Probabilistic learning and emergent coordination in a non-cooperative game with heterogeneous agents: An exploration of "Minority game" dynamics

G. Bottazzi *
G. Devetag †

August 10, 1999

Abstract

In this paper we present results of simulations in which we use a general probabilistic learning model to describe the behavior of heterogeneous agents in a non-cooperative game where it is rewarding to be in the minority group. The chosen probabilistic model belongs to a well-known class of learning models developed in evolutionary game theory and experimental economics, which have been widely applied to describe human behavior in experimental games.

We test the aggregate properties of this population of agents (i.e., presence of emergent cooperation, asymptotic stability, speed of convergence to equilibrium) as a function of the degree of randomness in the agents' behavior. In this way we are able to identify what properties of the system are sensitive to the precise characteristics of the learning rule and what properties on the contrary can be considered as "generic" features of the game.

Our results indicate that, when the degree of "inertia" of the learning rule increases, the market reaches a higher level of allocative and informational efficiency, although on a longer time scale.

1 Introduction

In this paper we make a first attempt to investigate if and how alternative hypotheses about learning may influence the aggregate long-term properties of a simple non-cooperative game. The general framework analyzed is a multi-agent game called the "Minority game", whose structure is intended to capture, although in a highly stylized and abstract way, some basic properties of speculative market interactions.

*S.Anna School of Advanced Studies, Pisa, Italy.

†S.Anna School of Advanced Studies, Pisa, Italy. Preliminary draft. Please, do not cite without permission, but comments are very welcome. We wish to thank Giovanni Dosi for useful comments and suggestions on earlier versions of this work. Financial support from the Italian Ministry of University and Scientific and Technological Research is gratefully acknowledged

In the "original" version of this game, first introduced by [1], a population of N artificial agents (where N is an odd number) must each simultaneously and independently choose between two sides, say 0 and 1. The side chosen by the minority of the agents, i.e. the "minority" side is the winner, and agents who choose it are awarded one point each, while those who choose the majority side win nothing. Each agent is initially endowed with a fixed number of strategies (which will be defined more in detail later), and updates them throughout the game according to a deterministic algorithm.

The game intends to reproduce, at least in first approximation, the core of speculation activities in financial markets, where agents form beliefs about the market future outcomes (determined by the behavior of the majority of agents operating on it) and try to "beat" it by acting in an opposite way. This type of speculative activity is sometimes referred to in the financial literature as "contrarian investment strategy", meaning the practice of trying to speculate on perceived investor sentiment [11]. From a game-theoretic point of view, the game is a multi-agent coordination game with several asymmetric equilibria in pure strategies, and a unique symmetric mixed-strategy equilibrium. As the population playing a minority game is supposed to be generally large as to have no chance to communicate, one can investigate the conditions under which repeated interaction among players cause some forms of aggregate self-organization to emerge spontaneously. The major results already obtained in this vein will be discussed later in the section.

Besides simulation studies, some experimental studies have also been conducted on similar types of coordination games[2] The data coming from experiments suggest that the degree of self-organization generally depends on the characteristics of the game being played in terms of, e.g., payoff function, amount of information available to players, number of repetitions and size of the population. These variables, in fact, not only modify the incentive structure involved but, more importantly, determine the type of adaptive behavior (or learning) that players will exhibit throughout the game. This latter point is especially relevant in the context of a multi-agents game of the "minority" type, where the kind of adaptive dynamics and, to a lesser extent, the information available to agents are likely to substantially modify both the long-term outcome and the "collective" adjustment process itself.

Our scope in this paper is to study the variation of the asymptotic properties and dynamics of a population playing a minority game when the learning rule of the agents is modified. In particular,we adopt a probabilistic learning algorithm for the agents and leave any other parameter of the original game unmodified. The choice to adopt a probabilistic learning rule is supported by the available evidence on human learning in games, which suggests that learning processes in various interactive contexts can be quite accurately described by simple probabilistic models [6, 4, 5]. We demonstrate that such simple variation in the agents' learning algorithm produces important modifications to the asymptotic properties of the system, suggesting that some features of the original game are not generally valid but strongly depend on the particular behavioral assumptions made.

In particular, such modifications all yield improvements in the system asymptotic performance; the improvement is highest for certain values of the game parameters and for higher degrees of "inertia" in the learning algorithm (measured by a parameter β), although at the expense of a longer adjustment phase.

The remainder of the section briefly illustrates the basic features of the original game, and describes the major results of the system asymptotic properties present in the literature. Section 2 introduces the probabilistic learning rule adopted for our simulations, and introduces the inertial parameter β.

In section 3 the length of the "training phase" is analyzed. In fact, although we are mainly concerned with studying the system optimality properties in the long run, a non secondary aspect concerns the duration of the adjustment phase, which is generally increased by the introduction of a probabilistic rule. The "transient length" issue, although primarily a technical one, has important theoretical implications for our model, in that it highlights a tradeoff between longer times and better performances, which is common to all probabilistic search algorithms.

We subsequently analyze the system aggregate performance and in doing so we define several measures of efficiency. Section 4 defines the notion of "allocative efficiency" as in [1, 3], strictly connected to the size of the minority; in fact, the smaller is the winning minority the more points are "left on the table" instead of being distributed over the population. The influence of the parameter β on the degree of the system allocative efficiency is analyzed. We then compare the degree of efficiency so obtained with the level of efficiency theoretically attainable by perfectly rational and perfectly informed players who "solve" the game analytically.

Section 5 analyzes the effect of β on the degree of "informational efficiency", connected to the existence of arbitrage opportunities. At the end of this section, we analyze the influence of β on the system's degree of social optimality (uniformity of earnings ditribution over the population). Finally, Appendix A analyzes modifications to our results when a time discounting factor is added to the learning algorithm.

The results obtained in the various sections are strongly consistent and show that the introduction of randomness in the learning rule has a positive effect on all the three types of efficiency introduced. Besides, such positive effect is greater the greater amount of "inertia" is assumed. Section 6 contains some final remarks and suggestions for future research.

Let us start our analysis by briefly recalling the basic features of the "original" game.

In this game [1] all players after each round know only which side (0 or 1) was the winner, without knowing the actual "size" of the minority. The market signal is represented by the (history) H of the game, that is a time series modeled as a binary string specifying which side has won at every stage. The degree of rationality of the agents is determined once for all by the specification of two parameters homogeneous over all the population.

The first parameter is the amount of "memory"of the past that agents are able to retain, corresponding to the last m bits h_m of the game history H. The second parameter is the number s of strategies assigned to each agent.

A strategy is defined as a prescription on the action to take on the next round of play (i.e. to choose 0 or 1) after a particular history (that is, a particular sequence of m bits) has been observed up to that point. For example, in the simple case in which $m = 3$, a strategy is defined as follows:

that is, the "history" columns specify all the possible histories of the game in the last m periods; the "prediction" columns specify which action to choose on the next trial in correspondence to each particular history observed. The

history	prediction	history	prediction
000	1	100	1
001	0	10001	0
010	0	110	1
011	1	111	0

Table 1: Example of strategy with $m = 3$.

strategies are randomly drawn from a common pool consisting of the 2^{2^m} ways of assigning all the 2^m possible strings of length m to an action. Note that even if m and s are the same for all the population, heterogeneity follows from the random initial strategy assignment. Each strategy in play will be characterized by a value $q_i(t)$, which indicates the total number of points accumulated by strategy i at time t. Indeed, after each period of the game, all the strategies that have predicted correctly in that period (that is, all strategies prescribing the side ex post resulting the winning side) are assigned one point each. In other words, all strategies which, if played, would have been winning on a particular round, are all updated regardless of whether they were actually played or not. Note that the procedure of strengthening strategies that were successful in the past certainly sounds plausible and it is also the core of the so called "reinforcement learning" algorithms, which are widely used to model agents behavior in low-rationality, low-information environments. However, unlike the original minority game, reinforcement learning implies that *only* strategies that were actually played get strengthened. The learning rule in the original minority game, hence, differs from reinforcement learning *stricto sensu* in assuming on the part of the agents a higher degree of rationality.

Given strategies and updating rules, behavior at each stage is completely deterministic, in that each agent at each period plays, among the strategies he possesses, the one with the highest number of accumulated points.

In order to judge the system's degree of self-organization, it is necessary to introduce a measure of allocative efficiency. A natural candidate is provided by the average number of players belonging to the winning party. Such quantity measures the degree to which the system is in equilibrium. In fact, when the winning party is equal to $N/2$ the system finds itself in equilibrium in the sense that no player can do better by unilaterally deviating. Besides, the equilibrium configuration is also globally efficient, in that the maximum number of players win and the highest number of points are distributed over the population. Otherwise if this number is near 0 (or N), few players win and less points are allocated. Instead of averaging the size of the winning party one can choose to compute an associated quantity, the mean squared deviation from the half population σ. Let N be the number of agents and $N_0(t)$ the number of agents attaining side 0 at time step t, then in a given simulation of length T the mean squared deviation is computed as

$$\sigma = \frac{1}{T} \sum_{\tau=0}^{T} (N_0(\tau) - \frac{N}{2})^2 . \tag{1}$$

Note that σ is also a measure of the fluctuations around the $N!/(((N-1)/2)!)^2$ game Nash equilibria in which exactly $(N-1)/2$ players form the winning

minority.

Before starting to describe the results from the simulations performed in [1, 3], few remarks are appropriate.

Any simulation depends, other than on the parameters N, m, and s, on the initial distribution of strategies among agents and on the initial history, both generated randomly for the system. Therefore, if not stated otherwise, all the quantities shown in the plots are obtained via an averaging procedure over 50 independent simulations with randomly generated histories and strategy distributions. This averaging procedure is performed in order to produce asymptotically stable quantities, i.e. a different resampling with different initial histories and strategies will produce an equal "asymptotic" state for the system.

Moreover at the beginning of each simulation the system is left evolving for a "training phase" of length T_0 in order to eliminate any eventual transient effect on the subsequent averaging procedure. The quantities so obtained can be considered "asymptotic" properties of the system as long as T_0 and T are chosen high enough to provide a good approximation of the $T \to \infty$ limit. As we will see later, a sensible choice for T and T_0 is far from trivial in a generalized setting.

The dependence of the volatility σ on N, m and s for the original minority game has been studied in many works [1, 3] and is summarized in Fig. (1) for $s = 2$. As noticed by [3] the parameter $z = 2^m/N$ turns out to be, at least in first approximation, the relevant one and the curves for various N collapse if plotted in this variable. Various explanation of this peculiar feature has been proposed [1, 3]. Notice that even if the actual number of possible strategies is 2^{2^m}, their relative strengths are completely defined in term of the frequency $P(0|h_m)$ with which, in history, a 0 follows a given $m-$length string h_m and there are 2^m of such variables. So, z can be interpreted as the density of agents in the strategy space degrees-of-freedom.

Looking at Fig. (1) tree different "regimes" of the system can be identified: a "random regime" occurs when z is large (the agent are sparse in the strategy space), and the system can hardly organize. In fact its behavior can be described as a collection of random agents that choose their side with a coin toss. In fact suppose the past history be a given h_m and suppose there are $N_d(h_m)$ agents whose strategies prescribe differently based on that history while there are $N_0(h_m)$ and $N_1(h_m)$ agents whose strategies prescribe the same party (we restrict ourselves to the $s = 2$ case), respectively 0 and 1. If the agent in N_d choose randomly the variance is $\sigma(h_m) = N_d(h_m)/4 + (N_0(h_m) - N_1(h_m))^2/4$. The average over the possible h_m will then give $\sigma = N/4$. Notice that there are two different contributions to σ: a fluctuation in the choices of agents able to choose and a fluctuation in the initial distribution of strategies.

The second regime is the "inefficient regime" for $z << 1$. Here the agents densely populate the strategy space and they in fact "coordinate" in the sense that their actions are strongly correlated. This coordination however leads to a worsening of the overall performance due to a "crowd" effect [7]: the agents in fact are too similar to each other and they tend all to choose the same party based on the information available.

The third regime for $z \sim 1$ is where the coordination produces a better-than-random performance. Here the agents are enough differentiated so as not to produce "crowd" effects but sufficiently distributed over the strategy space so as not to produce a random-like behavior. The point where σ is minimum

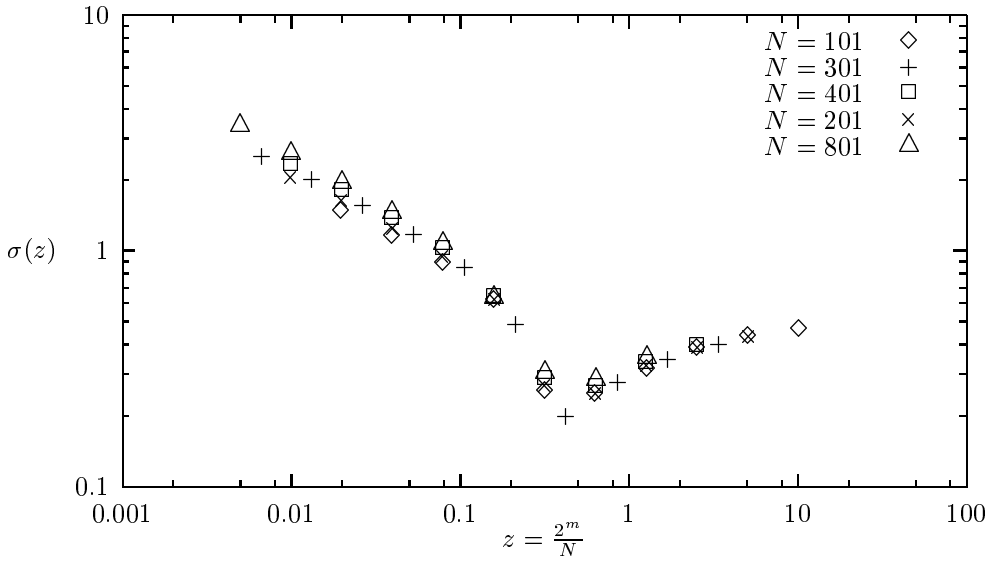

Figure 1: The volatility $\sigma(z)$ for $s = 2$ and different values for N and m

is referred to in the literature as the "critical point" z_c suggesting that a major change in the system behavior happens when this point is crossed. As we will see in the following sections, this "criticality" survives and in some sense more clearly appears when generalizing the learning rule initially proposed.

From now on we will restrict ourselves to the case $N = 101$ and $s = 2$ so we will speak of the "optimal" value for memory length m_o referring to the value of m which minimize σ[1]

while we prefer to drop the word "critical" as it brings to mind special features of physical systems which are not clearly perceived in our simulations.

2 Learning dynamics

The notion of strategy in the minority game is relatively unusual in standard or evolutionary game theory, and it requires some interpretation in terms of the behavioral and cognitive characteristics of the human players it intends to describe.

According to us, each strategy may been seen as a particular "mental model" or "hypothesis" about the world (the "world" may include values of the fundamentals of the market in question, or, as it seems appropriate in this case, the beliefs and behavior of the other players). Each general hypothesis then translates into specific predictions on which will be the winning action for each particular history observed so far. In this respect, a strategy in this game resembles the notion of "repeated game strategy" in standard game theory (see,

[1] Note that the values chosen for m and N conform to what found in [1, 3]; the choice to set s=2 is justified by the fact that the system exhibits the same qualitative properties for any $s \geq 2$, while reducing to a trivial case for s=1

e.g.,[13] for an introduction).

Each agent initially has a number of different strategies, that is a number of different (and competing) hypotheses. After each period, more evidence is collected and all the hypotheses consistent with the evidence are updated through a process that is very similar in spirit to Bayesian updating.

From a behavioral point of view, both the definition of strategy and the choice of the learning rule to adopt are particularly demanding in terms of the degree of rationality of the agents. In fact, players not only are supposed to form several hypotheses about the game, but they also update them consistently, de facto applying sophisticated conterfactual forms of reasoning. On the other hand, as previously stated, literature on experimental games has shown that behavior of human subjects in games can often be well approximated by simple adaptive learning rules which act probabilistically. For example, the reinforcement learning model [4, 6], originally developed in psychology, has been shown to accurately describe medium and long run behavior in a large class of games in which agents have limited information and feedback. A more recent probabilistic model developed by Camerer [5], more in line with the algorithm initially proposed for the minority game, extends the updating mechanism also to actions that were not played but which would have been successful, according to what he calls the "law of simulated effect".

In the present paper we introduce a simple modification of the standard learning rule: the updating mechanism is left unaltered (that is, all winning strategies are updated regardless of whether they were played or not), but the choice between strategies in each period is probabilistic instead of deterministic.

Remember the definition of $q_i(t)$ as the total number of points strategy i would have won if played until time t then each agent chooses among his strategies based on the following probability distribution:

$$p_i(t) = \frac{e^{\beta q_i(t)}}{\sum_j e^{\beta q_j(t)}} \,. \tag{2}$$

where the sum on j is over all the strategies possessed by the player [2]. Note that in general, different players will assign different probabilities to the same strategy due to a different strategy endowment.

Our model bears similarities with a discrete time replicator dynamics [8]. The parameter β can be considered as a sort of "willingness to choose": when it is high, the agents are sensitive even to little differences in the virtual score of their strategies and in the $\beta \to \infty$ limit the usual minority game rule is recovered. On the contrary for low values of β a great difference in strategy strengths is necessary in order to obtain significant differences in probabilities.

The connection of (2) with the replicator dynamics is straightforward if one looks at the probability updating equation associated with it:

$$p_i(t+1) = p_i(t) \frac{e^{\beta\, \delta q_i(t)}}{\sum_j p_j(t) e^{\beta\, \delta q_j(t)}} \,. \tag{3}$$

where $\delta q_i(t) = q_i(t+1) - q_i(t)$ are the points won by strategy i at time t. If one thinks of a continuous process $\delta q_i(t) = \dot{q}_i(t)\delta t$, where $\dot{q}_i(t)$ is the instantaneous "fitness" of strategy i, then the continuous time replicator dynamics equation is recovered keeping only the first terms in δt expansion.

[2] For our $s = 2$ case, the summation will contain two terms

3 Transient length

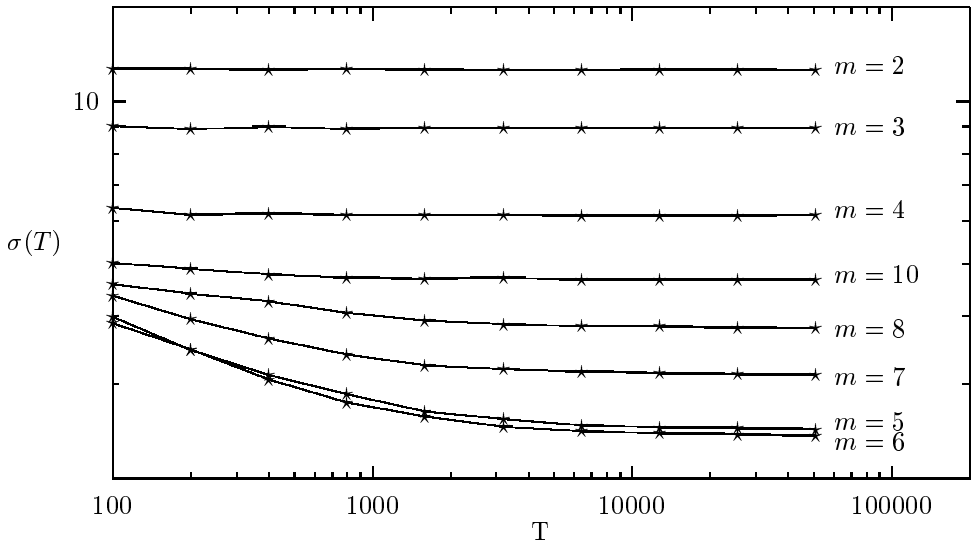

Figure 2: The mean σ as a function of the run length for different m. The points are averages over a sample of 30 independent runs with $N = 101$ and $s = 2$

Let us go back to the problem of defining the correct values for T_0 and T in (1). The central question is: How long must the system be left evolving before it reaches the asymptotically stable dynamics? In the minority game analyses found in the literature [1, 3], the general answer is "long enough", where "enough" is typically 10.000 to 100.000 time steps for an agents population ranging from 100 to 1000 units.

Fig. (2) plots the average σ value for the original minority game as a function of the time length T over which this average is taken with a transient $T_0 = T$. As it can be seen from the graph, the values used in the literature on the minority game are generally sufficient to obtain a prediction correct to a few percent. However, two things are worth noticing:

- For low values of m, in the "inefficient regime", and for high value of m, the "random regime",the system reaches a stable dynamic quite fast. On the contrary, for values of m near the optimal value m_o, the system takes a longer time to self-organize.
- The system approaches the asymptotic value from above, which suggests the intuitive interpretation that the system "learns" to self-organize with time.

Consider now the case in which the learning rule is the one described in (2). For high values of β this learning rule approaches the standard one, and accordingly, the transient length is similar to the one found in the previous case.

However, as β decreases, such length generally increases. The increase is most dramatic for values of m near the optimal value m_o, and it progressively disappears for higher values of m, as can be seen in Fig. (3). Such a result is somewhat intuitive if one considers the meaning of β in terms of the learning rule. Supposing a non trivial dynamics for m near m_o, the parameter β sets the time scale on which such dynamics is attained.

As a suggestive explanation of the statement above, consider the following argument:

Let be $r(t) = p_1(t)/p_2(t)$ the ratio of the probabilities that an agent associates to her two strategies, and $\Delta q(t) = q_1(t) - q_2(t)$ the difference in their respective strengths. From (2) it follows that $r(t) = e^{\beta \Delta q(t)}$. Assuming that the difference in the two strategies performance holds constant over time, assumption which is generally true in the initial transient regime where agents' behavior is substantially randomic, we obtain $\Delta q(t) \sim t$; hence, from the equality above, a given difference in probability is obtained at a time which is inversely proportional to β.

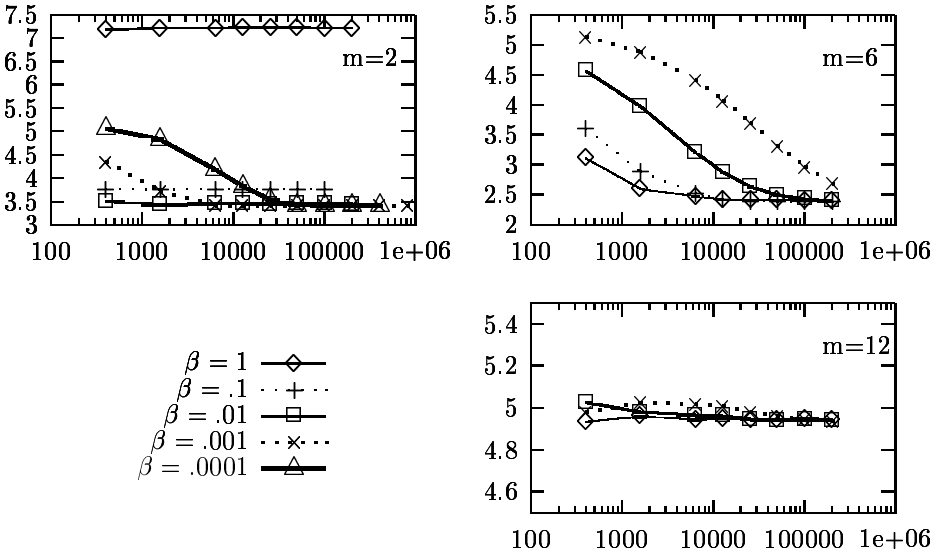

Figure 3: σ as a function of run length T for different β. The points are average over a 30 runs sample with a transient time $T_0 = T$.

In order to estimate the time scale over which stability is attained, we use the following procedure: Holding all the parameters and the initial conditions constant, the system volatility can be expressed as a function of both the "transient" phase duration, and of the time length over which it is averaged, i.e. $\sigma = \sigma(T, T_0)$.

Starting from a reference time T_r,[3] we compute the mean volatility progres-

[3]Note that the chosen value for T_0 is irrelevant as long as it is small compared to the typical time scale.

sively doubling t and t_0, and thus obtaining a series of values $\sigma_n = \sigma(2^n T_r, 2^n T_r)$.

When the relative variation $|\sigma_n - \sigma_{n-1}|/\sigma_n$ falls below a fixed threshold ϵ, we stop and take the last computed value of σ as an estimate of its asymptotic value. The corresponding time length $\hat{T}(\epsilon)$ will be an estimate of the time implied by the system to reach this asymptotic stability.

As can be seen in Fig. (3) the increase in $\hat{T}$ when β is lowered is mainly concentrated around m_o, with shapes that suggest the presence of a discontinuity.

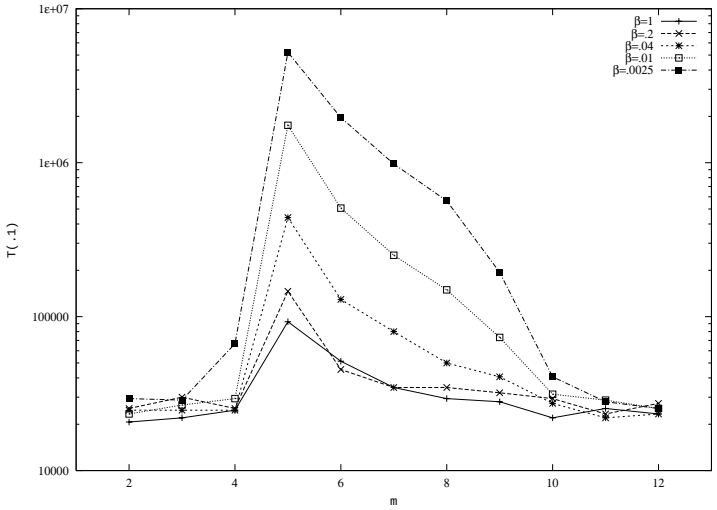

Figure 4: A rough estimate of the time $\hat{T}$ needed by the system to reach the stable asymptotic σ value with an error not greater than a few percent. The plot is made against m for different value of β. see the text for a description of the method.

4 Allocative efficiency

In order to analyze the asymptotic properties of $\sigma(m)$ for different β, we use the same procedure described above regarding the calculation of $\hat{T}$, i.e. we leave the system evolve until stability is reached. The simulation results are plotted in Fig. (5). As can be seen when β decreases, the system performance level generally increases. Such increase is larger the lower the value of m, and it becomes negligible for $m \geq m_0$. The observed behavior is consistent with the idea that for high values of m, the system dynamics is completely determined by the initial distribution of strategies among players, and the players have no opportunities to attain a higher performance by adjusting their behavior. Therefore, the particular learning rule used is largely irrelevant. On the contrary, for low values of m, the original learning rule ($\beta = \infty$) produces a "crowd effect" [9] (consisting in large groups of agents choosing the same side) that, due to homogeneity in the initial strategy endowments, prevents the system from attaining a high degree of efficiency.

In some sense, one can interpret the crowd effect as a collective form of "overreaction" [10]. Of course, introducing a probabilistic learning rule for the

strategy choice acts like a brake that dumps the amplitude of such correlated fluctuations. At the individual level, this can be interpreted as the presence of higher degree of "inertia" as agents update their probabilities more slowly. In other words, as β decreases each agent behaves as if he was applying a sort of "fictitious play" approximation [**?**][4], indeed assuming stationarity on the distribution of other agents choices. This assumption is in fact consistent: a decrease in β makes the behavior of the population as a whole change at a slower pace.

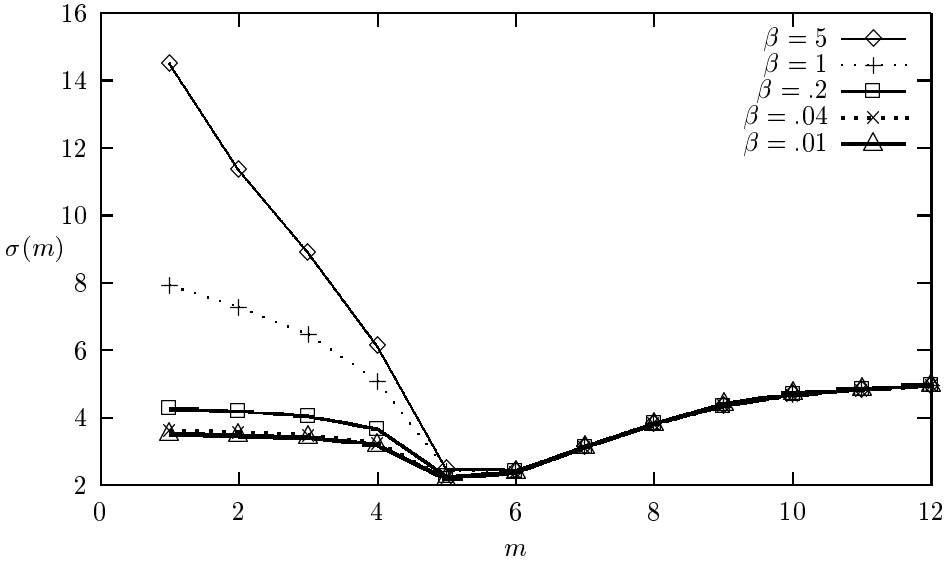

Figure 5: The volatility σ for $s = 2$, $N = 101$ and various m and β. The run are performed doubling the time length T until the last two value are different by less then 1%.

The double effect of a slower probability updating at the individual level and the resulting more stable collective behavior implies that σ is a non increasing function of β. In fact, if the system reaches a dynamical stability via an averaging procedure over the past outcomes, increasing the time scale over which the averaging procedure is taken cannot rule out previously attainable equilibria.

However, ote that if one performs the simulations with a fixed time length, when β becomes small the system behavior resembles the behavior of a random system. This finding is due to both the increase in the transient length and the purely randomic starting dynamics which occur when β is decreased. Here we are facing a double limiting problem: we are interested in the value of volatility in both $\beta \to 0$ and $T \to \infty$ limit and therefore it is necessary to specify which limit is taken first. The results of the fixed time simulations are plotted in Fig. (6) and are in line with [14].

[4]Note that fictitious play implies that a player always best responds to the observed fre-

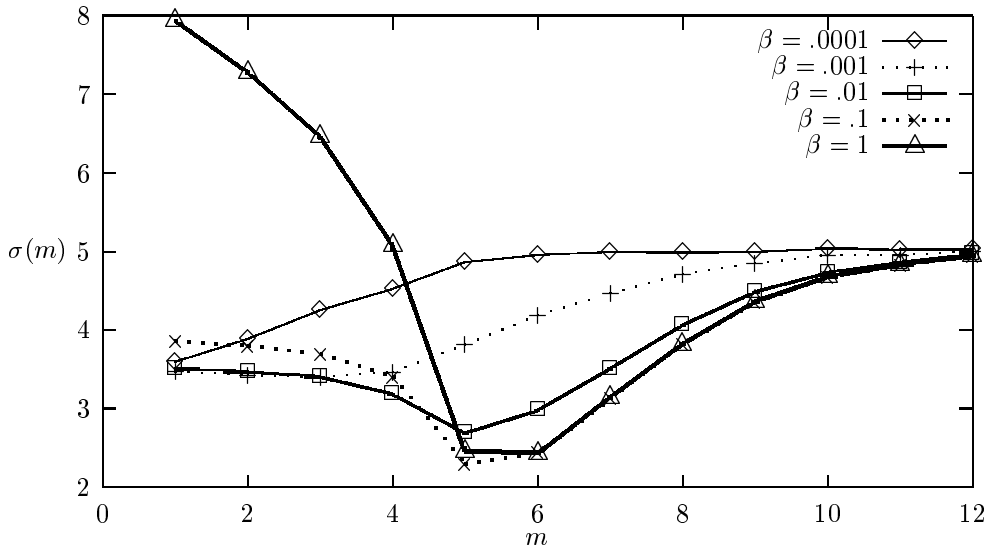

Figure 6: The volatility σ for $s = 2$, $N = 101$ and various m and β. The runs are performed with a fixed time length $T = T_0 = 10000$. When $\beta \to 0$ the system approaches a collection of randomly choosing agents.

The performance attainable in the minority game via a dynamical organization of agents with limited information and ability to choose is actually surprisingly high, compared to the efficiency attainable with more informed and more rational agents who are endowed with a greater flexibility in choice.

Consider for instance a collection of agents characterized, in line with the original minority game, as follows: each agent is assigned $S = 2$ strategies, and a vector of length 2^m containing the probability $p(h_m)$ of playing according to the first strategy after the appearance of h_m. Moreover, for each h_m, each agent knows the values of $N_0(h_m)$, $N_1(h_m)$ and $N_d(h_m)$ indicating respectively the number of agents for which their strategies prescribe both to play 0, both to play 1 or to play differently.

Assuming that the game structure and the amount of information available to agents is common knowledge and assuming the agents are perfectly rational the problem completely factorizes and for each h_m every agent in $N_d(h_m)$ will solve the game analytically choosing $p(h_m)$ in order to minimize

$$\frac{(N_1(h_m) - N_0(h_m))}{2} - p(h_m)N_d(h_m) \tag{4}$$

i.e. to make the average value of people choosing a given side nearer to $N/2$ as possible. This choice will produce a volatility $\sigma \sim N_d/4 = N/8$[5] which is roughly similar to what obtained in simulation Fig. (5) in low m low β region.

quency of opponent's play

[5]We are assuming $\Delta N = N_1(h_m) - N_0(h_m) < N_d(h_m)$. Notice that for random agents

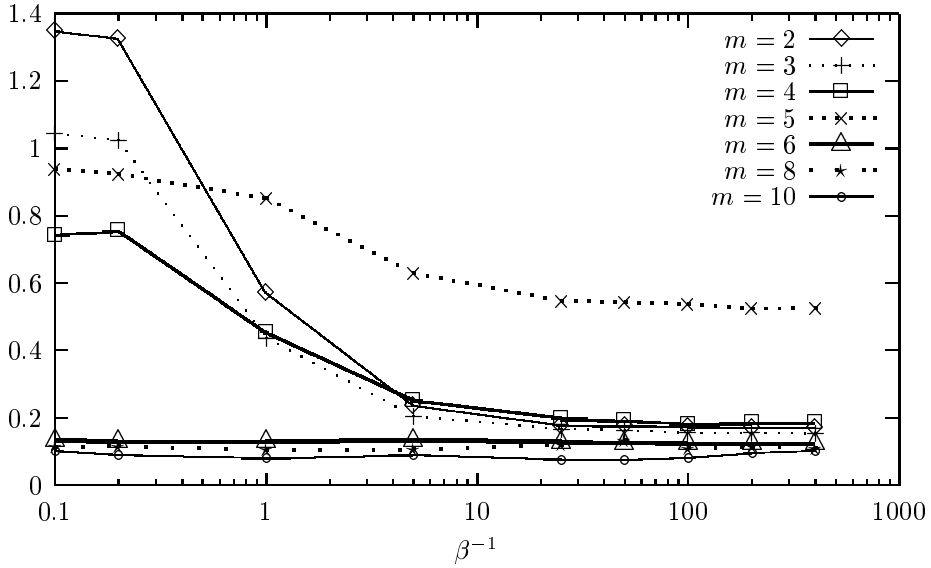

Figure 7: The variance of the distribution of σ over a sample of 50 independent runs. As β becomes small the point $m \sim m_0$ maintains a significantly larger variance.

A final remark concerns the variance of the distribution of σ as a function β for various m plotted in Fig. (7). The graph shows that when β decreases the variance of σ decreases for any m, however it remains three times greater for $m = m_o$ suggesting a stronger dependence of the asymptotic performance on the initial strategy assignment which the system is not able to rule out.

5 Informational efficiency

In this section we analyze the informational content of H, the binary string of successive winning sides. Relatedly, with informational efficiency we mean here the extent to which the future system outcome is unpredictable, i.e. the absence of any arbitrage opportunity.

Let $p(0|h_l)$ be the probability that a 0 follows a given string h_l of all the possible 2^l strings of length l.

The analysis performed in [3], for the original game leads to the identification of two regimes: a "partially efficient" regime for $m < m_o$ in which $p(0|h_l) = .5$, as long as $l \leq m$; thus no informational content is left for strategies with memory less or equal to the one used by the agents. For $m > m_o$ an "inefficient" regime is entered in which the distribution of $p(0|h_l)$ is not flat, even for $l \leq m$, meaning there are "good" strategies that could eventually exploit the market signal to

$\Delta N \sim \sqrt{(N)}$ and $N_d \sim N$ and we can neglect $\Delta N/N_d$ terms in the solution of (4)when N is large.

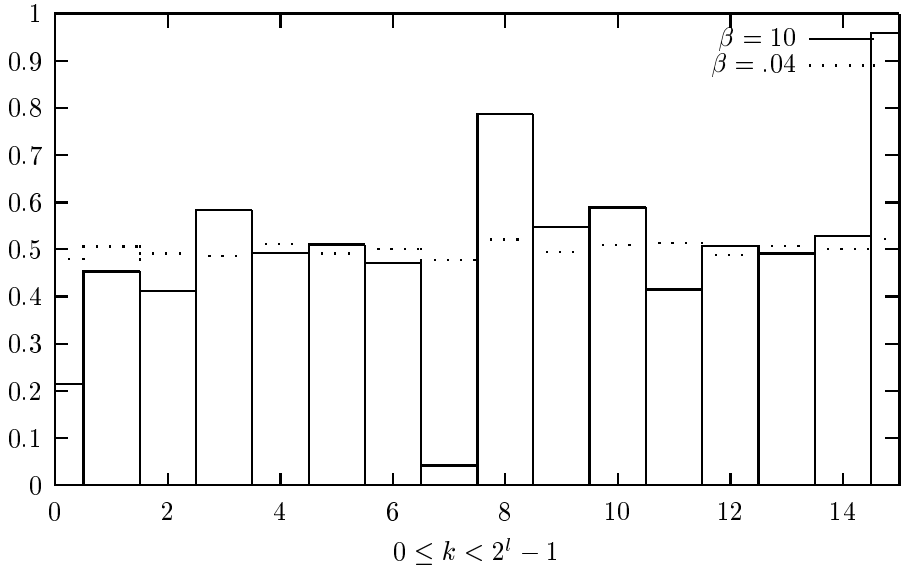

Figure 8: The probability $p(0|h_l)$ of obtaining 0 following a given binary string h_l in system history for $m = 3$ and $l = m + 1 = 4$. When β is reduced the distribution "flatten" and any structure is lost.

obtain higher profits. For $l > m$ both the regions show a non trivial distribution $p(0|h_l)$ with an increasing degree of "roughness" as l increases.

The effect of introducing "randomness" through the parameter β leads to the obvious effect of reducing the "roughness" of $p(0|h_l)$ (see Fig (8)).

In order to study the behavior of the system as β changes we introduce two related quantities which can be used to characterize the informational content of the time series. The first is the conditional entropy $H(l)$ defined as:

$$H(l) = -\sum_{h_l} p(h_l) \sum_{i \in \{0,1\}} p(i|h_l) \log p(i|h_l) \tag{5}$$

where the summation is intended over all the possible string of length l and $p(h_l)$ is the frequency of a given string in the system history H. The maximum value $H(l) = 1$ is reached for a flat distribution $p(0|h_l) = .5$, and is interpreted as impossibility of forecasting (in probability) the next outcome starting from the previous l outcomes. The idea that the information content can be used to "make money" leads us to the definition of a second quantity $A(l)$:

$$A(l) = \sum_{h_l} p(h_l) \max \{p(0|h_l), p(1|h_l)\} \tag{6}$$

which is the average fraction of point won by the best strategy of memory l. This is a measure of the reward obtained by the best arbitrageur with memory l (where if no arbitrage opportunities are present $A(l)$ is equal to .5.)

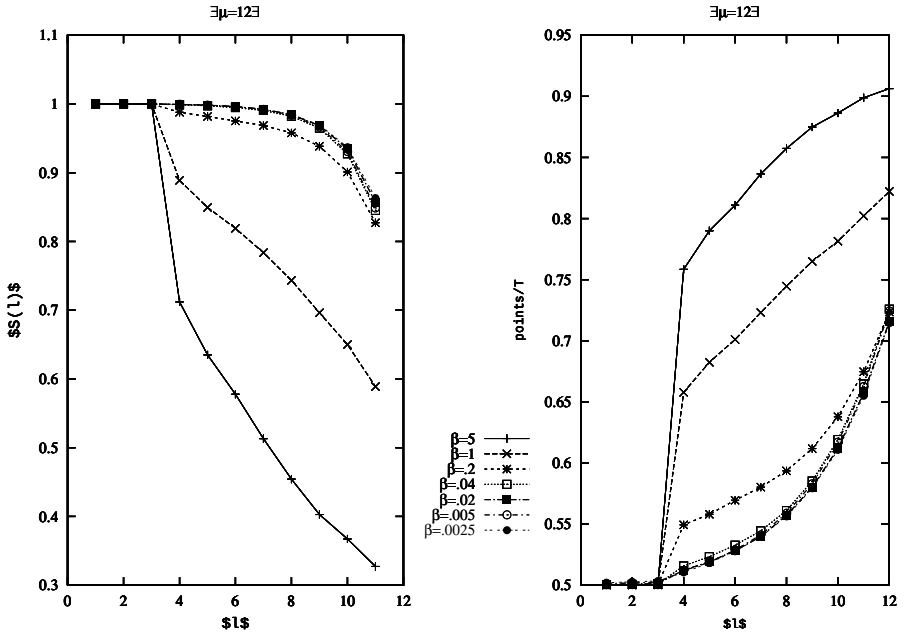

Figure 9: The conditional entropy $S(l)$ (left) and arbitrage opportunity $A(l)$ (right) as a function of time depth l for $m = 3 < m_o$.

Before analyzing the behavior of these quantities when β is varied, let us start by analyzing the properties of a population characterized by the two opposite models of "perfectly-informed, perfectly rational agents" and of "random agents" discussed before.

Under the former characterization the problem factorizes for each past history and the dependence on m disappears. The history produced by such a system is a random series of 0 and 1. Indeed the number of agents choosing one side is distributed according to a binomial around $N/2$ with different widths for different h_m. This in particular means that in this limit the "memory" loses any predicting power and no arbitrage opportunity is left for agents with longer memory, i.e. no residual information is left in the time series and the behavior of agents makes the market perfectly informationally efficient. Under this assumption we expect $S \sim 1$ and $A \sim .5$.

Under the opposite characterization of "random agents", due to the unbalance in the initial strategies endowment we expect a non trivial structure to appear for every l; thus $S < 1$ and $A > .5$.

In Fig. (9) we plot $S(l)$ and $A(l)$ for histories generated with a value of $m > m_0$, in the "partially efficient" regime. The effect of decreasing β shows up when $l > m$ but the information content for high l is never completely eliminated. The market becomes less efficient the larger is the time scale l at which it is observed. In fact it can be shown under very general assumptions that certain strings in the history are more abundant than others [3] and the long-range correlation that was responsible for the "crowd effect" at high β survives as a non trivial structure in $p(0|h_l)$ for high l. To an agent with memory $l \leq m$

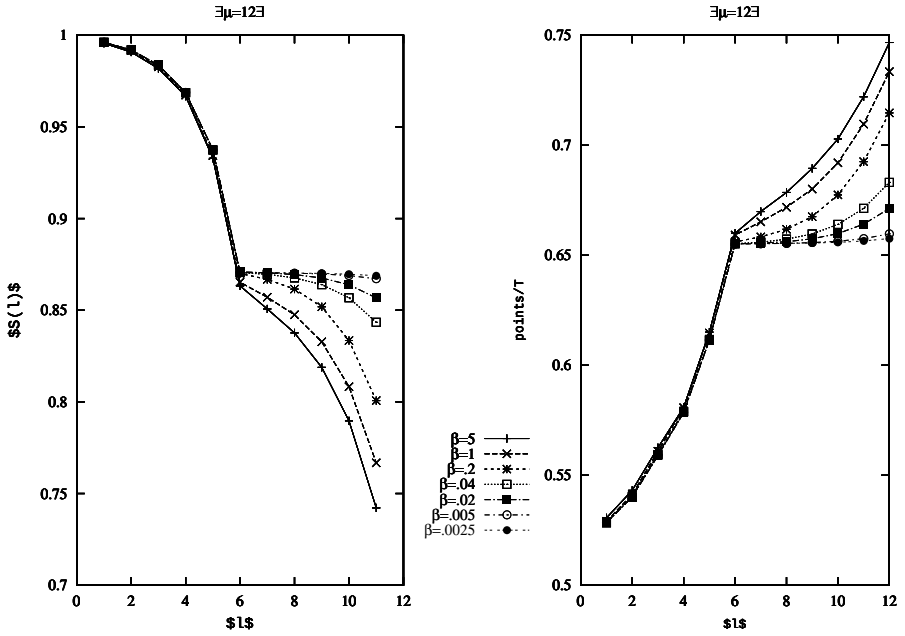

Figure 10: The conditional entropy $S(l)$ (left) and arbitrage opportunity $A(l)$ (right) as a function of time depth l for $m = 6 > m_o$.

the market appears perfectly efficient regardless of the β value.

For values of m in the "inefficient phase" the effect is in some sense reversed. As can be seen in Fig. (10) the effect of decreasing β is again negligible for $l \leq m$ but in the limit $\beta \to 0$ the curve becomes flat for $l > m$. This last result deserves some comments: the flatness in $l \geq m$ means that no gain is achieved from inspecting the time series with a very long memory $l >> m$ because no more arbitrage opportunities are open for a smarter (i.e. with a longer memory) agent than the best possible agent of memory m. The market can be called again "partially efficient" in the sense that it generates an upper bound on the maximal attainable arbitrage capability which does not depend on the arbitrageur memory.

The particular form of the conditional entropy in Fig. (10) suggests that in the limit $\beta \to 0$ the system can be described as a Markov chain of memory m. Notice that following its very definition, the system is conceived as one in which the past is not discounted (however, see Appendix A for an analysis of the system properties when a time discount factor is introduced), in the sense that agents weigh their strategies on the basis of all the game outcomes starting from the beginning of the simulation. The present result can be explained by noticing that when β is small only great differences in the past performances of strategies are relevant and in the limit $\beta \to 0$ only infinite differences stay relevant. Stated otherwise, the frequency of victories of the various strategies becomes constant implying the formation of a static hierarchical structure in the strategy space which at the end is responsible of the Markov character of the resulting history.

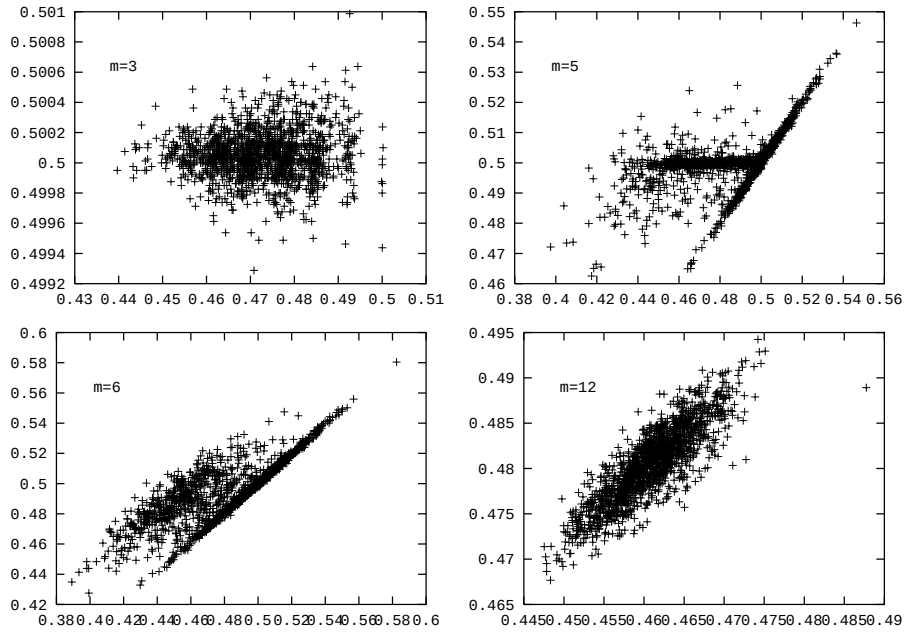

Figure 11: For each player we plot the scoring rate of its best strategy toward its own winning rate. The population is made of 30 independent runs of 101 players. The value of β is .04.

The appearance of "best strategies" in $m > m_o$ region is supported by plotting the average points scored by the best strategy versus the average point scored by the player (see Fig. (11)).

A correlation appears between the performance of a player and the performance of its best strategy for $m \geq m_o$. In the $m \sim m_o$ region a sub population showing the same kind of high correlation coexists with a population that presents no correlation, constituted of agents possessing two equally performing strategies.

We can say that the low m region is the one possessing the characteristics of "social optimality" where no strategies are preferred to others and no player is bound to lose due only to his initial strategy endowment.

Notice however that perfect equivalence between strategies does not necessarily imply equivalence in agent performances. As a further analysis we have plotted the variances and the supports of the points distribution for different value of *beta* and m in Fig. (12). It appears that only for low m and low β does equivalence in strategy performance imply a more uniform distribution of points over the population. We can then identity this region with the "socially optimal" one.

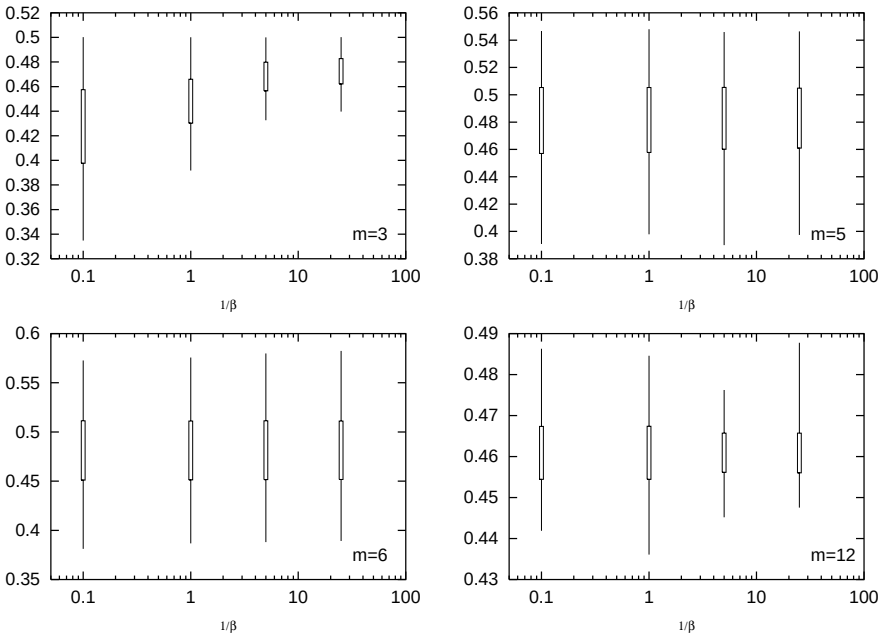

Figure 12: Variance (rectangle) and support (straight line) for the scored points distribution on a population of 30 independent runs with $N = 101$ and $s = 2$. Notice that while in the high β simulations the distributions are similar in width for any m, when β in reduced the low m region emerges are the "social optimal".

6 Conclusions and Outlook

Our results show that introducing some degree of randomness in the behavior of the low-rational agents who play the minority game has a positive effect on performances both in term of allocative and informational efficiency. The system indeed attains better resources exploitation and creates smaller, even if not negligible, arbitrage opportunities. Moreover the "social optimality" of the system, expressed as the inverse of the variance or analogously of the support, of the earnings distribution over the population increases with the "inertia" in the players behavior.

The major effect of randomness is that of acting like a brake on the system dynamics, thus preventing groups of players who densely populate the strategy space from acting in a strongly correlated way and from producing a "crowd" effect which worsens the system performance. The introduction of randomness in individual behavior is only one of possible ways to introduce heterogeneity in players' behavior. For instance, the same effect has been obtained in [15] substituting the "global" evaluation of strategies on the system history H with a "personal" evaluation in which each agent uses the binary string made up of its own record of victories. A "diversification" mechanism is again at work breaking the correlation among agents.

On the same line it is interesting to analyze the effect on the game of introducing a reinforcement learning model which, due to the "update only what

you play" prescription, will introduce a personal history for each player which presumably will unlock the crowd formation. The adoption of a reinforcement learning model, moreover, would be justified by it being the "zero-level" model in terms of degrees of rationality and information required, which renders it particularly well suited to model a wide array of real interactive situations. In fact, while more sophisticated models may be more easily violated by human players (and a growing literature indeed demonstrates that they often are), the "law of effect" underlying reinforcement models is almost never violated by human subjects. Results obtained by adopting this learning rule should therefore be considered quite robust. This analysis will be conducted in a forthcoming publication, together with the exploration of a "linear" model for the assignment of probability to strategies.

The reason to analyze the system aggregate properties under different "learning rules" is testing the "robustness" of the model: in fact, the characteristics of the system that are independent or weakly dependent on the particular behavior of the individual agents can be considered as general features of a multi-agent system like the minority game. In particular, our modification of the original model has been in the direction indicated by the experimental literature on learning in games [4]. From a more theoretical point of view such a study can be seen as an effort to decouple the peculiar features of a social self-organizing system from the exact rules governing the individual choices, in the spirit of trying to identify, at least in first approximation, the variables that determine its universality class.

7 Appendix A

Many authors especially in the experimental literature [4] introduce one more parameter in the description of learning, connected to the idea that agent weigh more the information they received in the recent past that the one coming from the far past. This parameter takes typically the form of a discount factor. If $\epsilon_i(t)$ are the points scored by strategy i at time t and $0 < \alpha \leq 1$ the information discount factor the updating rule for the total strength becomes

$$q_i(t+1) = \alpha q_i(t) + \epsilon_i(t) \tag{7}$$

and the associated updating rule for the probabilities:

$$p_i(t+1) = p_i^{\alpha}(t) \frac{e^{\beta q_i(t)}}{\sum_j p_j^{\alpha}(t) e^{\beta q_j(t)}} \,. \tag{8}$$

The effect of introducing such a memory leakage is twofold: on one hand it puts an upper limit to the maximal strength any strategy could reach, namely $1/(1-\alpha)$, and on the other hand in presence of no information flux the equiprobability between strategies is steadily restored. This effect will implies that if one takes the $\beta \to 0$ limit keeping constant the value of α, the system will converge to a collection of random agents. This can be interpreted saying that agents have to collect a large amount of information before they start behaving as an organized collection. The effect of introducing "forgetting" in the learning rule is easily understood: if the agents forget more rapidly than they learn they are always bounded to a suboptimal behavior. Indeed ,as can be seen from Fig. (13),

if the value of α is decreased the optimality of the system is proportionally reduced.

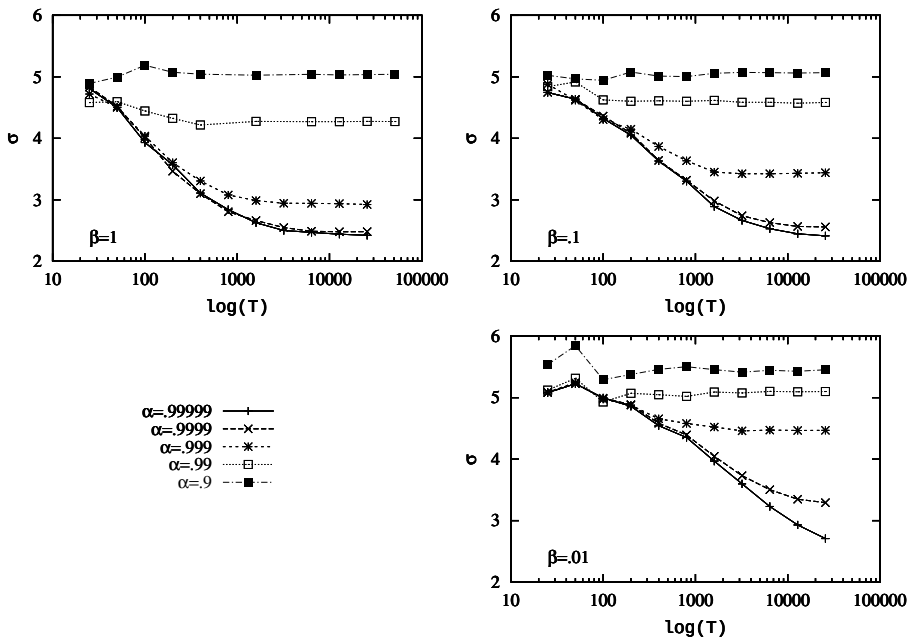

Figure 13: σ as a function of run length T for different values of β and α. The simulations are performed with $m = 6$ where a greater sensitivity of the transient time length toward "learning" parameter β and α is expected, see Sec. (3).

References

[1] D. Challet and Y.-C. Zhang "Emergence of cooperation and organization in an evolutionary game", Physica A 226, 407 (1997)
D. Challet and Y.-C. Zhang "On the minority Game: Analytical and Numerical Studies" Physica A 256,514 (1998)

[2] Rapoport, A., Seale, D., Erev, I., and Sundali, J., (1998), "Equilibrium Play in Large Group Market Entry Games", *Management Science*, vol. 44, No. 1, 119 – 141.

[3] R. Savit, R. Manuca and R Riolo "Adaptive Competition, Market Efficiency, Phase Transition and Spin Glasses" adap-org/9712006
R.Manuca, Yi Li, R.Riolo and R.Savit "The structure of Adaptive Competition in Minority Game" adap-org/9811005

[4] Erev, I., and Roth, A., (1997), "Modelling How People Play Games: Reinforcement learning in experimental games with unique, mixed-strategy equilibria", University of Pittsburgh Working Paper.

[5] Camerer, C., and Ho, T., (1999), "Experience Weighted Attraction Learning in Normal Form Games", Caltech Working Paper.

[6] Bush, R., and Mosteller, F., (1995), "Stochastic Models for Learning", New York, Wiley.

[7]

[8] J. W. Weibull "Evolutionary Game Theory", M.I.T. Press

[9] N.F.Johnson, M.Hart, P.M.Hui "Crowd effects and volatility in a competitive market" preprint cond-mat/9811227

[10] Thaler, R.H. (1993)(eds) "Advances in Behavioral Finance", Russel Sage Foundation, New York.

[11] De Bondt, W.F.M., and Thaler, R.H. (1995) "Financial Decision-Making in Markets and Firms: A Behavioral Perspective", in R. Jarrow et al., Eds., Handbooks in OR & MS, vol. 9, Elsevier Science.

[12] Fudenberg, D., and Levine, D.K. (1998), "The Theory of Learning in Games", MIT Press, Cambridge, MA.

[13] Osborne, M.J., and Rubinstein, A. (1994), "A Course in Game Theory", MIT Press, Cambridge, MA.

[14] A. Cavagna, J.P. Garrahan, I.Giardina and D. Sherrington "A thermal model for adaptive competition in a market" preprint cond-mat/9903415

[15] M.A.R. de Cara, O.Pla and F.Guinea "Learning, competition and cooperation in simple game" preprint cond-mat/9904187