spatial configuration on the outcome of the game[4]. The goal of this research is to conduct further analysis on the possible confounding factors of these experimental settings. In particular, we investigate on the effect than different machine learning approaches for learning-rule inference (such as genetic algorithms and particle swarm optimisation) have on the correlation between the dependent variables of previous controlled experiments (the number of players, the scale of interaction and the initial percentage of cooperators and defectors) and the evolution of cooperative behaviour. The data that are relevant to answer the research question are gathered by means of repeated controlled experiments that aim to give insight on the effect that the factors under analysis have on the convergence of the environment to a state of almost full cooperation.

1 Introduction

The need for a tool to study the evolution of cooperative and competitive behaviour in many real world scenarios (such as socio-geographic and economical environments or biological systems) has lead to the formulation, in the context of Game Theory, of new extensions to the renowned Prisoner's Dilemma. The ***n-person iterative prisoner's dilemma*** (IPD) consists of multiple, consecutive rounds of the classic Prisoner's Dilemma, each played by $n > 2$ players. Similarly to the classic game, each player can either choose to *defect* or *cooperate* at each round. The defect option is the dominant strategy (*Nash equilibrium*), although it intersect in a subpar outcome. In fact, if all agents choose to cooperate the resulting outcome is preferable. In the classical formulation of the game, since no agent is motivated to deviate unilaterally from defecting, cooperative behaviour is highly unlikely to emerge[10]. Nonetheless, the introduction of multiple consecutive rounds in the iterative extension of the game is proven to lead to a more significant evolution of cooperative behaviour among the agents[5]. This new variant of the game is indeed deemed to have grater generality and more applicability in real-life scenarios[9].

Due to its wide suitability to model relevant decision-making processes in realistic environments, the n-person iterative prisoner's dilemma has been thoroughly studied during the past two decades. In particular, the effect of numerous factors on the evolution of cooperative behaviour was the matter of extensive analysis. The past twenty years of research have indeed shown that parameters such as the *number of players*, the *mobility* of the agents and the *initial percentage of cooperators* and defectors are all variables that can determine the predominance of cooperative or competitive behaviour at the end of game[4][5][2]. The aforementioned conclusions, although thoroughly proven under the assumption of an evolutionary approach to strategy inference, were drawn without considering the impact that the method employed to infer the strategy might have to the dynamics of the game.

The research on the development of new learning-rules for NIPD agents has indeed led to the proposal of multiple ml-based approaches to the decision making processes of players. Among many, solutions employing Genetic Algorithms, Particle Swarm Optimisation and Q-Learning have proven themselves to be valid alternatives that favour the development of effective NIPD strategies[7][6][8].

The development of these new solutions raises new questions on the dynamics behind the evolution of cooperation in the NIPD. The purpose of this work is indeed to investigate on whether the approach to strategy inference can be considered a confounding factor of previous analysis on the development of cooperative behaviour. The hypothesis is that, although generic patterns might be confirmed, the ML algorithm chosen to implement the agent's computational intelligence has bearing on the speed to which the population converges to a predominance of cooperators.

To prove so, a set of controlled experiments is designed in 3.4, each will be conducted under the assumption of a different ML-based approach to strategy inference. The algorithms to be employed are explained in 3. The results of the experiments are presented in 4, and conclusions are drawn in 6 along with suggestions for future work. A meditation on possible ethical implications and on the reproducibility of the work is included in 5.

2 Related Work

A spatial version of the n-person iterative prisoner's dilemma has proven itself to be a powerful tool to analyse and reason upon a variety of realistic socio-economic scenarios [5] [2]. Nonetheless, while the 2-players version of the game has been extensively studied for decades, a more general formulation of the game, allowing $n > 2$ players to play rounds of the IPD, is deemed to have greater generality and applicability to to real life situations[9].

In his work on Game Theory, Colman[10] defines the n-player Prisoner's dilemma by three main properties:

- each player can either defect or cooperate;
- the defect option is dominant for each player, meaning each player is better off defecting than cooperating, no matter the percentage of cooperators in the n-players cluster;
- the defect strategy intersect in a non-optimal equilibrium; if all players choose their non-dominant strategy of cooperation, the resulting outcome is preferable, from each player's perspective, to the one resulting from every player choosing to defect. Nonetheless, no one is motivated to deviate unilaterally from defecting.

Extensive work has been done to investigate and interpret the effect of different factors on the evolution of cooperation across multiple rounds of a spatial prisoner's dilemma[1]. Previous studies have indeed analysed the impact of the number of players in the Prisoner's Dilemma Game on the evolution of cooperation, along with how the latter is influenced by the initial percentage of cooperators and the magnitude of the interaction of agents[5][2].

As far as the first one of the aforementioned factors (the number of players) is concerned, [5] argues that, although cooperation can still be present in a version of the n-IPD with $n > 2$, it is more difficult to evolve cooperation as the group size increases, while [2] extensively reports on how the initial percentage of cooperators has little bearing on the emergence of cooperation but the mobility of the automata, thus the scale of their interaction, was a central factor that favoured cooperation.

Both the experimental settings, as many others in the context of analogue analysis for 2IPD games, were based on the assumption of evolutionary learning rules, i.e. strategies developed throughout the game in accordance to a mimic behaviour of the most successful player. Nonetheless, multiple alternative machine learning approaches were proposed to replace the go-to evolutionary algorithm in strategy inference. In particular, an evolutionary model based on Genetic Algorithms is explained in [7] while a model based on Particle Swarm Optimisation is proposed in [6] and it is stated to enhance the cooperation rate. On the other end, in an empirical study of reinforcement learning in the iterated prisoner's dilemma, [11] investigates the ability of Q-Learning agents to play against other unknown opponents. It was indeed concluded that although learners faced difficulties when playing against other learners, agents with longer history windows, lookup table memories and longer exploration schedules fared best in the IPD games.

As the research on new machine learning based strategies advances, the question of whether those approaches affect the conclusions drawn in previous studies on the effect of spatial configuration on the evolution of cooperation affirms itself as a legitimate research question.

This work intends to elaborate on whether the assumption of different ML techniques for learning rules inference is to be considered a confounding factor in the analysis that was conducted on the effect of spatial configuration on the evolution of cooperation in the NIPD.

3 Methodology

The research data in this thesis is drawn from three main sources; a spatial iterative prisoner's dilemma game is simulated by means of agent-based modelling involving agents situated on a square grid. The players are therefore allowed to learn new strategies to play the game using three different frameworks for computational intelligence: a vanilla evolutionary algorithm to mimic the efforts of the most successful agent in the neighbourhood, a genetic algorithm to produce an evolutionary fortunate strategy and a particle swarm optimisation approach to encourage Pareto optimal behaviour.

This section describes the procedures and algorithms used in this investigation, including the design of the controlled experiments that were performed to answer the research question.

3.1 Baseline Algorithm

The algorithm at the base of much of the related work on the evolution of cooperation in spatial NIPD is a simple ***mimic-last-move evolutionary algorithm***.

In fixed size grid, each agent is placed to occupy a single cell. The n-person iterative prisoner's dilemma is played for $n > 2$ rounds and at each round the agent can either cooperate or defect. Furthermore, the agent keeps track of its own currently gained payoff; information about an agent's *payoff* can be shared at the moment of an interaction. Interactions happen systematically at every round with a fixed number of nearby agents.

At the beginning of the game, each agent is instructed to randomly pick a move to play during the first round (either cooperate or defect). The first round is therefore arbitrarily played, each player can either be a cooperator with probability p (initial percentage of cooperators) or a defector with probability $1 - p$.

From the second round on, the agent picks its next move according to the following decision-making process: it retrieves information about the nearby agents' current payoff during its interactions with them. It identifies the 'best fit' member of the neighbourhood as the player with the highest payoff and decides upon its next move mimicking the player's last move.

Although it is a very simple strategy, the mimic-last-move evolutionary algorithm has proven itself to guarantee a fast convergence to a state of full cooperation amongst the players in the grid.

3.2 Genetic Algorithm

Genetic algorithms are *search algorithms* that exploit the mechanics behind natural selection and genetics. The initial state of the algorithm includes a sample of the search space constituted by *random* solutions. The fitness of these solutions is then evaluated according to a designed *fitness function* and a form of 'natural selection' is performed based on the fitness score achieved by each of the solutions. The selected solutions are therefore combined (through *cross-over* and *mutation*) to produce a new generation of solutions [7].

The first step to develop a Genetic Algorithm for strategy inference in the NIPD game is to figure out how to encode a strategy as a string. In [7], Haider A. suggests to reason upon the possible combinations of previous games, assuming that each player can remember up until one previous game. A strategy is then a rule that specifies an action in the case of each one of these possibilities.

The instrument that maps each possibility to a consequent decision upon the next game is a *lookup-table*. To use a string as a strategy the player records the moves made in the previous game and retrieves from the lookup-table the value that corresponds to the case. The retrieved integer value i (from 0 to the number of possible combinations of previous game moves) is then used to select the *ith* letter in the strategy string. The letter (C for cooperate and D for defect) will therefore be the agent's next move of choice.

A random strategy is therefore a random sequence of Cs and Ds, with length equals to the number of possible previous games scenarios. At each round of the game, the agents play the NPD and record their current payoff. The payoff is the estimate of each agent's strategy's fitness. The 10% of the population with the highest payoff is therefore selected to generate the next generation of strategies. The new generation is produced by randomly picking two parents from the 10% elite. The two parent-strategies are then crossed-over with probability 0.95. The cross-over happens at a random point of the strategy string. With probability 0.02 the resulting children-strategies are mutated in two different points. The mutation point is picked at random and the letter corresponding to the selected point is switched (Cs are replaced with Ds and Ds are replaced with Cs). The old generation is finally replaced by the new generation of strategies; the new strategies are randomly placed in the grid.

The algorithm performs a thorough scan of the solution space and eventually converges to a state of full cooperation between the agents in the grid.

3.3 Particle Swarm Optimisation

Swarm Intelligence is a computational technique inspired by the behaviour of animal *swarming*, where a coordinated behaviour can be reached by means of a small set of simple local interactions between members of the flock or between the individuals and the environment.

Particle Swarm Optimisation is the most popular approach in SI. Searching the best solution in PSO is carried out by endowing each member of the swarm to wonder the search space and to adjust its velocity at each step based on the best local solution and the swarm (*global best*) solution.

In the adjusted implementation of the PSO employed in this work, each particle's position (*strategy*) is represented by means of a bi-dimensional normalised vector representing the probability of choosing to cooperate or defect. At each round, while searching the strategy solution space, the position of each particle is updated according to the following equation:

$$X_i(t+1) = X_i(t) + V_i(t+1) \tag{1}$$

where:

$$V_i(t+1) = wV_i(t) + c_1 r_1(t)(y_i(t) - x_i(t)) + c_2 r_2(t)(\hat{y}_i(t) - x_i(t)) \tag{2}$$

as presented in [6] and normalised afterwards; the coefficients c_1 and c_2 have an initial value of 2.0 and 2.5 and are respectively decreased and increased by 0.1 at each iteration to allow the global solution contribution to weight more as the game proceed. The parameter w, the *inertia weight*, is meant to balance the global exploration and the local exploration of PSO; it is set to be within the range [0, 1] and it is decreased to 0.1 during the game.

Along with the specifications of the PSO implementation, a set of communication ***topologies*** are described in [6] as a mean to enhance the communication between players during the rounds of the game.

To the purpose of this paper, the PSO strategy is implemented assuming a *random topology* that, given n members of a neighbourhood, establishes n random symmetrical connections between pairs of individuals.

As we have reasons to think that this component of the algorithm has a strong correlation with the effect that the *scale of interaction* has on the performance of the algorithm, results in subsection 4.2 must be interpreted with caution. It is important to bear in mind the possible alterations in the outcome of such experiments under the assumptions of topologies that dictate sparser connections between neighbours (such as the *ring topology*) or an higher number of pair communication between members of the neighbourhood (such as the *star topology*).

Nonetheless, the algorithm is able to converge quickly and to reach a state of almost full cooperation in the grid at the end of the game.

3.4 The controlled experiment

The previous sections have attempted to provide a brief summary of the possible alternative algorithms that can be exploited to infer strategies to play the n-person iterative prisoner's dilemma.

As was pointed out in the introduction to this paper, the purpose of this work is to elaborate on whether the choice of a certain approach can represent a confounding factor in the analysis conducted on the effect of spatial configuration on the development of cooperative behaviour. This is a preliminary study that aims to justify further, meticulous and statistically rigorous research on the possible spurious association caused by this confounder.

In order to accomplish the goal, a series of *controlled experiments*, involving each of the aforementioned algorithms, were carried out to test the effect of the *number of players*, the *scale of interaction* of the agents and the *initial percentage of cooperators* and defectors under the assumption of different learning rules.

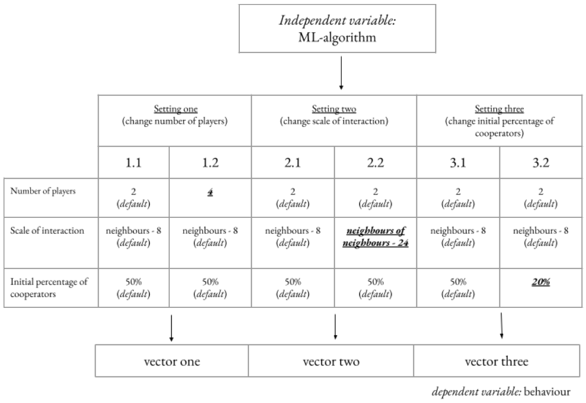
The *independent variable* in this scenario is indeed the learning rule of choice. The variable is assigned three possible different values: the agents' computational intelligence can be based on either a simple *mimic-last-move evolutionary algorithm* (the baseline), a *genetic algorithm* or a *particle swarm optimisation algorithm*, according to their implementations in 3.1, 3.2, 3.3.

The *dependent variable* to be observed is the behaviour of the agents, intended as their tendency to cooperate rather than defect while playing the game. In particular, the change in the final portion of cooperators as the number of players increases, as their mobility is increased and when we change the spatial configuration to modify the initial collaborators/defectors ratio.

The dependent variable can therefore be intended as a *3-tuple* of *vectors*, each vector representing the evolution of cooperative behaviour when one of the aforementioned spatial factors is tweaked. The vectors' direction is regarded as the general trend (cooperation increases or decreases), while the magnitude amounts to the intensity of this effect.

Other components of the game, such as the size of the grid (and thus the total number of the agents) and the number of rounds to play, are kept constant across all instances of the experimental setting.

The size of the grid is a 50X50 matrix, each cell containing only one agent. The agents play a total amount of fifty rounds each time. A more detailed explanation of how values are assigned to the variables of the experiment during the simulations can be found in Figure 1.



Independent variable: ML-algorithm

	Setting one (change number of players)		Setting two (change scale of interaction)		Setting three (change initial percentage of cooperators)	
	1.1	1.2	2.1	2.2	3.1	3.2
Number of players	2 (*default*)	***4***	2 (*default*)	2 (*default*)	2 (*default*)	2 (*default*)
Scale of interaction	neighbours - 8 (*default*)	neighbours - 8 (*default*)	neighbours - 8 (*default*)	***neighbours of neighbours - 24***	neighbours - 8 (*default*)	neighbours - 8 (*default*)
Initial percentage of cooperators	50% (*default*)	50% (*default*)	50% (*default*)	50% (*default*)	50% (*default*)	***20%***

vector one	vector two	vector three

dependent variable: behaviour

Figure 1: Specifications of the experimental setting

The code for the simulations is written in *python3* and the experimental setting is implemented exploiting the *mesa* library for agent-based modelling (ABM).

4 Results

As mentioned in Section 3.4, in order to provide insight on the evolution of cooperative behaviour in the n-person IPD, an estimate of the change in the number of cooperators in the grid was produced for each experimental setting illustrated in Figure 1.

Simple statistical analysis was used to approximate the numbers produced by the experiments in order to mitigate the role that randomness plays in the machine learning algorithms under analysis.

With the aim of generating a reliable value for such quantities in a feasible amount of time, the algorithms implemented a simple form of *antithetic variable variance reduction technique* whenever a uniform distribution was used initialise a random variable in the process.

The experiments were therefore run three hundreds times each, and the *average* of the results is used as an estimate of the value to be computed.

The results are presented in Tables 1, 2 in the form of *estimate* ± *error*, where the error is the *standard deviation* of the gathered data.

4.1 The number of players

The first set of analyses examined the changes in the magnitude of the cooperative behaviour as the number of player is increased. Cooperation is measured as the portion of agents in the grid that opted for a collaborative deportment in the last round of the game. These agents are represented in the visuals of the grid with the colour blue.

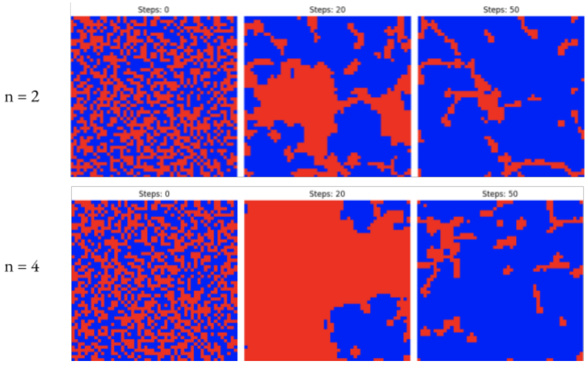


Figure 2: Comparison of the evolution of cooperative behaviour in 2IPD and 4IPD conditional to MLM-EA

Figures 2, 3 and Figure 6 compare the evolution of cooperative behaviour and its correlation with the number of contends playing the n-person iterative prisoner's dilemma under the assumption of the three algorithms presented in Section 3. The top half of visuals 2 and 3 represent the state of the grid across 50 rounds of 2IPD when the agent's decision process is based respectively on the mimic-last-move evolutionary algorithm (MLM-EA) and the genetic algorithm (GA). The second row of the figures aims to report about an analogue series

Table 1: Development of cooperation in the controlled experiments

	MLM-EA	GA	PSO
Experimental Setting 1.1	1119.316 ± 47.15	395.67 ± 98.46	1242.89 ± 24.92
Experimental Setting 1.2	895.09 ± 49.55	394.11 ± 91.63	1242.34 ± 25.97
Experimental Setting 2.2	1409.69 ± 29.95	447.54 ± 92.89	1246.50 ± 12.25

of simulations with the number of players increased from two to four (thus the game consists of 50 rounds of 4IPD). As the convergence to a state of almost full collaboration is much quicker under the PSO assumption, Figure 6 shows the evolution of cooperative behaviour (in terms of number of cooperators) at each round of the NIPD. Experimental setting 1.1 and 1.2 represent respectively the case of a 50-rounds 2IPD and 50-rounds 4IPD under the assumption of PSO (full details about the experimental settings can be found in Figure 1). The algorithms are implemented as per their description in Section 3 .

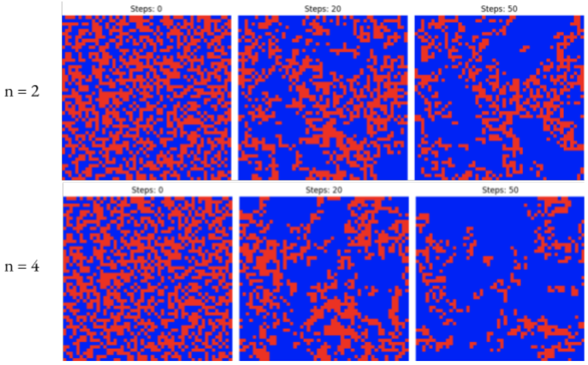


Figure 3: Comparison of the evolution of cooperative behaviour in 2IPD and 4IPD conditional to GA

Experimental setting 1.1 and Experiemental setting 1.2 in Table 1 compare the delta of collaborators when the game is played respectively by two and four players, employing the three algorithms analysed.

What stands out in the illustrations is how increasing the number of players affects and discourages cooperation when the MLM-EA is involved in the players' decision-making process, while it has less impact on the development of collaboration when the latter is replaced by the GA. Ultimately, it seems to have little to no effect when the approach that is used is the PSO.

Together these results provide important insights about the correlation between the number of players and their collaborative attitude and how this can be widely dependent on the nature of the computational intelligence framework to be employed. The magnitude of the effect established to be caused by the variation of the number of players by [4] has proven to vary a lot across the three scenarios presented, this might suggest a possible total cancellation (or even a shift in the trend) of such effect under the assumption of a different reinforcement learning approach.

4.2 The scale of interaction

The second set of analyses examined the effect of increasing the size of the neighbourhood on the evolution of cooperation between the agents. As mentioned in the previous subsection, the cooperation is measured as the cooperators/defectors ratio after 50 rounds of n-person prisoners' dilemma.

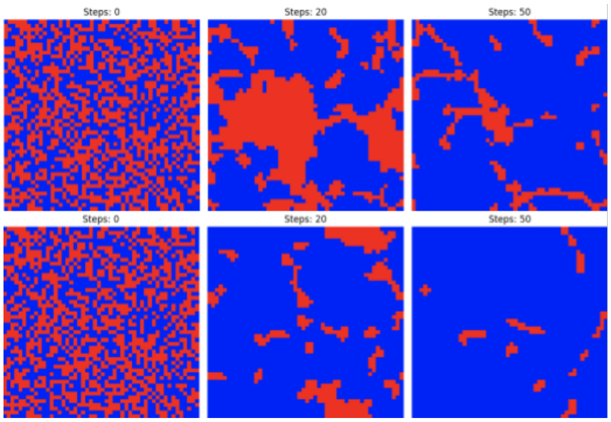


Figure 4: Comparison of the evolution of cooperative behaviour changing the scale of interaction of the agents, conditional to MLM-EA

To that end, Figures 4, 5 show the development of cooperation across 50 rounds of NIPD; as it was the case for the previous set of results, the cooperators are depicted in blue and the defectors are in red. The top-halves of visuals 4 and 5 report the state of the grid after 50 rounds of 2IPD when each agent's neighbourhood consists of its immediate contiguous peers (once again, the results are produced varying the approach to strategy inference and employing respectively MLM-EA, GA). The bottom-halves of the visuals assume one agent's neighbourhood to be defined as the set of its immediate neighbours and their nearest neighbours.

Results under the PSO assumption are depicted in Figure 6 for reasons analogue to the ones elaborated in the previous subsection.

Figure 4, reproducing the results of previous relevant studies, shows how an increase in the neighbourhood size leads to a quicker convergence and an higher final percentage of cooperators. The delta of the number of cooperators, as reported in Table 1, is much smaller in Experimental setting 1.1, where the neighbourhood consists of near at hand agents, than the equivalent measurement in Experimental setting 2.2, when the neighbourhood consists of the immediate neighbours and their neighbours.

Table 2: Final state of the grid in Setting 1.1 (base case) vs Setting 3.2

	MLM-EA	GA	PSO
Experimental Setting 1.1	2368.31 ± 37.9	1644.17 ± 94.21	2492.23 ± 9.50
Experimental Setting 3.2	2407.70 ± 35.01	1662.10 ± 89.33	2486.91 ± 12.60

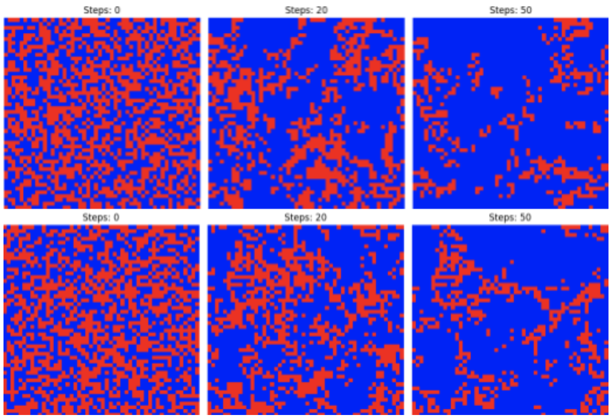


Figure 5: Comparison of the evolution of cooperative behaviour changing the scale of interaction of the agents, conditional to GA

Although an analogue conclusion can be drawn while observing the behaviour of the agents that implemented the GA (Figure 5), the magnitude of this effect is hardly the same in the latter case. In fact, although values in Table 1 relative to Experimental setting 1.1 and 2.2 are indeed comparable in their rapport to the ones reported under the MLM-EA assumption, the error coupled with such estimates is significantly higher. The scale of interaction seem to have less consistent influence on the evolution of cooperation. The fact that the two parameters are loosely coupled can be caused by the *elitist* nature of the genetic algorithm. Independently by the interactions amongst members of the neighbourhood, the algorithm contemplates a complete replacement of the old generation of strategies with a new generation that is the result of a cross-over and mutation process between members of a *global elite*. This justifies the lower correlation between the two variables in this experiment.

Figure 6 reports the results of increasing the scale of interaction of the agents under the assumption of the PSO, comparing the number of cooperators across the fifty rounds in Experimental setting 1.1 and Experimental setting 2.2.

The single most striking observation to emerge from the data comparison was that the algorithm seems to converge similarly in both cases, to reach a state of almost full collaboration in the grid at the end of the game. Further inquiry on the effect that the topology of choice in the PSO algorithm can bear on such phenomenon is in need to establish with certainty the consistency of this result.

In summary, these results show that, analogously to what was established in the previous subsection regarding the number of players in the game, the effect of the scale of interaction is influenced by the reinforcement learning technique of choice and the details of its implementation.

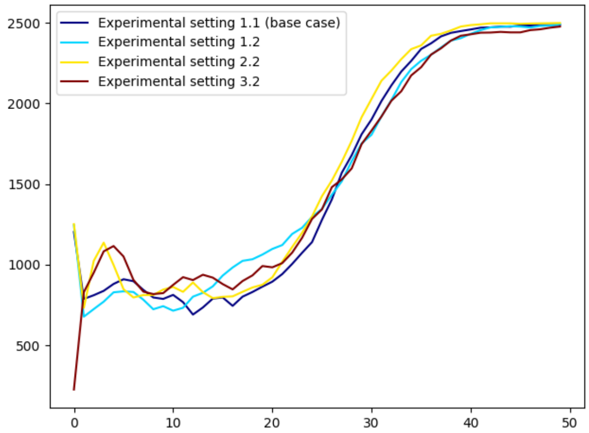


Figure 6: Comparison of the convergence to cooperation in different experimental settings, conditional to PSO

4.3 The initial ratio

The third set of analyses examined the effect of biasing the initialisation of the grid to have majority of defectors playing the first round of the game.

The initial percentage of cooperators/defectors is deemed by Power [2] to have little bearing on the development of cooperative behaviour in the n-person iterative prisoners' dilemma.

The top-half of Fig 7 shows the initial and final state of the grid (after 50 rounds of 2IPD) assuming a uniform distribution of cooperators and defectors in the initialisation of the grid. In other words, the state of the grid at step zero consists of 50% cooperators and 50% defectors.

The bottom-half of Figure 7 reports the results of an analogue setting assuming an initial percentage of cooperators equals to 20%.

Furthermore, estimates of the number of cooperators at the end of the NIPD employing respectively MLM-EA, GA and PSO are reported in Table 2. The data reported in the first row of the table are relative to Experimental setting 1.1, when the grid is initialised to contain an equal percentage of cooperators and defectors. This data is compared to Experimental setting 3.2, when the initialisation of the grid is biased to include only 20% of cooperators.

The results in presented in the visual and the table partially confirm the conclusions drawn by previous studies, and show how it is indeed of minor importance how the early state of the grid is biased towards a general attitude to cooperate or defect.

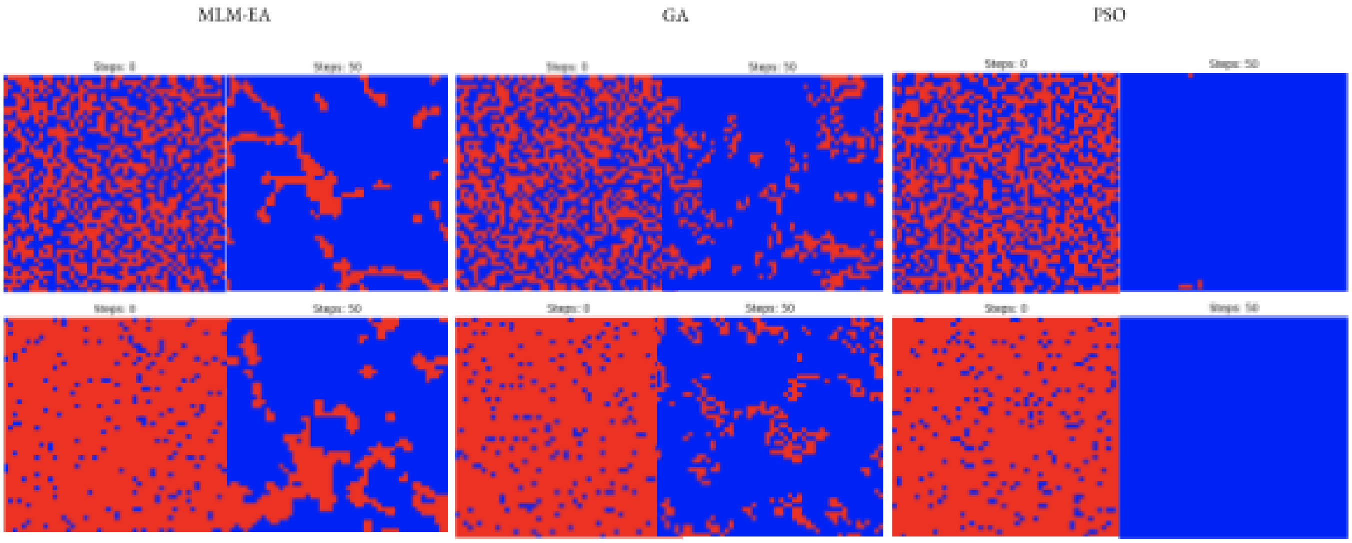


Figure 7: Impact of grid initialisation on the evolution of cooperation in NIPD, conditional to different ML approaches to strategy inference

5 Responsible Research

The strong reliance that the scientific community, and arguably society itself, place on experimental results rises a number of questions about the role of reproducibility as an essential practice in every scientific discipline.

In the past years the difficulty to replicate machine learning research was a topic for discussion in the context of many computer science and ML conferences.
This shows a need to be explicit about exactly what is meant by the word ***replicability***.

The term *replicability* has come to be used to refer to the characteristic (of an experiment or scientific setting) of being able to be copied exactly; in the context of machine learning it is often associated to the practice of providing the code to give third parties the possibility to replicate the experiment as it was performed during the process of answering the research question(s). Whereas replicability refers to the operations of performing the exact steps (in the exact order) performed by the research team, *reproducibility* refers to the extent to which consistent results are obtained when multiple experiments are conducted to answer the same research question.

In his attempt to distinguish the term '*reproducibility*' and the term '*replicability*' Drummond [12] claims there are important differences between the two: reproducibility implies changes while replicability discourages them. For this reason, while reproducibility is generally desirable, replicability is not worth having.

In the task of defining reproducibility for machine learning based research is indeed in order a look-back to more traditional sciences. As stated by Sonnenburg et al. in [14], "in many areas of science it is only when an experiment has been corroborated independently by another group of researchers that it is generally accepted by the scientific community." To this end, a mere replication of the experiment by means of the replication package might not provide the necessary endorsement to the drawn conclusions.

As noted by Drummond [12], an attempt to a discussion about scientific results is far more cost effective than precise records of the original experiment, especially in light of the great trust placed in these results. Collecting scripts and software as the basis of experimental results is a good practice to enhance the reviewing process, but reproducibility should not be the reason why we record experiments.

Also, as it is the case for the implementation that is the basis of this work, the performance of computational intelligence and systems based on machine learning algorithms designed to learn by trial and error is widely influenced not only by the exact code used, but to the random numbers that are used to initialise the learning environment [13].

To mitigate the effects of the latter, the conclusions drawn in this paper are presented to the reader as the results of a statistical estimate of the quantities under analysis; variance reduction techniques (such as antithetic variables) were used to estimate the values as precisely as possible in feasible time.

Although a replication package including the code used to run the experiments is provided (upon request), the implementation is thoroughly described in Section 3 to enhance reproducibility as previously defined in this passage.

6 Conclusions and future work

The aim of the present research was to examine the extent to which different approaches for strategy inference can affect the behaviour of the agents playing the n-person iterative prisoner's dilemma as the spatial configuration of the game changes. In particular, the effect of factors such as the number of players, their scale of interaction and the initial predominance of cooperators or defectors is investigated under the assumption of different machine learning algorithms shaping the agent's learning rules. The present study was designed to determine if the choice of such different ML approaches can be considered a confounding factor in previous analysis on the correlation between the aforementioned spatial features and the development of cooperative behaviour.

Previous studies have proven that, while the initial cooperators/defectors ratio has little bearing on the final predominance of cooperators in the context of the NIPD, the number of players to participate in the game is of key importance, and the greater it is, the less prone the contenders are to collaborate. Furthermore, the number of interactions between agents and thus the amount of information an agent is able to gather from the grid is a relevant factor that favours joint effort and cooperation. The most obvious finding to emerge from this study is that the general pattern identified by previous related work is indeed confirmed, although the end result of adapting the parameters under analysis is contingent on the algorithm of choice.

This finding provides some support for the conceptual premise that learning approaches are relevant constituents of NIPD simulations that should be regarded as much as possible while drawing conclusions on experimental results of such nature. Being limited to an analysis of three main ML approaches, this study lacks generality. Notwithstanding these limitations, the study suggests that some algorithms might bear a more consistent attitude to collaboration than others and that the effect of some of the spatial constituents that were accounted for in this study might be cancelled of shifted under the assumption of a different model for the agent's computational intelligence.

More information on the effectiveness of alternative algorithms would help us to establish a greater degree of accuracy on this matter. Scientific research lead by means of NIPD simulations is as the basis of many sociological, sociogeographic and socio-economic notions; the insight provided by this work can help in the task of confidently defining circumstances under which human collaboration is enhanced.

A natural progression of this work is to analyse, in an analogue manner, other types of machine learning approaches for strategy inference.

The findings reported in this work provide insights for future research: although the generic pattern of the effect of spatial configuration on the evolution of cooperation is confirmed under the hypothesis of the algorithms under analysis in this paper, the effect is very much dependent on the approach to strategy inference. This could suggest a potential cancellation (or even shift) of such effect(s) when a different ML-based approach is employed.

Furthermore, large randomised controlled trials could provide more definitive evidence to endorse the conclusions of this work. To this end, implementing multiple ML alternatives to drive the learning process of the agents is of paramount importance.

References

[1] Schweitzer, F., Behera, L., & Mühlenbein, H. (2002). Evolution of cooperation in a spatial prisoner's dilemma. Advances in Complex systems, 5(02n03), 269-299.

[2] Power, C. (2009). A spatial agent-based model of N-person prisoner's dilemma cooperation in a socio-geographic community. Journal of Artificial Societies and Social Simulation, 12(1), 8.

[3] Oliphant, M. (1998). Evolving cooperation in the non-iterated prisoner's dilemma: The importance of spatial organization.

[4] Suzuki, R., Arita, T. (2003). Evolutionary analysis on spatial locality in n-person iterated prisoner's dilemma. International Journal of Computational Intelligence and Applications, 3(02), 177-188.

[5] Yao, X., Darwen, P. J. (1994). An experimental study of N-person iterated prisoner's dilemma games. Informatica, 18(4), 435-450.

[6] Almanasra, S. (2019). Evolutionary model for the iterated n-players prisoners' dilemma based on Particle Swarm Optimisation. Journal of Theoretical and Applied Information Technology, 97(5).

[7] Haider, A. (2005). Using genetic algorithms to develop strategies for the prisoners dilemma.

[8] Kies, M. (2020). Finding Best Answers for the Iterated Prisoner's Dilemma Using Improved Q-Learning. Available at SSRN 3556714.

[9] Davis, J. H., Laughlin, P. R., Komorita, S. S. (1976). The social psychology of small groups: Cooperative and mixed-motive interaction. Annual review of Psychology, 27(1), 501-541.

[10] Colman, A. M. (2016). Game theory and experimental games: The study of strategic interaction. Elsevier.

[11] Sandholm, T. W., Crites, R. H. (1996). Multiagent reinforcement learning in the iterated prisoner's dilemma. Biosystems, 37(1-2), 147-166.

[12] Drummond, C. (2009). Replicability is not reproducibility: nor is it good science.

[13] Hutson, M. (2018). Artificial intelligence faces reproducibility crisis.

[14] Sonnenburg, S., Braun, M. L., Ong, C. S., Bengio, S., Bottou, L., Holmes, G., ... Williamson, R. C. (2007). The need for open source software in machine learning.