Towards Understanding the Role of Learning Models in the Dynamics of the Minority Game

Ricardo M. Araújo and Luís C. Lamb
Institute of Informatics, UFRGS
Porto Alegre, 91501-970, Brazil
rmaraujo@inf.ufrgs.br, LuisLamb@acm.org

Abstract

This paper reports experiments in a boundedly rational evolutionary game, namely the Minority Game, where agents apply a very simple learning algorithm to discard bad strategies and create new ones. The results show that even such simplified learning model presents qualitatively differences from the behaviour of the traditional game, where strategies are fixed and cannot be modified or discarded. We show that this results are qualitatively similar to other, more complex, learning approaches. Also, we study how the learning parameters of our model affects the dynamics of the game, evidencing a high dependence between the behaviour of the system and the way fitness is attributed to new strategies entering the game.

1. Introduction

Classical economics makes the assumption that economic agents involved in any system are perfectly rational. This usually means that each agent has knowledge of all relevant aspects of the environment, a logical and coherent preference scale, and enough computational power to process all this information in order to choose the best course of action to attain the highest, optimal point in his or her preference scale [12]. This assumption facilitates the use of analysis tools within game theory to predict the outcome of interactions among multiple agents, but is nevertherless a strong assumption over human agents. Many studies suggest that we, humans, often do not behave in a perfect rational way [14], limiting the predictive power of traditional tools used for analysis of social interactions and economics. In [12, 11] it is argued that a more useful anaysis tool would include in the model the procedures used by the agents in reasoning, including known limitations. That is, agents would be modeled with *bounded rationality*. Bounded rational agents have limits in their abilities to make choices, often making suboptimal ones and resorting to "shortcuts" in reasoning, i.e. *rules of thumb*.

In order to study the effects of bounded rationality in multi-agent decision making, Arthur [2] proposed a scenario known as the *El Farol Problem*. In this scenario, multiple, boundedly-rational agents choose whether or not to go to the El Farol bar in a given night. It turns out that an agent only enjoys the night out if the bar is not overcrowded - given an overcrowd threshold common for all agents. An agent thus goes to the bar if he believes that it will not be overcrowded. No communication is allowed among the agents. Whenever all agents commit their choice simultaneously, an interesting paradox arises: if many believe that few will go, then many will show up; if many believe many will go, then few will show up. This belief might be constructed by reasoning over what other agents intend to do, but one must consider that each agent would be reasoning in the same way about every other agent, creating lines of thought such as "if agent 2 believe that I believe that agent 2 believes..." and so on. Human subjects have difficulties following this line of reasoning, specially where several agents are involved. This problem is dealt in the El Farol Problem by considering that agents are not able to reason explicitly over what other agents are reasoning and choices are individually made based solely on a history of past attendances to the bar, which is common knowledge to all agents. In Arthur's model, agents are given a simple inductive learning algorithm to reason over this information.

This setup aimed at studying the effects of inductive learning as a procedure for boundedly rational agents reasoning in such competitive systems; such setups are useful since they can be seen as simplified models of a market economy. Results in [2] have shown that the system converges close to a game-theoretic equilibrium, fluctuating around the overcrowd threshold. This equilibrium is where the sum of gains is maximum, i.e. the system as a whole is using the given resources in an efficient way, by maximizing the number of agents satisfied.

In [15] a better formalization and generalization to

Arthur's proposal is developed and named *the Minority Game*. The Minority Game scenario closely follows the general setup of the El Farol problem, but with some simplified assumptions in order to make it more general and mathematically tractable. In the Minority Game (MG) an odd number (N) of agents iteratively choose whether or not to be in one of two possible groups (a more general setup with an arbitrary number of groups may be found in [6]) and those belonging to the minority group (the group containing the least number of agents) are positively rewarded. Thus, it is of interest of every agent to be in the minority group, creating the same kind of paradox as in Arthur's model.

In contrast to the El Farol scenario, in the MG agents do not receive a history of past attendances; they receive only the information of which group had the minority in previous runs (thus, the history is a binary string and not a history of integers as in the El Farol problem). This represents much less information available to the agents. Each agent is given a fixed set of strategies and chooses his or her best strategy at each turn to commit a choice, evaluating each by their performance in the past. An agent cannot modify his or her strategies, only choose his or her preference over the ones given at the start of the simulation. It was verified that this setup results in convergence towards a mean attendance of $N/2$ to the groups, which maximizes the sum of gains of the agents as a group [15]. However, the minority rule applied to an odd number of agents makes stable convergence impossible, forcing the agents to ever adapt towards a non-existing equilibrium, generating fluctuations around the mean attendance. In order to take this into account, the concept of *volatility* was introduced in [15]. Volatility is taken to be the variance (σ^2) or standard deviation (σ) of the attendance of a reference group, representing how spread are agents from the mean attendance, thus as volatility increases, so does the "waste" of resources, that is, more agents are losing the game by not being in the minority group and the farther the system is from an optimum group behaviour. Because of this, volatility is taken to represent the efficiency of group choices. It is well known that the volatility of a system is highly dependent on the agents' memory size and presents a phase transition when plotted against this parameter [10]. Whether or not it is dependent of the memory *content*, seems to remain an open question [3, 4].

The fixed nature of the strategies in the MG has led researchers to develop other learning paradigms without such bounds, in order to study the effects on the dynamics of the system. In [13], a genetic algorithm is applied to the agent's strategies and in [1] agents are able to learn through a reinforcement learning algorithm (Q Learning). In both cases, results have shown that the characteristic phase transition disappears. However, it is unclear what is the role of the learning algorithms used in producing such dynamics. In order to better understand this point, we shall present a very simplified evolutionary-based learning algorithm and study how the dynamics change when a learning parameter is varied. Section 2 presents the general framework for our numerical analysis. Section 3 shows the results and discuss them, while Section 4 concludes this paper.

2. Simulation Framework

We start by making some definitions and describing the MG and our modification to the learning algorithm. The MG model is an iteractive system composed of N agents, where N is an odd number, two groups here labeled *A* and *B*, and a global history recording which was the minority group in past rounds. An agent is a system that, at each round, reads a window of size M from the global history and outputs a decision either to attend to group *A* or to group *B* in the next round using solely the information contained in this window. We will refer to the number of agents attending to group *A* in time t as $\mathcal{A}(t)$. The group *B* is always the complement of *A*, $\mathcal{B}(t)$ = N - $\mathcal{A}(t)$, and we shall use group *A* as a reference group. We are interested in evaluating the variance of $\mathcal{A}(t)$ as a measure of how spread the agents are from the mean attendance and, thus, also as a measure of the efficiency of group choices. We use the standard formula to calculate the variance in a population:

$$\sigma^2 = \frac{1}{(T - t_0)} \sum_{t=t_0}^{T} (\mathcal{A}(t) - \langle \mathcal{A} \rangle)^2$$

where $\langle \mathcal{A} \rangle$ is an average over $\mathcal{A}(t)$, T is the total number of rounds executed and t_0 is the round after which the system reaches a steady state, i.e. the round after which fluctuations due to the random initialization are gone [9]. We take t_0 as being a much larger number compared to M, $t_0 >> M$, thus guaranteeing that the window read by agents is composed only of endogenous information, i.e. created by the agents themselves and not by some external process.

To make the decisions, each agent is given $S \geq 2$ strategies. Qualitative behaviour of the system is well known not to change with S if $S \geq 2$ [10] and thus we focus on the simplest case where $S = 2$. A strategy relates history windows to a decision and can be seen as a lookup table that maps each combination of past attendances to a decision. Figure 1 shows an example of such strategy for $M = 3$, where A is represented by "0 and B by "1. Thus, a strategy can be effectively represented as a binary vector of size 2^M, with each vector position representing the output of one unique possible combination of history's outcomes. Each agent randomly draws S strategies on its creation from the strategy space (i.e. the strategy vector is randomly initialized) and, in the traditional MG, agents must use only these

A(-3)	A(-2)	A(-1)	Decision
0	0	0	1
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	0

Figure 1. An example of a strategy as a lookup table with $M = 3$

strategies during the entire course of the game. After each round, agents update the *fitness* of each strategy, which is a measure of how well the strategy performed in the past. One possible fitness function, described in [5] and used here, is to test whether a strategy *would* have predicted correctly the group containing the minority in the last M rounds, giving points to each correct prediction. When making a decision, an agent uses his higher fitness strategy in order to commit the decision.

The above setup is known to present a phase transition in its efficiency, as a function of the agent's memory size. This is one of the main features of the MG, and suggests that agents, for some memory sizes, are able to organize themselves in order to increase system efficiency, reaching a better than random volatility [10]. Also, this behaviour is similar to some physical systems and has attracted numerous physicists to the area. In order to further study the effects of different learning algorithms in the dynamics of the system, we extend the MG model by adding a simple evolutionary-based learning procedure. This evolutionary algorithm follows the following steps:

1. Order the strategies by their fitness;
2. Discard the $\frac{S}{2}$ worst strategies;
3. Generate $\frac{S}{2}$ new random strategies and replace the discarded ones;

The above algorithm is invoked with period L for each agent in the game. This guarantees that, at every L rounds, new strategies are generated and the underperforming ones are eliminated. Values of L are integers from 1 to ∞. It is interesting to note that *all* agents update their strategies with the same periodicity, contrasting to [13] where only a fraction of the worst performing agents actually invoke the learning algorithm. We take this approach in order to avoid agents to receive any information about the performance of other agents, making any individual agent concerned only with its own performance. Another parameter of the algorithm is how fitness is attributed to new strategies. We here analyse two cases, one where all new strategies begin with zero fitness and another where new strategies inherit fitness from its predecessors, i.e. from the strategies removed from the set.

This kind of evolutionary learning, where underperforming results are discarded and new ones are generated for substitution, is well known and studied, presenting good results for several "conventional problems [8]. However, the problem faced here is somewhat unconventional: an agent must learn not a pattern generated externally by the environment, but one that is generated dynamically by himself and other similar agents, in a competitive environment. Also, we are not interested in how well this algorithm might perform for any individual agent. Our interest lies on the behaviour of the system as a whole, by the analysis of its dynamics.

3. Results

We have run all experiments using the following fixed parameter set: $N = 101$, $S = 2$, $t_0 = 1000$ and $T = 3000$. We begin by analysing the role of the learning periodicity parameter L in the extreme cases. When $L = 1$, at every round the learning algorithm is invoked and the worst performing strategies are discarded and new ones are generated. In this case, since fitness attribution is done once at every round, strategies are chosen to be discarded based only on their performances in the previous round. At the other extreme, when $L \geq T$, strategies are never discarded nor new ones are generated, thus the system behaves exactly like the traditional MG.

Figure 3 shows these two extreme cases when new strategies inherit fitness from its predecessor, plotting the variance attained against the agent's memory size. For $L \geq T$, the curve is exactly the same as the traditional MG, presenting the characteristic phase transition. That was expected since in this case the evolutionary algorithm is never invoked and the system becomes the traditional MG. However, it can be seen that when $L = 1$, the system has a qualitatively different behaviour, presenting no phase transition, with the minimum variance corresponding to the point where the memory is smaller, increasing with the increase of M. This is precisely the behaviour reported in [1], where agents are able to learn through a Q-Learning algorithm. For large M, both cases converge to the same variance, which is the variance expected when agents make independent random choices at every round, assuming a value of $\sigma^2_{rand} = \frac{N}{4}$ [7].

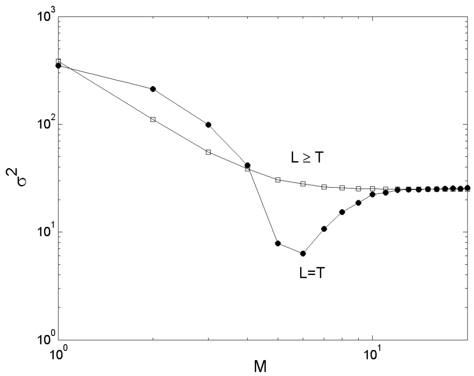

Figure 2. Variance σ^2 as a function of M when new strategies starts with zero fitness. Filled circles are for $L \geq T$ and empty squares for $L = 1$. Plotted values are averages over 100 independent runs.

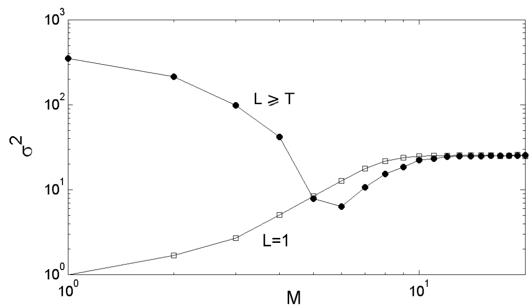

Figure 3. Variance σ^2 as a function of M when new strategies inherit its predecessor's fitness. Filled circles are for $L \geq T$ and empty squares for $L = 1$. Plotted values are averages over 100 independent runs.

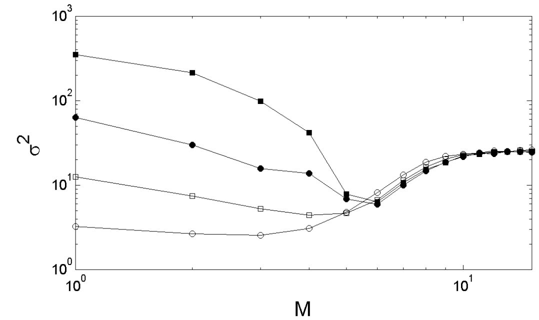

Figure 4. Variance σ^2 as a function of M when new strategies inherit its predecessor's fitness. Empty circles are for $L = 100$, empty squares are for $L = 500$, filled circles are for $L = 1000$ and filled squares are for $L \geq T = 3000$.

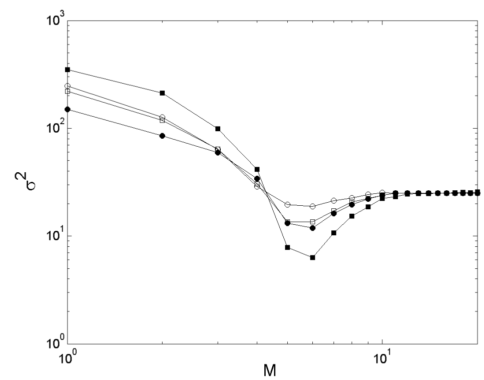

Figure 5. Variance σ^2 as a function of M when new strategies starts with zero fitness. Empty circles are for $L = 100$, empty squares are for $L = 500$, filled circles are for $L = 1000$ and filled squares are for $L \geq T = 3000$.

Figure 2 plots the results when the experiment is repeated but new strategies now starts with zero fitness. A completely different behaviour is observed. For $L = 1$, where there was, in the previous case, a strictly increasing efficiency with the increase of memory size, there is now a strictly decreasing function, with the most remarkable aspect being that it never behaves in a better than random way. This is also the observed behaviour in [13], where a genetic algorithm is used to produce new strategies.

This stands as evidence that the macroscopic dynamics found in [13] and [1] are not particular of the learning paradigm used; rather, these seem to be a more general effect of systematically ruling out bad strategies from the system, combined with a fitness attribution scheme for the strategies. One might question if the described behaviors might not be simply due to the periodical introduction of new strategies into the game. To account this, we have modified the algorithm to randomly choose the strategies to be replaced (not considering their fitness). In this case the described behaviors are not attained, rather the variance becomes equal to the random choice case for all memory sizes, for both fitness attribution schemes. Also, when the algorithm is further modified to replace the *best* strategies, the same random behaviour is attained. Thus, the selective removal of underperforming strategies plays an important role in our setup.

Next, we analyse how the variance behaves when we set $1 < L < T$. Figure 4 shows the variance behaviour for some values of L when new strategies inherit the fitness of its predecessors. Figure 5 shows the same experiment, but new strategies receive zero fitness. It can be seen that in both cases the behaviour smoothly transits between the two extreme cases. The phase transition becomes more evident as L increases, thus evidencing that the phase transition behaviour of the traditional MG is a consequence of not allowing agents to modify or discard their underperforming strategies throughout the game. When we allow the worst strategies to be discarded and substituted with maximum frequency, i.e. $L = 1$, the behaviour changes dramatically as was shown.

4. Conclusions

In this paper we have presented a simple learning algorithm for the Minority Game where, with periodicity L, the worst performing strategies of each agent are removed and new random strategies are created for replacement. This simplified learning procedure was used to further investigate the role of learning in this kind of game. We showed that for small enough L the phase transition of variance when plotted against memory size, characteristic of the original MG, disappears. Similar results are reported by other researchers and we have compared ours to those, showing that such dynamics might be attained using a much simpler learning rule. We attribute this change in behaviour to the possibility of agents to adjust their strategies by the systematic removal of underperforming strategies. Also, we have investigated how does the learning parameter L affects the efficiency of the system, showing that there is a smooth transition between the two extreme cases ($L = 1$ and $L \geq T$), evidencing that the phase transition is due the unability of agents to adjust their strategies in response to bad performance.

Moreover, we have found out that the fitness attribution scheme used have a high influence on general behaviour. When new strategies start with zero fitness the system never reaches a better than random variance, becoming a strictly decreasing function with memory size. However, when we allow new strategies to inherit fitness from its predecessor the variance becomes strictly increasing, with high efficiency present on a large range of memory sizes. This result shows that any learning algorithm applied to the MG should take fitness attribution into account, for this aspect alone might change dramatically the observed behaviour.

These results are important to better understand the role of the learning paradigm used in the MG and other evolutionary games. By showing that such a simple learning model is able to capture the main features offered by more complex models, and by analysing the effects of changes on simple parameters of this model, we believe we have taken a step closer to understanding how different learning procedures affect the dynamics of such games.

Acknowledgements This work has been partly supported by CNPq and FAPERGS.

References

[1] M. Andrecut and M. Ali. Q learning in the minority game. *Physical Review E*, 64(067103), 2001.

[2] W. B. Arthur. Inductive reasoning and bounded rationality. *American Economic Review (Papers and Proceedings)*, 84:406–411, 1994.

[3] A. Cavagna. Irrelevance of memory in the minority game. *Phys. Rev. E*, (59), 1999.

[4] D. Challet and M. Marsili. Relevance of memory in the minority game. *Phys. Rev. E*, (62), 2000.

[5] D. Challet and Y.-C. Zhang. Emergence of cooperation and organization in an evolutionary game. *Physica A*, (246):407–418, 1997.

[6] F. Chow and H. Chau. Multiple choice minority game. *Physica A*, 319:601–615, 2003.

[7] A. R. de Cara, O. Pla, and F. Guinea. Competition, efficiency and collective behavior in the ”el farol” bar model. *Eur. Phys. J. B*, 10:187–191, 1999.

[8] D. Fogel. *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*. IEEE Press, 2a. edition, 2000.

[9] D. Fogel, K. Chellapilla, and P. Angeline. Inductive reasoning and bounded rationality reconsidered. *IEEE Transactions on Evolutionary Computation*, 3(2):142–146, July 1999.

[10] E. Moro. The minority game: an introductory guide. In E. Korutcheva and R. Cuerno, editors, *To appear in ”Advances in Condensed Matter and Statistical Physics”*. Nova Science Publishers, 2004.

[11] A. Rubinstein. *Modeling Bounded Rationality*. Zeuthen Lecture Book Series. The MIT Press, Cambridge, Massachussets, 1998.

[12] H. Simon. A behavioral model of rational choice. *The Quarterly Journal of Economics*, LXIX, February 1955.

[13] M. Sysi-Aho, A. Chakraborti, and K. Kasti. Intelligent minority game with genetic-crossover strategies. *Eur. Phys. J. B*, 34(00234):373–377, 2003.

[14] A. Tversky and D. Kahneman. The framing of decisions and the psychology of choice. *Science*, 211.

[15] Y.-C. Zhang. Modeling market mechanism with evolutionary games. *Europhysics News*, March/April 1998.