Stochastic learning in co-ordination games: a simulation approach

Enrico Zaninotto, Alessandro Rossi and Loris Gaio

March 10, 1999

RAPPORTO INTERNO
DISA-015-99
MARZO 1999

DIPARTIMENTO DI INFORMATICA E STUDI AZIENDALI
Università degli Studi di Trento
Via Inama, 5
38100 Trento
http://www.cs.unitn.it

Abstract

In the presence of externalities, consumption behaviour depends on the solution of a co-ordination problem. In our paper we suggest a learning approach to the study of co-ordination in consumption contexts where agents adjust their choices on the basis of the reinforcement (payoff) they receive during the game.

The results of simulations allowed us to distinguish the roles of different aspects of learning in enabling co-ordination within a population of agents. Our main results highlight: 1. the role played by the speed of learning in determining failures of the co-ordination process; 2. the effect of forgetting past experiences on the speed of the co-ordination process; 3. the role of experimentation in bringing the process of co-ordination into an efficient equilibrium.

1 Introduction

Imagine a decision-maker, in a group of consumers, who must choose between two goods characterized by network externalities. In order to do so, he must evaluate similar choices made by other people in the group. He is then playing a co-ordination game and clearly his choice depends on the way in which he forms his expectations about other people behaviour.
This problem has been examined assuming both perfect rationality and some relaxed versions of rationality. Arthur's work on these matters is well known. Arthur [1989] supposes that choices are made sequentially and that each agent, randomly chosen from a population of agents with different preferences, selects the alternative that gives him the higher payoff, taking into account adoption choices observed until that moment. The payoff depends both on the agent's intrinsic preferences and on the market share gained by each alternative. Thus, the dynamic of the adoption process is driven by a random arrival mechanism which determines the evolution over time of market shares. In this way Arthur has highlighted some interesting features of sequential co-ordination processes, such as strong path dependence, unpredictability and non-ergodicity.
The model proposed by Kandori et al. [1993] is based on the same idea of myopic behaviour. But, in this case, each agent plays an iterative stag hunt co-ordination game (where one equilibrium is Pareto-dominant and the other is risk dominant) with a population of opponents: he chooses the strategy that maximizes his payoff, computed with respect to decisions observed in the previous period (best reply dynamic). In this case the process is deterministic and agents co-ordinate a strategy that depends uniquely on the initial strategies of the population. The same authors consider the effect of some evolutionary force (players which, at each stage, mutate from their best choice and play at random with a positive probability). The stochastic component of the process is thus based on the

small margin that evolution allows for chance. This small margin is enough to move the process from the best reply deterministic path: Kandori et al. find in fact that when evolutionary forces are at work, the system spends most of its time in the risk dominant equilibrium. Ellison [1993] enlarges the evolutionary framework of Kandori et al. to show that local externalities accelerate the equilibrium selection process towards the risk dominant equilibrium. Evolution, of course, is a source of innovation that shuffles the cards and tends to weaken the force of history. Öchssler [1997] has also worked along these lines, modelling evolutionary forces as a small probability, assigned to agents playing in a given group, of changing their group. He shows that, in this case, the process tends to the Pareto-efficient equilibrium.
Finally, another line of research has been pursued by Kaniovski and Young [1995], whose paper models a co-ordination process through fictitious play. Here the stochastic component of the process is introduced by sampling: at each stage, two players are randomly selected and, on the basis of information extracted from a sample of previous players, they choose their best expected strategy. Kaniovski and Young show that players almost invariably converge on a stable Nash equilibrium.
In short, the literature has highlighted three main ways to introduce a stochastic component into a myopic co-ordination process among N agents: random arrivals; random mutation; random information sampling. This paper seeks to highlight a different force, namely adaptive learning, which has, in some sense, an effect complementary to mutation or information sampling. Adaptive learning agents make use of their past experience (in terms of good and bad outcomes) to redirect their future behaviour. This effect is known as the *law of effect*.
In our model the stochastic component depends on the fact that, although players behave randomly, they adjust the probability they assign to their possible behaviour on the basis of experience. Hence, while the game-theoretic framework of our model closely resembles Kandori's, agents do not adopt best-response strategies; rather, they adjust their choices on the basis of the reinforcement (payoff) received during the game.
The main goal of this paper is to study, with the help of simulations, some general properties of learning in a co-ordination game. In other words, we investigate whether there is a "good way" to learn from experience. It is quite natural, in fact, to claim that a workable learning process in co-ordination games must:

- drive agents towards compatible choices;
- permit selection of the most efficient alternative, when there are different ways to achieve compatibility;

- be effective within a reasonable time span: very long learning times might not have a clear operational meaning.

It is not obvious how learning should affect these characteristics of co-ordination. On the one hand, when there is an opportunity to experiment with alternative outcomes, we would expect the whole population of agents to learn the best way to co-ordinate. On the other, we would expect a very fast learning process to trap agents in the strategies explored at the very beginning of the game, preventing them from searching for alternative (and maybe better) solutions. We could then expect some sort of trade off to arise between the learning time horizon and the efficiency of the learning process.
As we shall see, our simulations confirmed these expectations; although we shall also see that single parameters used to model different aspects of the learning process play special roles.
The learning algorithm that we employed derives from the studies of Roth and Erev [Roth and Erev, 1995, Erev and Roth, 1997]. The reasons for choosing this special algorithm are discussed in section 2, where we present some general features of adaptive models of behaviour and illustrate the Roth-Erev algorithm for stochastic learning. In section 3 we present the co-ordination model; section 4 is devoted to the presentation of the simulation results. Section 5 sketches some conclusions and outlines the progress of our research.

2 Stochastic learning algorithms

When adaptive behaviour takes place, individuals have capabilities which they effectively employ to discriminate among environmental stimuli (incentive structure, state variables, behaviour of other agents, and so on), and they modify their behaviour as a consequence of their elaboration of these environmental stimuli.
In the broad class of adaptive models of behaviour we can, nevertheless, distinguish between two different categories: reinforcement learning algorithms and beliefs-based algorithms. [1]

[1] IIt is surprising to find that the majority of studies on adaptive behaviour in games have focused on only one or the other class of learning models, without attempting to integrate the two approaches. Clearly, each class of adaptive model only takes account of one side of the adaptive behaviour of real decision-makers: while reinforcement learning neglects the role of beliefs in influencing behaviour, belief-based models do not consider the effect of past earned payoffs on future behaviour. Only recently have some scholars proposed an integration of the two models in order to overcome these shortcomings. For instance, both Camerer and Ho [1996] and Erev and Roth [1997] suggest an integrated approach to the modelling of adaptive learning, but they do not agree on how different learning algorithms perform in tracking experimental data. In particular, Roth and Erev point out that reinforcement learning models outperform belief-based ones and that

For the purpose of our research we decided to employ an adaptive model related to pure reinforcement learning; this class of adaptive learning processes, which also goes under various other labels, such as *choice reinforcement* or *stochastic learning*, is characterized by the following specific features:

- *stochastic behaviour*: strategies are selected by a stochastic mechanism (which associates a measure of propensity with each strategy);
- *reinforcement*: strategies are reinforced (inhibited) by previous positive (negative) payoffs;
- *payoff driven*: the only relevant feedback in updating propensities is the individual payoff;
- *behavioural focus*: no attention is paid to the decision-maker's beliefs or other internal mental states.

In particular, we used an algorithm derived from the studies of Roth and Erev [Roth and Erev, 1995, Erev and Roth, 1997]. Although the baseline version of their model is extremely simple, it is clearly psychological grounded in that it embeds the following fundamental properties of human learning:

Law of Effect: actions that have led to good outcomes (positive payoffs) in the past are more likely to be repeated in the future.[2]

Power Law of Practise learning curves tend to be steep initially and then to become increasingly flatter over time. [3]

While the reader is referred to the works of these authors for their general models [Roth and Erev, 1995, Erev and Roth, 1997], in what follows we shall focus on a particular instance of the adaptive model, which was adjusted to the purposes of the game that we studied.
In each period $t = 1, 2, \ldots$, player i chooses one of two possible actions, $d_{i,t} \in \{A, B\}$, with probabilities, respectively, of $p_{i,t}$ and $(1 - p_{i,t})$ and at the end of each stage the player receives a payment of $u_i(d_{i,t}, \cdot)$.
During each period t each player i independently and simultaneously chooses action A with probability $p_{i,t}$ and B with probability $(1 - p_{i,t})$.

the additional contribution of a mixed model to the explanation of experimental data is probably not worth, given the increased complexity of the algorithm, while Camerer and Ho show how their unified model (EWA) performs better than both belief-based and reinforcement based models.

[2]The law was originally formulated by Thorndike [1898].

[3]Like the previous law, this empirical claim too dates back to the early psychological literature on human and animal learning, and at least to Blackburn [1936].

In order to compute the probabilities $p_{i,t}$, the following measure of propensity

$$q_{i,t}(X), \qquad \forall X \in \{A, B\}$$

is defined for each of the two available strategies. Initial propensities (at period $t = 1$) are given and may assume any positive real number. The probability $p_{i,t}$ (of selecting strategy A) is then defined as the ratio between the propensity related to strategy A and the sum of the propensities related to the (two) available actions, as follows:

$$p_{i,t} = \frac{q_{i,t}(A)}{q_{i,t}(A) + q_{i,t}(B)} \tag{1}$$

The distinctive feature of the model is that reinforcement acts at the level of propensities: in period $t + 1, \forall t > 0$, each player i updates his propensity $q_{i,t+1}(d_{i,t})$ on the basis of the payoff earned in the previous period t as a result of having chosen strategy $d_{i,t}$, as follows:

$$q_{i,t+1}(d_{i,t}) = q_{i,t}(d_{i,t}) + u_i(d_{i,t}, \cdot) \tag{2}$$

Finally these updated propensities are used to compute the new probability values $p_{i,t+1}$ for the period $t + 1$.
Roth and Erev have suggested a more general version of the algorithm, allowing for two other well known robust features of human learning pointed out by the psychological literature [Skinner, 1953, Watson, 1930]:

Local Experimentation: (also know as Generalization or Error) positive past payoffs experienced with one strategy reinforce not only the strategy selected but also similar choices;

Gradual Forgetting: (or Recency) past experience is gradually forgotten and a more salient role is played by recent experience.

These two features can be easily incorporated into the baseline model, assuming that when player i in period t chooses $d_{i,t}$, then he updates both his propensities in the following way:

$$q_{i,t+1}(d_{i,t}) = (1 - \varphi)q_{i,t}(d_{i,t}) + (1 - \varepsilon)u_i(d_{i,t}, \cdot), \tag{3}$$

$$q_{i,t+1}(\neg d_{i,t}) = (1 - \varphi)q_{i,t}(\neg d_{i,t}) + \varepsilon u_i(d_{i,t}, \cdot), \tag{4}$$

where $\neg d_{i,t}$ is the strategy not chosen by player i in period t, ε is the generalization parameter, which prevents the propensity of the strategy not chosen from going to zero and φ is the *forgetting* parameter which set an upper bound on the value that a propensity can take. Obviously, when $\varphi = \varepsilon = 0$ we have the baseline model outlined in eq. 2.

3 The co-ordination model

Image a group of N players. Each player repeatedly plays a 2×2 co-ordination game (with one mixed and two pure equilibria) with the remaining $(N-1)$ players. One of the two pure strategy equilibria is Pareto-dominant. Payoffs are normalized to one, so that each two-person game has the strategic form shown in Figure 1.

	A	B
A	$1, 1$	$0, 0$
B	$0, 0$	$0.5, 0.5$

Figure 1: A baseline two-person co-ordination game.

In each period $t = 1, 2, \ldots$, player i chooses one of the two possible actions, $d_{i,t} \in \{A, B\}$, and at the end of each stage he receives as payment a compounded sum of the payoffs earned in each of the $(N-1)$ two-person game. This can be represented using the following payoff function:

$$u_i(d_{i,t}, d_{-i,t}) = \sum_{j \neq i} \pi_{i,j} g(d_{i,t}, d_{j,t}) \tag{5}$$

where the payoffs g are those of the previously highlighted 2×2 co-ordination game, and where $\pi_{i,j}$ is termed the *matching rule* and can be interpreted as the probability that players i and j will be matched in a given period of the iterated game. One can imagine (following Ellison [1993]) many different specifications of the matching rule: in what follows we adopt the simplest one, i.e. the so-called *uniform matching rule*:

$$\pi_{i,j} = \frac{1}{N-1} \qquad \forall j \neq i. \tag{6}$$

At the first stage, players play randomly, extracting their strategy from a given distribution (from now on we adopt the extreme hypothesis that players have no prior information, so that the strategy can be thought of as extracted from a rectangular distribution). Each player then observes the payoff of the stage game and correspondingly adjusts probabilities. Since the basic game is a co-ordination game, the greater the number of agents selecting the same strategy, the higher the reinforcement that strategy will receive; moreover, since the co-ordination game has a Pareto-dominant equilibrium, the Pareto equilibrium strategy (A) will receive stronger reinforcement than the B strategy, assuming that they are equally chosen by the population.
The state of the system at t is the sequence of choices each agent has taken to that moment. For instance, at time 3, we should have a set of choices like: $((A, B, A)_1, (A, B, B)_2, ..., (B, A, A)_N)$. The joint sequence of choices

of each agent at time t uniquely determines his probability $p_{i,t+1}$ to undertake strategy A at time $t+1$.

4 Results

4.1 The simulation plan

We run extensive computer simulations for the model previously described with a population of $N = 300$ agents.
In each simulation run, the stage game was iterated either for 10000 periods (which was considered sufficient time to observe convergence to the steady state) or until the following stopping rule was satisfied:

$$\sum_{i=1}^{n} |p_{i,t-1} - p_{i,t}| < 10^{-5}.$$

We chose to perform sensitivity analysis on two dimensions:

- learning rate, which depends on the initial strenght of propensities, defined as $S_{i,1} = q_{i,1}(A) + q_{i,1}(B)$. The higher $S_{i,1}$ is set, the lower the learning rate; following Erev and Roth [1997], we decided to set $S_{i,1}$ equal to 3 times the agent average payoff in the baseline version (this will be noted as $S_{i,1}(3)$). We then investigated two alternative settings, one with a higher level of learning rate ($S_{i,1}(0.3) = 0.3\times$ the average payoff) and the other one with a lower level of learning ($S_{i,1}(30) = 30\times$ the average payoff);
- initial probabilities $q_{i,1}$ (given at the beginning of the simulation run). These values affect the probabilities that strategies A and B will be played at the beginning. We decided to set $q_{i,1} = .\bar{3}$ in the baseline version, since at this point, if the population is large enough, around one third of the agents will choose strategy A, and this will result in similar payoffs for the whole population. We then moved from this baseline case to explore higher ($q_{i,1} = .3\bar{6}, q_{i,1} = .4$) or lower initial probabilities ($q_{i,1} = .2\bar{6}, q_{i,1} = .3$).

As a result, the whole sensitivity plan consisted in a 3×5 factorial design as depicted in Figure 4.1, where each cell gives the corresponding values of initial propensities ($q_{i,1}(A)$, $q_{i,1}(B)$).
With respect to the Roth-Erev algorithm (equations 3-4), we took four different parameterisations:

plain model: where neither experimentation nor forgetting are introduced ($\varphi = 0$ and $\varepsilon = 0$);

experimentation model: where $\varphi = 0$ and $\varepsilon = 0.05$;

	$p_{i,1} = .2\bar{6}$	$p_{i,1} = .3$	$p_{i,1} = .\bar{3}$	$p_{i,1} = .3\bar{6}$	$p_{i,1} = .4$
$S_{i,1}(.3)$	$.03, .0825$	$.03375, .07875$	$.0375, .075$	$.04125, .07125$	$.045, .0675$
$S_{i,1}(3)$	$.3, .825$	$.3375, .7875$	$.375, .75$	$.4125, .7125$	$.45, .675$
$S_{i,1}(30)$	$3., 8.25$	$3.375, 7.875$	$3.75, 7.5$	$4.125, 7.125$	$4.5, 6.75$

Figure 2: Sets of initial propensities according to initial parameters

forgetting model: where $\varphi = 0.01$ and $\varepsilon = 0$;

experimentation and forgetting model: where both forgetting and experimentation are allowed ($\varphi = 0.01$ and $\varepsilon = 0.05$).

Thus, we run simulations over a total set of 60 different parameterisations. [4] For each parameterisation we run 500 simulation trials, and at the end of each trial we recorded the shares a_{10}, a_{100}, a_{1000}, and a_x; with $a_t = N_{A,t}/N$ and where N is the size of the whole population, $N_{A,t}$ is the number of A adopters at epoch t and x is the last epoch (10000 or less if the stopping rule is satisfied).

4.2 Learning co-ordination in the long run

Before analysing the results of our simulations, it may be helpful to recall some results from deterministic best-reply dynamics as a benchmark. As pointed out by recent studies in evolutionary game theory [Kandori et al., 1993, Ellison, 1993], the properties of deterministic best-reply learning algorithms in 2×2 games with two symmetric strict Nash equilibria and one mixed strategy equilibrium, as in the case of the co-ordination game studied here, are already well understood. We may imagine a population of agents playing a best reply strategy as follows: in period t agent i is randomly selected, observes the behaviour of his $N-1$ opponents in $t-1$, and then includes his own behaviour to maximise his expected payoff. This best reply dynamic has two steady states that reached in finite time (or three if $N/3$ is an integer): $d_i = A, \forall i = \{1, \ldots, N\}$, $d_i = B, \forall i = \{1, \ldots, N\}$ (and, if $N/3$ is an integer, also $N/3$ agents playing $d_i = A$ and $2N/3$ agents playing $d_i = B$). Clearly, the final outcome is path dependent, since it crucially depends on the initial condition of the system: if more than $N/3$ agents initially choose $d_i = A$ then $d_i = A$ will be the steady state, otherwise $d_i = B$ will be the final state. We may thus interpret the dynamic process as having stable attractors $a = 0$ and $a = 1$, whose basins of attraction have a boundary for $a = 1/3$. It is then easy to show that this process converges with probability 1 to one of the two attractors in finite time.

[4] Three magnitudes of initial propensities $S_{i,1}$, five initial probabilities $p_{i,1}$, two forgetting (φ) conditions and two experimentation (ε) conditions.

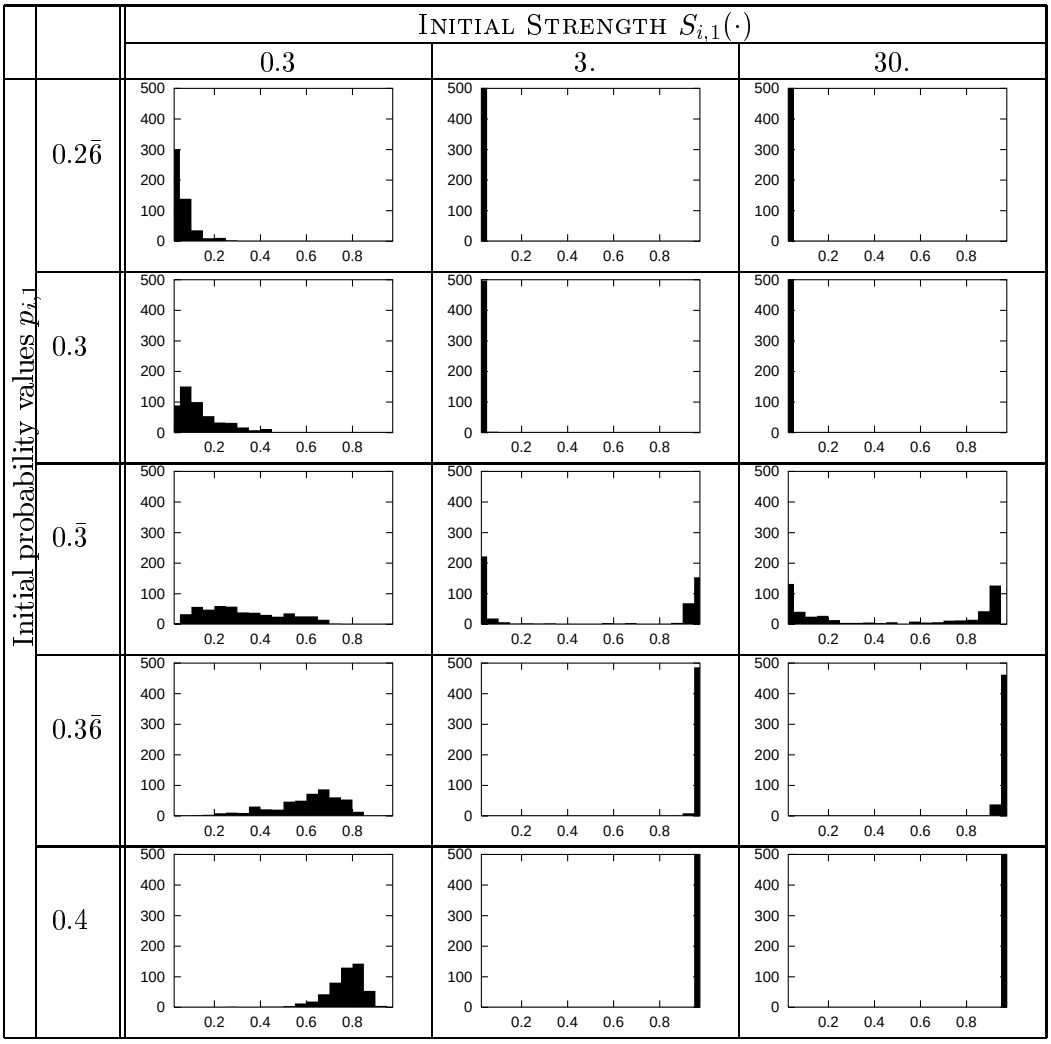

Figure 3: Shares for A Technology – Plain Model ($\varphi = 0, \varepsilon = 0$)

Commenting on our main results requires introduction of Figures 3 to 6, each presenting, for one of the four learning models (plain model, experimentation m., forgetting m., and experimentation with forgetting m.), the shares a_x computed for all the 500 simulation runs. The rows show the different initial probability conditions ($p_{i,1}$) and the columns show the different levels of learning rate (in term of $S_{i,1}(\cdot)$, the sum of initial propensities $q_{i,1}(A) + q_{i,1}(B)$).
Figure 3 sets out the results obtained with the plain learning model, where both gradual forgetting and local experimentation were not allowed ($\varepsilon = 0, \varphi = 0$).
The central row presents the results obtained when the starting point was a 1/3 probability of choosing technology A. Around this point, we would

expect the process to have equal probabilities of being driven towards one or other equilibrium, according to "small events" of the process. On the other hand, we would expect, with initial probabilities lower than 1/3, the process to be driven towards a null share of A, and with initial probability to chose A higher than 1/3, the population to co-ordinate, in the long run, on the use of technology A.
It is evident that our expectations concerning the convergence of the learning process are fulfilled only when learning rates are not too high. The second and third columns show, in fact, that starting from an initial condition of 1/3, the process has equal probabilities of being driven towards one or other technology; on the other hand, when initial probabilities are higher than 1/3, the full population co-ordinates on technology A; finally, when initial probabilities are lower than 1/3, the full population co-ordinates on technology B. We thus observe the usual path dependent phenomena and the "small events" effect around the mixed equilibrium point.
Matters are different when learning rates are very high. When initial probability values are around 1/3, in the long run the population splits in two sub-populations which co-ordinate on different technologies. The way in which the population splits is very different: the shares of choice A are between 0.1 and 0.7. Also when the starting point is a probability of A different from 1/3, there is not convergence on a single technology: instead the modal share of adopters moves towards 1 when initial probabilities are higher than 1/3, or towards 0 when initial probabilities are lower than 1/3, but full co-ordination is never reached. First moves reinforce initial choices, impeding search in subsequent trials from being rewarded by a good outcome. The tendency of the population to move toward a full co-ordination strategy, when initial values are distant from 1/3, is thus halted by a tendency for the system to "freeze in its path" as a consequence of the fact that a large number of agents stick to previous choices.
The situation changes when experimentation is introduced (Figure 4). In this case it is possible to observe that:

- the population more frequently co-ordinates on a single choice, even when learning rates are high. Experimentation appears to impede agents from being frozen early on in strategies played in the first steps of the game;
- the Pareto-dominant strategy is selected even when initial probabilities are near the mixed equilibrium. To illustrate how experimentation modifies the dynamic of co-ordination, imagine that one third of the population chooses technology A and two third chooses B: in this case players have equal payoffs and the two strategies are equally reinforced, leaving things unchanged. But, with

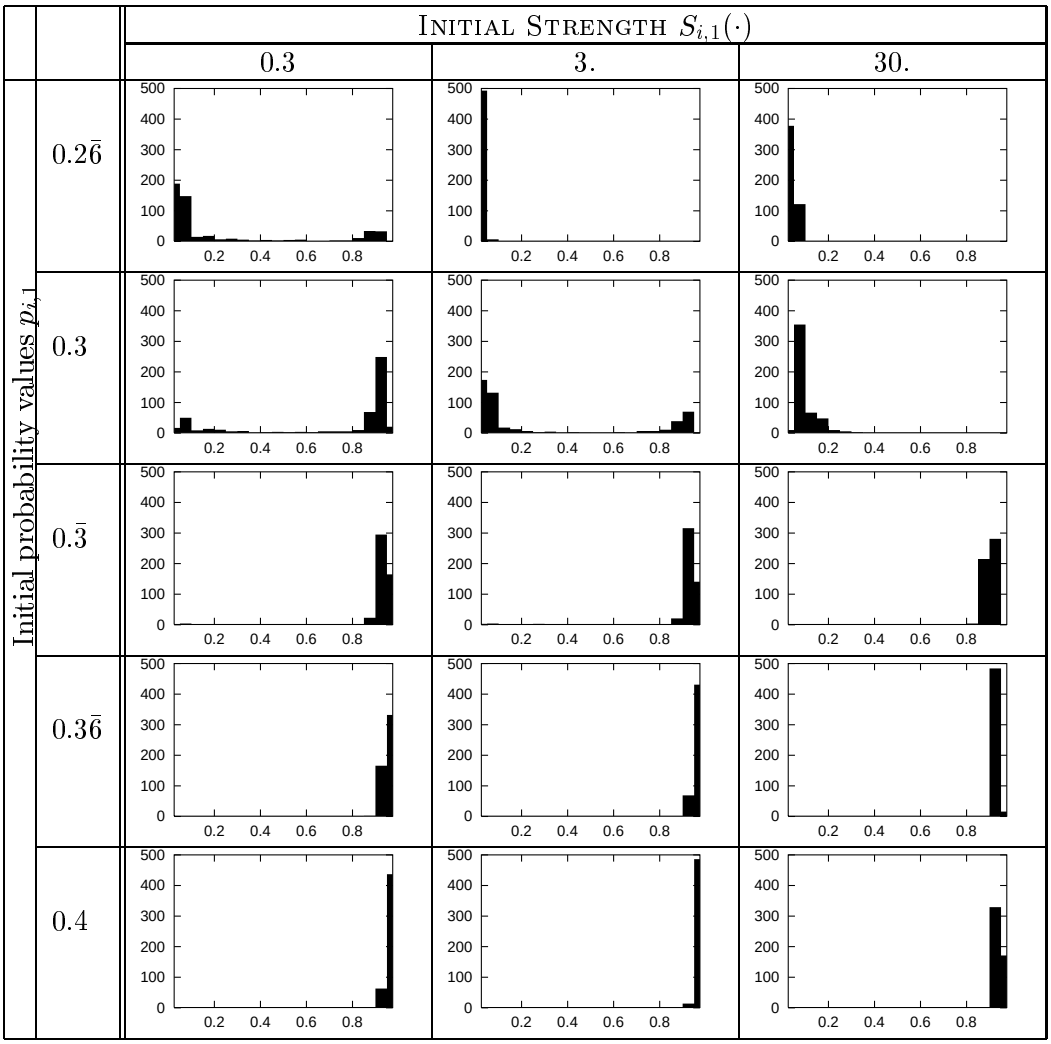

Figure 4: Shares for A Technology – Experimentation Model ($\varphi = 0, \varepsilon = .05$)

experimentation, the two thirds of players that have played strategy B give a small reinforcement to strategy A. The reverse is also true, but in this case only one third of players reinforce B. In general, when there is a Pareto dominant strategy and experimentation, the share of players that, near the indifference point, reinforce the Pareto dominant strategy is necessarily higher than the share of players that reinforce the alternative strategy;

- ordination is often not complete, due to the experimentation mechanism by which a small probability of choosing the alternative strategy is assigned even when the large majority co-ordinate on one choice;

Forgetting does not change the situation (Figure 5) with respect to the basic case, when learning rates are middle or low. Some new effects appear when learning rates are high. In this case the distribution of shares tend to be bi-modal. The role of forgetting is to reduce, after some time, the weight of old preferences: so that the population tends to be attracted by extreme solutions; nevertheless the high level of learning rate still traps the population before it achieves full co-ordination.
Finally, Figure 6 presents the case in which both experimentation and forgetting are at work. The results in this case are not particularly different from what was observed in Figure 4, when experimentation only was allowed. Thus, with respect to the final outcome of the simulations, the effect of experimentation seems to overcome the role of forgetting.

4.3 The timing issue: how long is the long run?

In the previous subsection we analysed the long run results of the simulations. The stopping rule we devised when running the simulations reflected the focus of our analysis, which was to investigate whether or not in the long run people co-ordinate on some equilibrium, and to what extent this equilibrium is unique, or alternatively whether co-ordination failure may occur.
The results presented in the previous subsection discriminate between conditions which result in full co-ordination rather than in partial co-ordination (co-ordination failure), and the conditions that result in co-ordination on a specific equilibrium.
However, in order to conduct adequate comparison among the four learning models, and to complete our analysis, we must also examine the time spent by the population of agents in achieving the steady state.
To illustrate our findings we introduce Figures 7, 9, 10 and 11, each of which presents, for one of the four learning parameterisations, the average concentration index $C_t = (\frac{N_{A,t}}{N})^2 + (1 - \frac{N_{A,t}}{N})^2$ computed over 500 simulation trials at periods $t = 10, 100, 1000, \bar{x}$ (where $N_{A,t}$ is the number

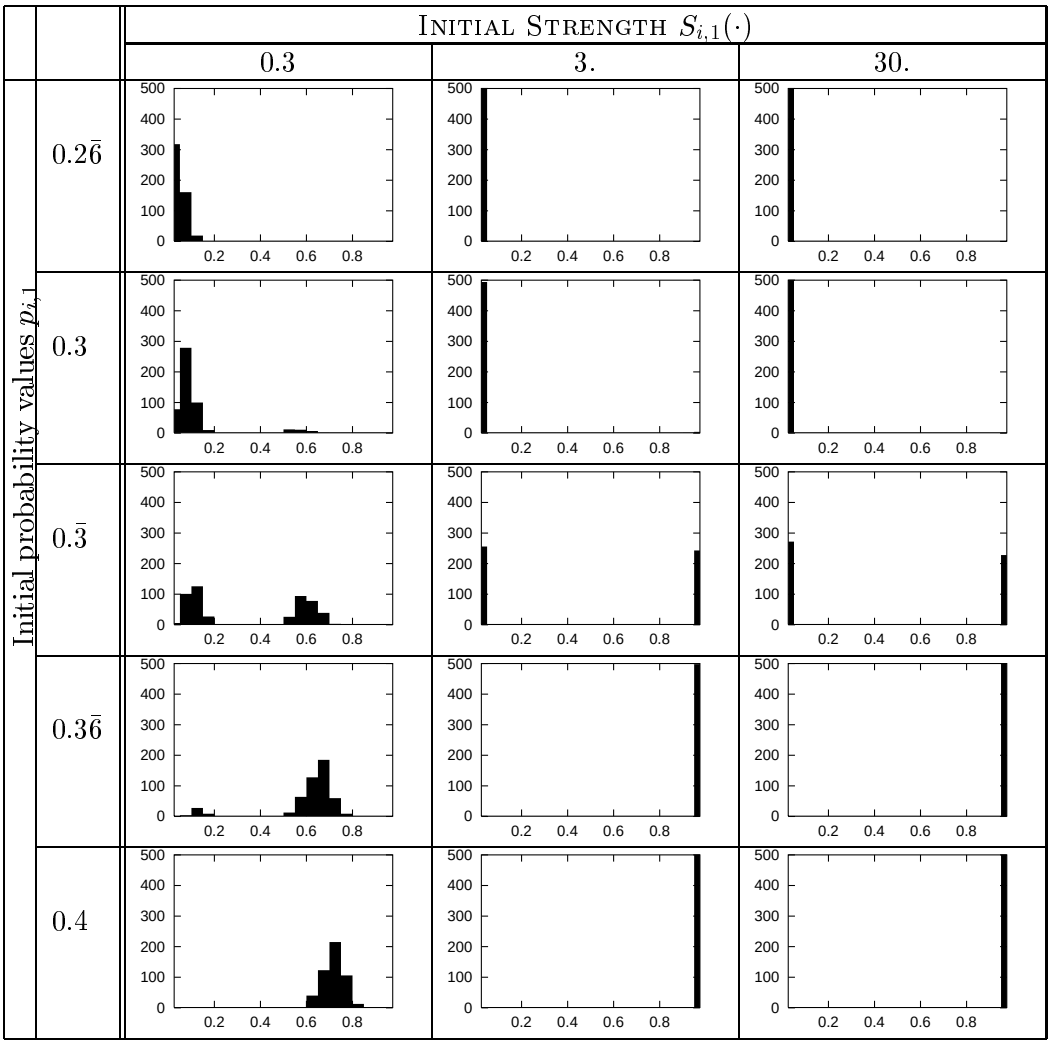

Figure 5: Shares for A Technology – Forgetting Model ($\varphi = .01, \varepsilon = 0$)

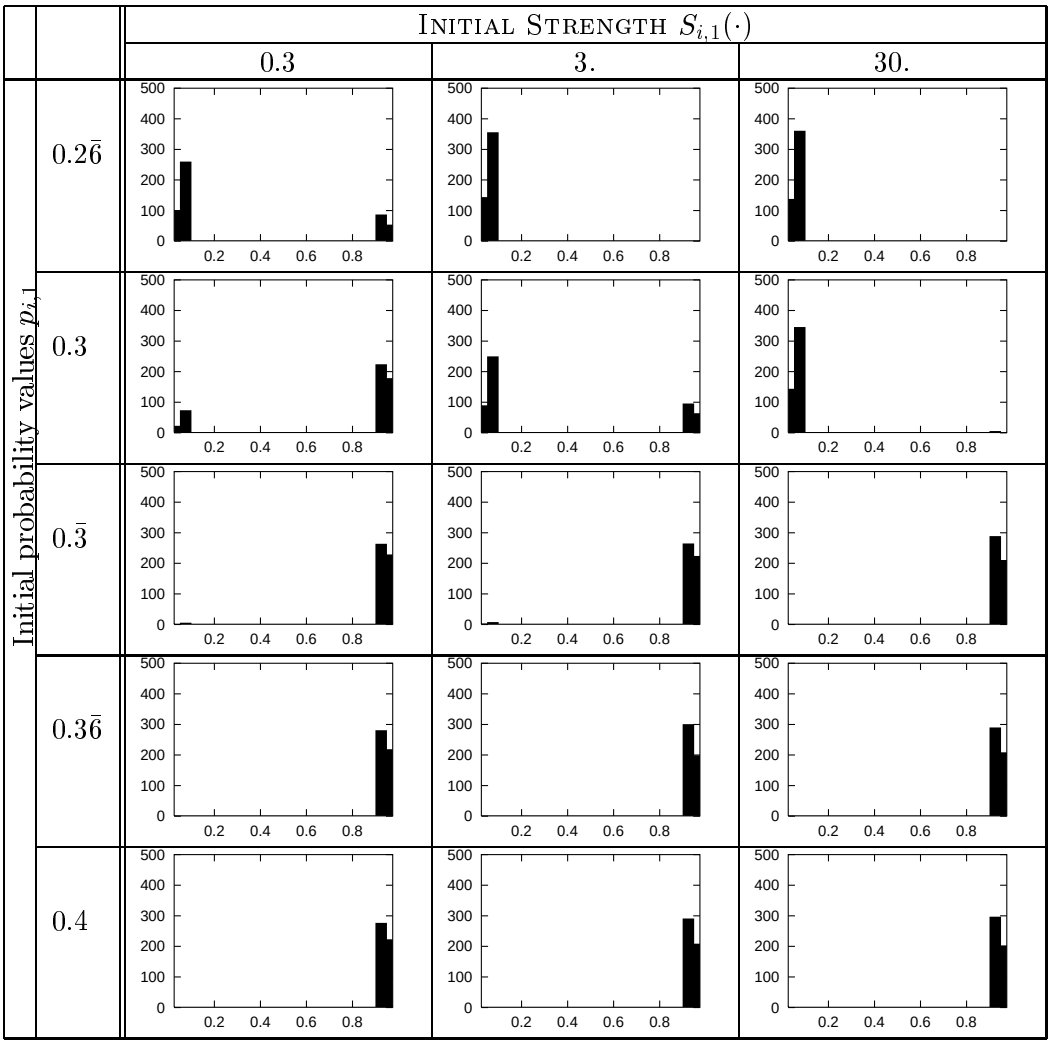

Figure 6: Shares for A Technology – Experimentation and Forgetting Model ($\varphi = .01, \varepsilon = .05$)

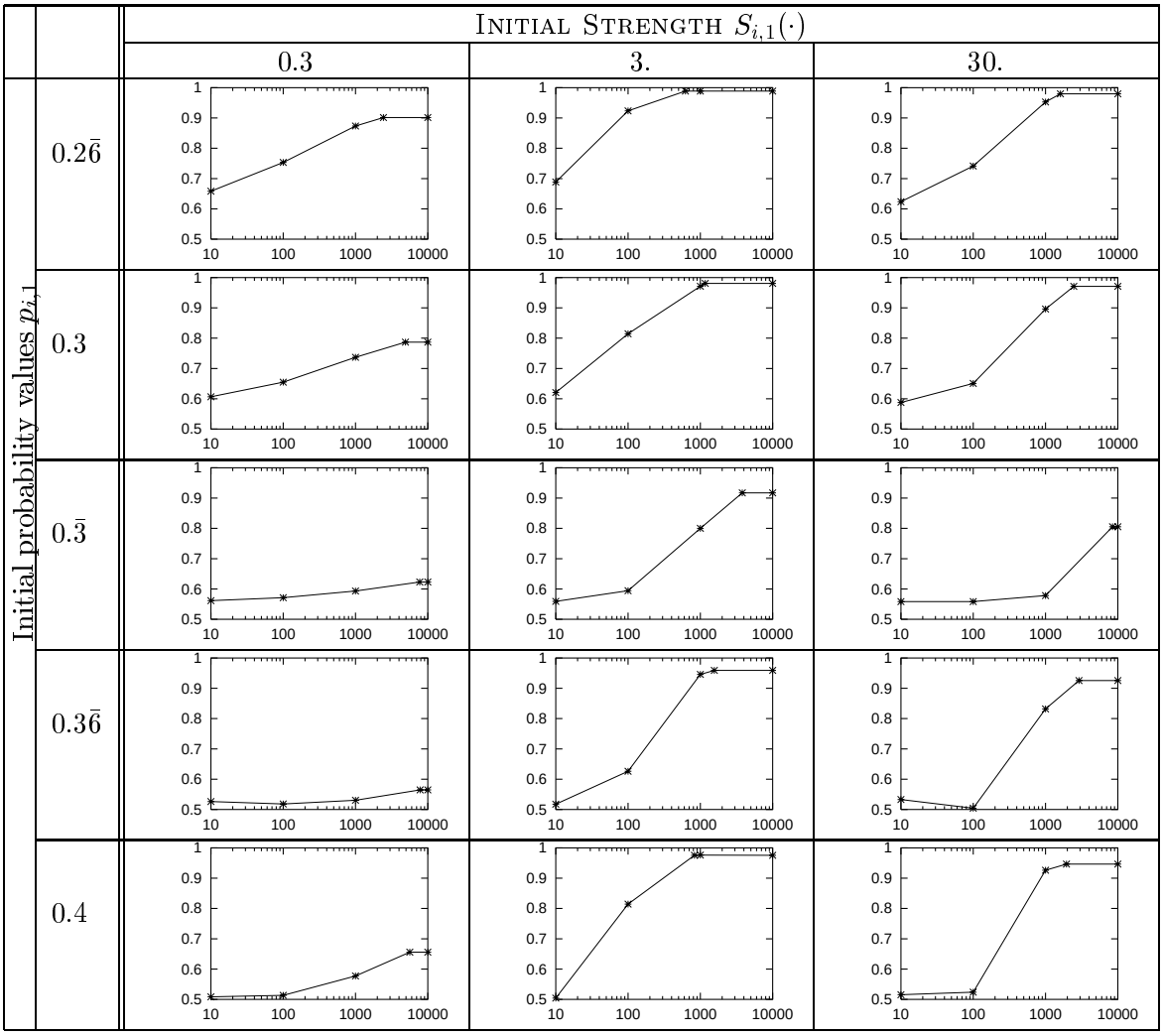

Figure 7: Concentration Index – Plain Model ($\varphi = 0, \varepsilon = 0$)

of adopters of technology A at period t and $\bar{x}$ is the average last period according to the stopping rule). As in the previous figures, here the rows show different initial probability conditions, and the columns show different levels of learning rate.

We begin with analysis of convergence times for the plain case. In this case we would intuitively expect the average convergence time to be negatively correlated with learning rate (the higher the learning rate, the lower the average convergence time). This is coherent with the idea that higher learning rates trap the system more quickly, while lower learning rates result in a very slow process of probability updating. Moreover, it should be easy to predict that a non-monotonical relationship will link convergence times and initial probability values (the closer to the

indifference point $p_{i,1} = 1/3$ at the beginning, the higher the convergence time); which reflects the intuitive insight that less time is needed to achieve full co-ordination when the population of agents is distant from the indifference value of the initial probabilities ($p_{i,1} = 1/3$), since the choice of one strategy is better rewarded than the opposite one. However, only the latter hypothesis was confirmed by simulations (see Figure 7), while a non-monotonic relationship between learning rates and convergence times was found. Indeed, convergence times both in the high learning rate and low learning rate treatments were longer than in the medium one. This non-intuitive result (longer convergence times for high levels of learning rate) may be explained as follows (for the sake of simplicity we shall restrict our analysis to the case of $p_{i,1} = 1/3$). In the medium learning rate treatment players gradually update their propensities and early small events drive the whole population smoothly into a self reinforcing lock-in process towards one of the two pure equilibria (the same dynamics can be observed, although at a slower pace, in the low learning rate treatment). Conversely, in the high learning treatment we can distinguish between early and late adopters, viz. players converging rapidly versus players converging relatively slowly to one single decision (technology A or B). In fact, owing to the high level of learning rate, many players lock into their behaviour in the very early periods of the simulation. Since probabilities in early periods are updated strongly, if one agent repeatedly selected the same decision at the beginning, he would so strongly reinforce his behaviour that the probability of extracting the opposite option would rapidly be driven close to zero. By contrast, a relatively small fraction of players who do not get stuck at the beginning of the simulation starts playing a co-ordination game where the large part of the population has already locked its behaviour into technology A or in technology B. The emergence of early and late adopters in the high learning rate treatment can be observed in Figure 8, which collects the number of players that change their decision over time, contrasting a typical high learning rate with medium and low learning rate simulation runs. Thus we observe that the closer the share of early adopters of technology A to 1/3, the higher the convergence time, since a late adopter receives similar reinforcement if he chooses A or B, and the magnitude of the reinforcement declines over time. The relatively low pace of convergence of the simulation towards a steady state in the high learning condition is thus the joint result of the separation of the population of agents between early and late adopters and the decline, over time, of learning rates for the late adopters.
Convergence times in the second treatment (experimentation) are slightly different in comparison with the baseline condition (Figure 9), since the non-monotonical relationship between convergence times and initial probability values is not found. This is essentially due to *exploration*, which drives the population towards technology A, since the more distant

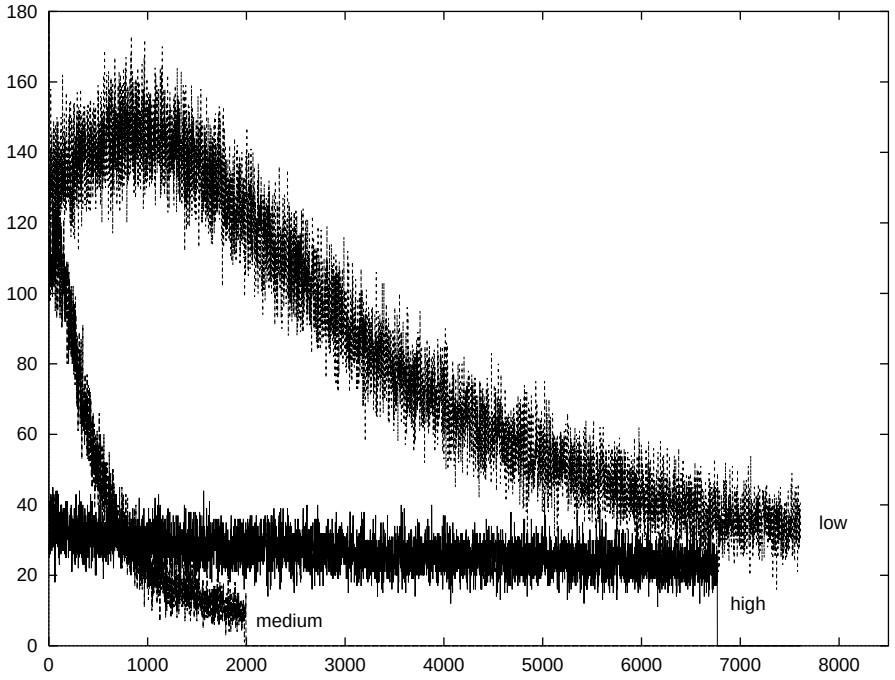

Figure 8: Number of players shifting from A to B or viceversa over time in three typical simulation runs (high- medium- and low-learning rate, plain model, $p_{i,1} = 1/3$).

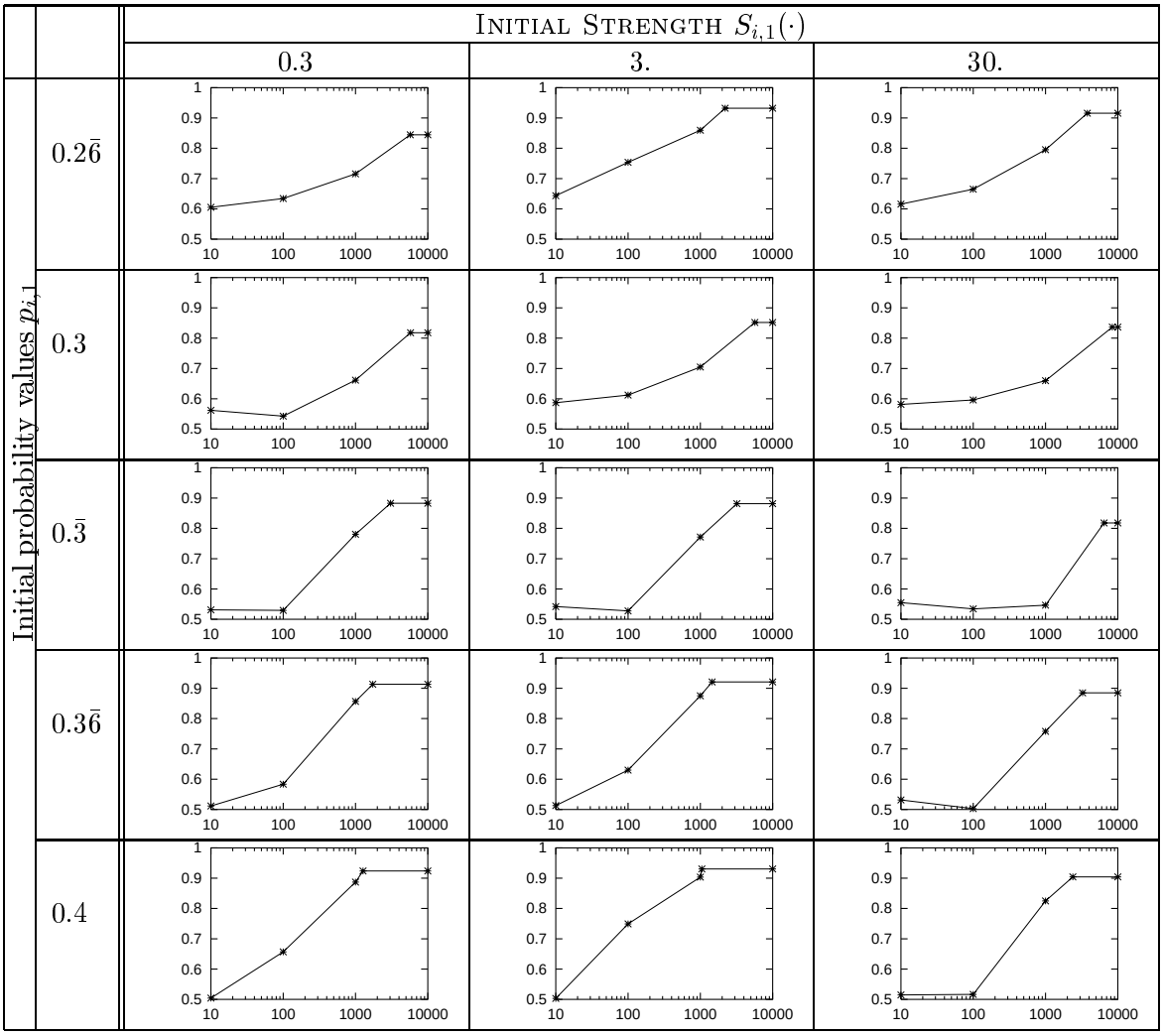

Figure 9: Concentration Index – Experimentation Model ($\varphi = 0, \varepsilon = .05$)

are agents' initial probability values, the longer is the time needed by the population to reach the steady state. Convergence times are also quite shorter in the high learning condition compared to the baseline treatment, since the previously described mechanism of a shift towards the Pareto-dominant strategy A at the indifference point drives the population of late adopters more rapidly towards technology A.
In the third treatment (Figure 10) the convergence times are strongly compressed, as the result of the *forgetting* parameter, which imposes an upper bound on the value of $q_{i,t}(A)$ and $q_{i,t}(B)$. Thus learning rates decrease up to some point and then remain relatively stable, rather than converging to zero.
Finally, in the last learning treatment (with both *experimentation* and

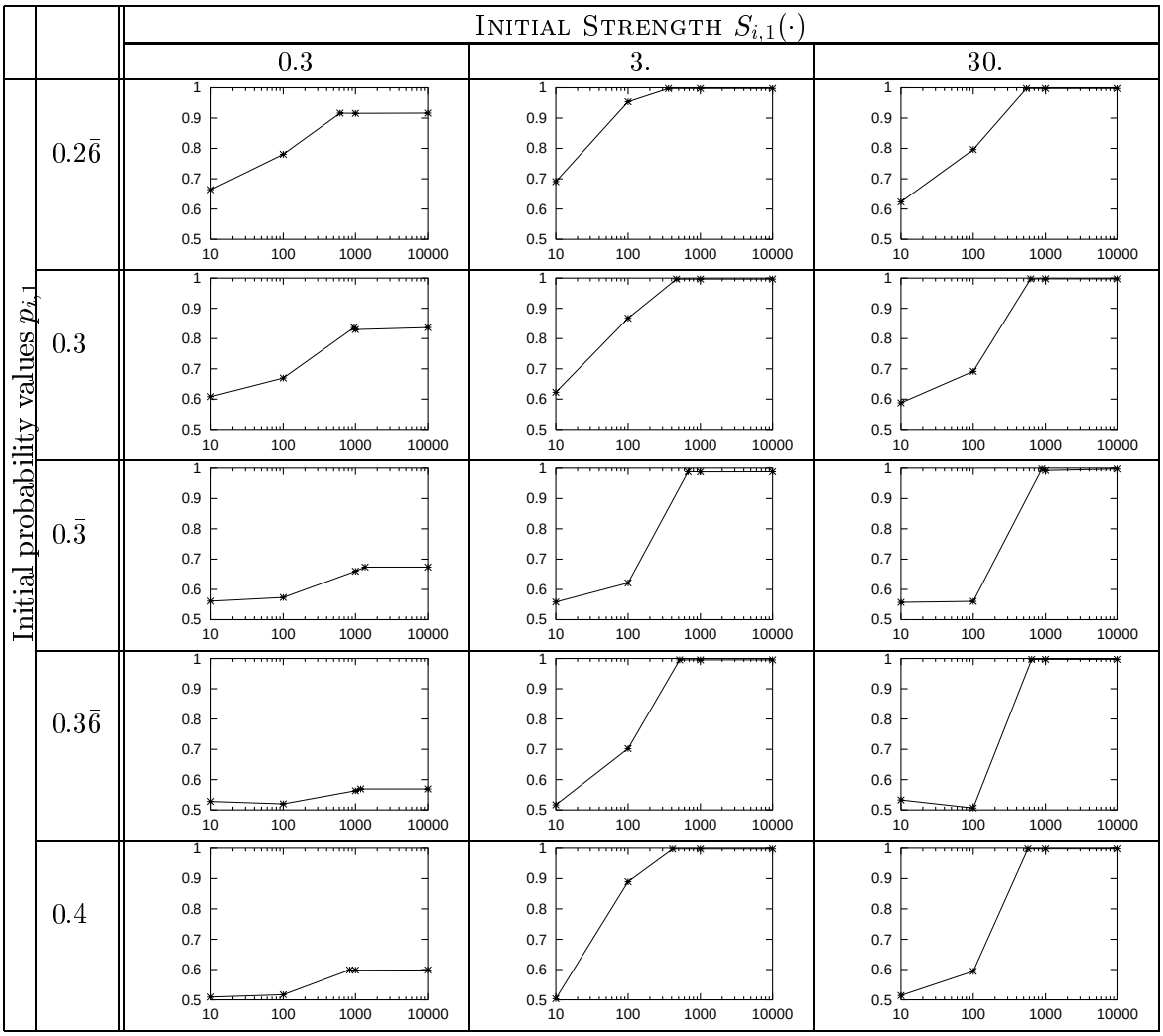

Figure 10: Concentration Index – Forgetting Model ($\varphi = .01, \varepsilon = 0$)

forgetting, Figure 11), all simulation trials ended at period 10000, since the interaction between *exploration* and high levels of learning rate due to *forgetting* resulted in a considerable update over time of the probability values, so that the stopping rule was never matched.

5 Discussion

In this paper we have conducted preliminary analysis of the dynamic of co-ordination processes in a population of consumers when behaviour is myopic and people adjust their capabilities step by step in order to discriminate among alternatives on the basis of an environmental stimulus. Simulations are of course only a method to gather suggestions about the possibly relevant dynamics. Our feeling is that the results of this first attempt open the way to a possibly rich area of research.
Stochastic learning appears to be a powerful force in driving the co-ordination dynamic. The forces at work in the case of learning are of three kinds:

the rate of learning: when the rate of learning is higher, agents stick more closely to the experience that rewarded them initially. Learning, in some sense, reduces the curiosity of players and the good solution is replicated without looking for better outcomes. In this case the population splits into two clusters, each group being attracted by the progressive reinforcement of initial choices;

the persistence of ambiguity: the role of learning may be balanced by the space given to ambiguity, that is, the persistence of some opportunity to make different choices. In Roth and Erev's model this is due to *experimentation*, which impedes learning from wiping out initial ambiguity;

the initial conditions of the system: It is clear that if the initial conditions are close to a particular stable attractor, the learning forces must increase in strength in order to enable the system to escape from it.

We expect the next stage of our work to move in the following directions:

- we have shown some effects of learning on a pure symmetric co-ordination game with a Pareto-dominant equilibrium. It would be interesting to investigate what happens when the co-ordination problem has different characteristics, as in cases of asymmetric co-ordination games or when a Pareto dominant equilibrium is contrasted by a risk dominant equilibrium, as in stag hunt game. Initial simulations of this last case have shown that previously described effects are stronger;

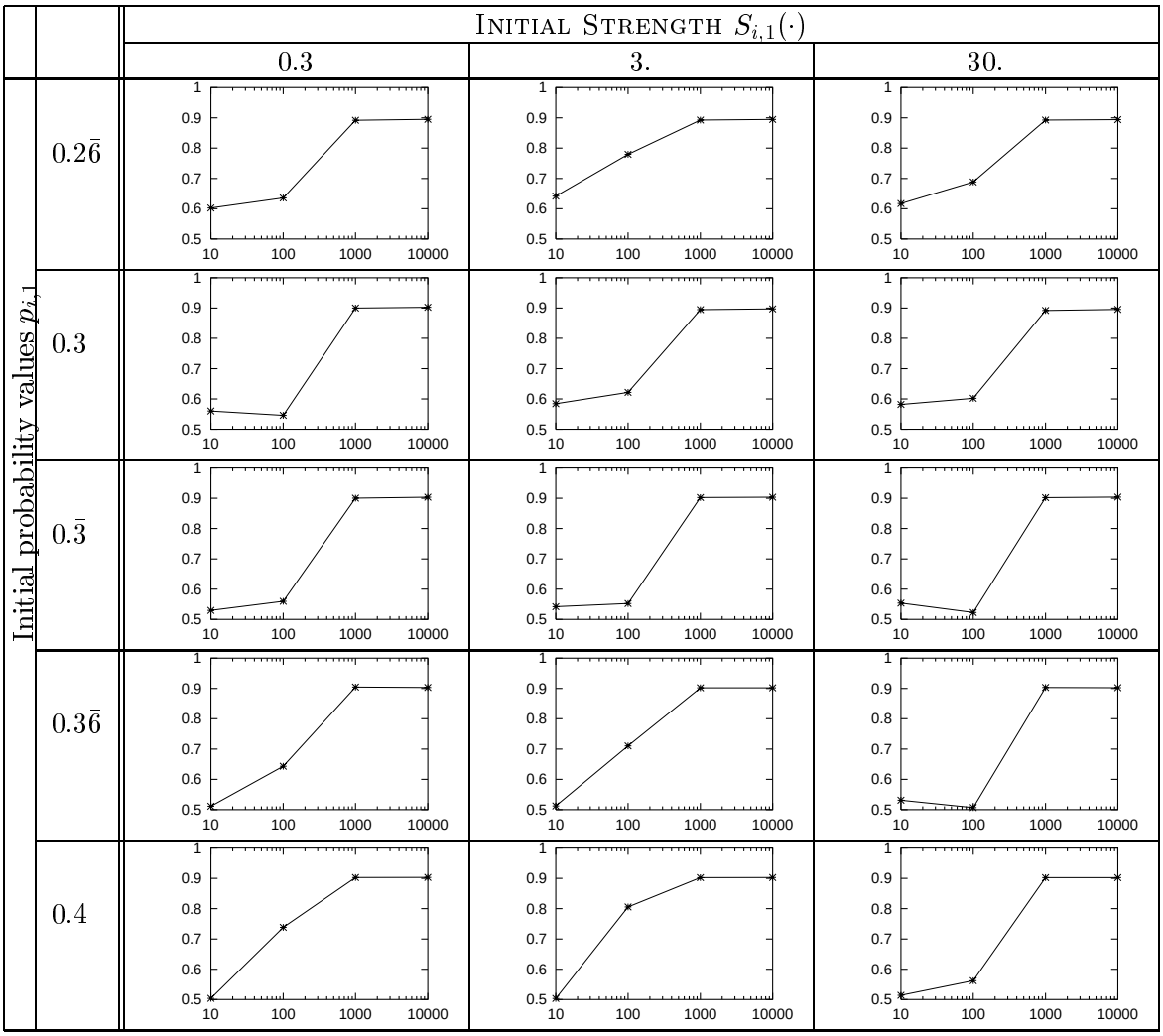

Figure 11: Concentration Index – Experimentation and Forgetting Model ($\varphi = .01, \varepsilon = 0$)

- spatial localisation could be taken into account. A local matching rule could be used to detect the effect of a differentiation of connections among agents. Our first simulations in this respect have shown that the non-uniform matching rule increases the effect of learning in producing local clusters;
- repeated co-ordination games are only a benchmark to test the effect of learning. In reality, it is implausible that the agents in a population adapt themselves only on the basis of repeated choices. It would be interesting to modify the model by introducing a different way of learning, taking into account, for instance, the experience of others [Lane and Vescovini, 1996, Narduzzo and Warglien, 1996];
- simulation is a good instrument insofar as it provides an instrument with which to detect the emerging properties of a complex system. But it is probably worth trying to verify with the help of formal analysis whether there are conditions under which stochastic learning processes display good properties, as a way to solve the co-ordination dilemma efficiently.

References

W.B. Arthur. Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal*, 99(394):116–31, 1989.

J.M. Blackburn. Acquisition of skill: An analysis of learning curves. Technical Report 73, IHRB, 1936.

C. Camerer and T. Ho. Experience-weighted attraction learning in games: A unifying approach. mimeo, 1996.

G. Ellison. Learning, local interaction, and coordination. *Econometrica*, 61 (5):1047–71, 1993.

I. Erev and E.A. Roth. Modeling how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. mimeo, 1997.

M. Kandori, G.J. Mailath, and R. Rob. Learning, mutation, and long run equilibria in games. *Econometrica*, 61(1):29–56, 1993.

Y.M. Kaniovski and H.P. Young. Learning dynamics in games with stochastic perturbations. *Games and Economic Behavior*, 7:330–63, 1995.

D. Lane and R. Vescovini. Decision rules and market-share: Aggregation in an information contagion model. *Industrial and Corporate Change*, 5(1): 127–46, 1996.

A. Narduzzo and M. Warglien. Learning from the experience of others: An experiment on information contagion. *Industrial and Corporate Change*, 5(1):113–26, 1996.

J. Öchssler. Decentralization and the coordination problem. *Journal of Economic Behavior and Organization*, 32(1):119–35, 1997.

E.A. Roth and I. Erev. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, 8(1):164–212, 1995.

B.F. Skinner. *Science and Human Behavior.* MacMillan, New York, 1953.

E.L. Thorndike. Animal intelligence: An experimental study of the associative processes in animals. Psychological Monographs, 2, 1898.

J.B. Watson. *Behaviorism.* University of Chicago Press, Chicago, 2nd edition, 1930.

The Rock Group

Rock Group (Research on Organizations, Coordination and Knowledge) is strongly committed to develop theoretical and empirical analyses of organizational key issues such as coordination among agents, decision making processes, coalition formation, replication and diffusion of knowledge, routines, and competencies within the organizational environment.
Rock Group is rooted in the Department of Management and Computer Science (DISA) of the University of Trento, which is characterized by a strong cooperative culture among scholars from different fields, from Management to Mathematics, from Statistics to Computer Science. Rock Group's activities, both research and education benefit by the whole range of competencies of DISA.

Mario Borroi, Ph.D. in Organization and Management (University of Udine). His main research interests are in the management and organization of innovation, technology transfer and the contribution of science to R&D activities. He has been visiting scholar at the Marshall Business School, USC, and had previous experience in the development of the Office for Technology Transfer at the University of Trento and in management consulting.

Paolo Collini, associate professor in Accounting and Management. Current interests involve cost management issues. He is member of the Editorial Management Board of the European Accounting Review.

Loris Gaio, assistant professor in Organization Economics and Management. Current interests involve coordination problems among economic agents, network economics and technology standards, with a particular attention for ITC issues. He has particular skills in

information technology, shaped by relevant experiences both as analyst and researcher.

Alessandro Narduzzo, Ph.D. in Management, currently post-doc, has been visiting scholar at the Cognitive Science Dept. of the University of California, San Diego, and at CREW, the University of Michigan. He studies organizational learning within a cognitive frame, creation and replication of tacit knowledge and the impact of IT on organizations.

Alessandro Rossi, Ph.D. candidate in Organization and Management, has been visiting scholar at the Wharton School, University of Pennsylvania. His interests are in behavioral game theory and managerial decision making and he studies how bounded rationality and experience affect decisions in organizations. He has also skills in field studies of regional manufacturing systems.

Luca Solari, Ph.D. in Organization and Management, temporarily appointed as professor in Human Resource Management at DISA. His main research interests are in population ecology and no-profit organizations. He has been visiting scholar at the Haas School of Business at U. C. Berkeley.

Enrico Zaninotto, full professor in Organization Economics and Management, and Dean of Faculty of Economics. His research interests are in coordination problems, explored both by a game-theoretical approach and field studies. Apart from the academic curriculum, he had some relevant experiences of management and consulting.

Affiliations and other group's members

Rock Group has close and systematic relationships with other scholars who are involved in Group's activity:

Vincenzo D'Andrea, assistant professor in Computer Science at Disa, his interests are in cellular automata, parallel and image processing, multimedia, tools for managing networked information.

Giovanna Devetag, is attending the Scuola Superiore of S. Anna program in Economics of Innovation (Pisa). Her research interests are in behavioral game theory, behavioral decision making, theory of mental models. She has been visiting scholar at Princeton, Department of Psychology.

Fabrizio Ferraro, Ph.D. candidate in Organization and Management (University of Udine), is involved in research activity at the University of Naples. He is mainly interested in research on organizational theory and in issues related to organizational impacts of quality certification practices (ISO 9000). He is currently graduate student at the Engineering School, University of Stanford.

Elena Rocco, Ph.D. in Organization and Management (University of Udine), visiting scholar at the Collaboratory for Research on Electronic Work at the University of Michigan. Her current research interests focus on bargaining and the organizational impact of computer mediated communication.

Massimo Warglien, associate professor in Organization Economics and Management at Ca' Foscari University of Venice. His interests are in bounded rationality and decision making, cognitive science and theory of the firm. He is editor of the Journal of Management and Governance.

Rock Group is affiliated to the Computable and Experimental Economics Laboratory (CEEL) of the University of Trento.

How to contact

Mail Address:	Rock Dipartimento di Informatica e Studi Aziendali Università degli Studi di Trento Via Inama, 5 I-38100 Trento (Italy) +39-0461-88-2142 (Ph) +39-0461-88-2124 (Fax)
E-mail:	`rock@cs.unitn.it`
Web URL:	`http://www.cs.unitn.it/rock`