ECONOMICS WORK

Economics Working Paper 24

On Adaptive Learning in Strategic Games

Ramon Marimon*
and
Ellen McGrattan†

June 1992

UNIVERSITAT POMPEU FABRA

Balmes, 132
Telephone (343) 484 97 00
Fax (343) 484 97 02
08008 Barcelona
e- mail econwp@upf.es

Abstract

We provide an overview of adaptive learning in strategic form games. We define a general class of adaptive learning rules and show how this class covers most rules studied in the literature. We also show how evolutionary models can be integrated in this general framework. Finally, following Marimon [60], we analyze the asymptotic behavior of adaptive learning algorithms and characterize the equilibria that these algorithms select.

1 Introduction

In most economic and game theoretical models, agents have perfect knowledge of the consequences of their actions. In order to gain this knowledge, agents are typically informed about the actions of other players or about a probability distribution over those actions. What happens when an agent does not have all of this information? Consider, for instance, a decentralized economy where agents only know their own choice sets and payoff functions. How does an agent choose the strategy that maximizes his expected payoff under these circumstances?

In this paper we provide an overview of the adaptive learning approach to answering this type of question. More specifically, we study infinitely played strategic-form games where agents behave myopically. We assume that the players do not consider the strategic consequences of their actions. We define a general class of learning algorithms and relate their asymptotic behavior to known and new solution concepts. The properties of these adaptive learning algorithms can be interpreted as providing a behavioral foundation for equilibrium theory.

An alternative approach that also attempts to provide a foundation for equilibrium theory is the rational choice-theoretical approach. Consider, for instance, a set of players that behave as Bayesian decision-makers. In this case, players postulate some distribution over the unknown elements affecting their payoffs. Further assumptions are needed, however, to reach an outcome in which agents' beliefs are fulfilled. For example, Aumann[4] has shown that if players are Bayes-rational in every state of the world and if they share a common prior, then the resulting outcome is a correlated equilibrium. All possible joint actions that might affect the payoffs of a player are included in the set of states over which the prior is defined.

A weaker requirement on prior coordination is that the structure of the game and the rationality of the players are common knowledge.[1] If this requirement is satisfied and if it is common knowledge that the players act independently, then they will only use rationalizable strategies. That is, the agents will only play strategies that survive the process of successive elim-

[1]For example, in a two player game, common knowledge of rationality means that player one is rational and knows that player two is rational, that player two is rational and knows that player one is rational, that player one knows that player two knows player one to be rational, and so on *ad infinitum*.

ination of strategies that are not a best response to the opponents' play. Assumptions such as common knowledge of rationality can be used to reduce the set of possible outcomes, but they might not suffice to eliminate nonintuitive outcomes. Joint restrictions on beliefs or mutual knowledge of opponents' strategies are needed. However, any form of joint restriction, such as a common prior, requires a degree of coordination that might not exist in a decentralized environment.[2]

If a game is played repeatedly, players might form their beliefs about the unknown elements of the game based on their experience. Common experience can help to coordinate agents' beliefs and, asymptotically, the game can have a well-defined outcome. Furthermore, in many environments, relatively simple decision rules can be good proxies for optimal behavior.

Milgrom and Roberts [68], and Gul [36] have explored the relationship between adaptive learning and rational choice. They show that when agents use adaptive learning rules, they play only *rationalizable* strategies in the long run. This result is remarkable in that it shows how repeated myopic learning achieves the same result that, in a static context, requires a rational deductive process. Furthermore, the inductive learning process does not require the strong common knowledge assumptions needed by the rational deductive process.

In this paper we follow the adaptive learning approach. We define a large class of adaptive learning rules, which includes most of the learning algorithms discussed in the literature but excludes some of the less intuitive ones considered by Milgrom and Roberts. The class of adaptive learning algorithms defined in this paper has three basic properties. First, every player is most likely to use the pure strategy with the best recorded performance at any point in time. We call this property *adaptation*. Second, every player experiments with every strategy in his strategy set with positive probability. We call this property *experimentation*. Third, the agents' learning processes must be imperfectly correlated. This last condition is achieved when every player continues to play the mixed strategy played in the last period with positive probability. We call this last property *inertia*.

One of our aims in providing a general characterization of the adaptive learning algorithms is to relate and integrate different results reported in the

[2]For a more complete treatment of the decision-theoretic approach to game theory see Brandenburger [10] [5].

rapidly emerging literature on learning[3]. In addition to Milgrom and Roberts [68], some of the current work on adaptive learning includes Fudenberg and Kreps [30], Krishna [58], Arthur [1] [2] [3], Jordan [40] [41] [42] [43], Brock [11], Crawford [18] [19] [21], Miller [69], Blume [7], Blume and Easley [8], Kalai and Lehrer [45] [46] [47], and Nyarko [75] [74]. In this paper, we do not attempt to survey all of this work, but we do point out some elements of adaptive learning algorithms shared by most of the learning models cited. Of the Bayesian learning models, we only discuss those that can be integrated with adaptive learning. We do show, however, how models of evolutionary dynamics can be integrated in a general adaptive learning framework. In this way, we relate adaptive learning with the expanding literature on evolutionary game theory and more specifically with the recent papers of Young [95] [24], Kandori, Mailath and Rob [50] [51] and Fudenberg and Harris [32]. Other related studies of evolutionary dynamics include Friedman [27], Canning [15], Fudenberg and Levine [31], Matsui and Rob [65], Samuelson [79] [80] [81], Cabrales and Sobel [14], Dekel and Scotchmer [22], Swinkels [86] [87], Nachbar [72], Selten [83], Ellison and Fudenberg [23], Robson [77] and [78] and Gilboa and Matsui [35].

We build on the work of Marimon [60], who refines the results of Milgrom and Roberts [68] by showing that when players use adaptive learning rules with inertia and experimentation, either the strategies played converge to a *robust equilibrium* or they cycle around a set of correlated strategies that belong to a *robust-recurrent set*. These robust-recurrent sets are proper subsets of the set of rationalizable strategies, and in games, such as the battle of the sexes[4], where the entire game is rationalizable, the only robust-recurrent sets are the two pure strategy equilibria. In such a game, the play of adaptive learners converges, with probability one, to one of the pure strategy equilibria. In this paper, we illustrate these convergence results and we show that, within our general class of adaptive learning rules, the evolution of play is sensitive to the specific algorithm used and to its parameterizations.

The paper is organized as follows. In Section 2 we discuss some general features that characterize adaptive learning. In Section 3, we formalize our general class of learning algorithms. In Section 4, we show how evolutionary

[3]In taking this approach, we follow David Kreps's advice [57] contained in his comments to our earlier paper on learning using Holland's Classifier Systems [61].

[4]See Section 2 for a description of the battle of the sexes game.

environments in which agents are matched can be mapped into the framework described in Section 3. In Section 5, we analyze the relationship between the asymptotic dynamics of adaptive learning and the refinements of Nash equilibria; we illustrate this relationship with different simulated examples. Finally, Section 6 concludes the paper.

2 General Features of Adaptive Learning

Before proceeding with the specifics of the learning algorithms, we illustrate some aspects of adaptive learning with a simple example. Game 1, Γ_1, is the strategic form game commonly known as *the battle of the sexes*. The payoff matrix for Γ_1 is described below, where $a > 1$. In this game, coordination is achieved either by an explicit mechanism or by learning from past experience when the game is repeated.

Γ_1	a_2	b_2
a_1	a,1	0,0
b_1	0,0	1,a

Milgrom and Roberts [68] consider a class of learning algorithms that converge to the serially undominated strategy profile. In the battle of the sexes example, every strategy is serially undominated. Milgrom and Roberts' class of algorithms is described as "adaptive" because players only assign positive weight to strategies that are the best response to strategies observed during a finite past history of the game.[5] They show that this is enough to guarantee that only rationalizable strategies will be played asymptotically. Further, if the sequence of play converges to a pure strategy profile then this profile must be a Nash equilibrium. This not withstanding, the class of rules that Milgrom and Roberts consider also includes some algorithms that can hardly be called adaptive in the sense of conforming to the best response. Consider, for example, that player i observes his opponents' play and uses the following rule: in period t, player i checks the past history of his opponents' plays, $\{c_{-i,t-m}, \ldots, c_{-i,t-1}\}$, where $c_{-i,t}$ is the pure strategy profile played by i's opponents in period t. Player i then assigns uniform weight to every

[5] An alternative definition of the adaptive learning analyzed in this paper is provided in Section 3.

strategy which is a best response to some strategy profile previously played in the given finite history of the game. This rule is adaptive in the Milgrom-Roberts sense, but it postulates a form of behavior that may not conform to the best-response map. Consider the battle of the sexes game, if player 1 observes that player 2 has chosen strategy a_2 $m-1$ times out of the last m plays and strategy b_2 only once, he still plays strategies a_1 and b_1 with equal probability. It seems more reasonable to think that proper adaptive behavior would assign a higher probability to a player repeating the strategies most frequently played. This last type of behavior its more likely to coordinate plays and beliefs.

But just conforming to the best response map may not be enough to result in this coordination of plays and beliefs. For example, if both players follow a rule that instructs them to play "best response to the previous play" may result in lack of coordination or cycles. In the battle of the sexes if the players start by choosing strategy (a_1, b_2) the following period they will play the best response to this strategy which is (a_2, b_1) to which they will respond by playing again (a_1, b_2) and so on. As a result, they never coordinate their responses and they always receive a payoff of zero. In this case, both players are being extremely reactive to each other's play but they are not realizing that they are being continuously misled. The problem, however, is not that the players are overreacting in the sense that they are only taking into account the previous period play. If they were to respond to a frequency distribution of plays, the same lack of coordination could also arise (see, for example, [95] and [43]).

Further, this lack of coordination is robust to perturbations as long as both players revise their strategies concurrently. To illustrate this point, consider the following modification of the Cournot rule. In period t, player i plays, the (pure) strategy that he played the previous period with probability $1-\rho_t$, and he revises his strategy with probability ρ_t. Whenever he revises his strategy, he chooses the best response to the opponents' last move with probability $(1-\epsilon_t)$ and any other available strategy with probability ϵ_t. In this algorithm, $1-\rho_t$ denotes the player's inertia and parameter ϵ_t the experimentation rate. Table 1 shows the transition probabilities that obtain in the battle of the sexes when both players follow this modified Cournot rule. The states are the four possible joint plays (c^{ij}) where c^{ij} is the state in which player 1 plays strategy i and player 2 plays strategy j.

Table 1. Transition probabilities for the modified Cournot rule in Γ_1.

$p(c^{ij}\|c^{kl})$	c^{11}	c^{12}	c^{21}	c^{22}
c^{11}	$(1-\rho_t)^2$ $+2\rho_t(1-\rho_t)(1-\epsilon_t)$ $+\rho_t^2(1-\epsilon_t)^2$	$\rho_t(1-\rho_t)\epsilon_t$ $+\rho_t^2\epsilon_t(1-\epsilon_t)$	$\rho_t(1-\rho_t)\epsilon_t$ $+\rho_t^2\epsilon_t(1-\epsilon_t)$	$\rho_t^2\epsilon_t^2$
c^{12}	$\rho_t(1-\rho_t)(1-\epsilon_t)$ $+\rho_t^2\epsilon_t(1-\epsilon_t)$	$(1-\rho_t)^2$ $+2\rho_t(1-\rho_t)\epsilon_t$ $+\rho_t^2\epsilon_t^2$	$\rho_t^2(1-\epsilon_t)^2$	$\rho_t(1-\rho_t)(1-\epsilon_t)$ $+\rho_t^2\epsilon_t(1-\epsilon_t)$
c^{21}	$\rho_t(1-\rho_t)(1-\epsilon_t)$ $+\rho_t^2\epsilon_t(1-\epsilon_t)^2$	$\rho_t^2(1-\epsilon_t)^2$	$(1-\rho_t)^2$ $+2\rho_t(1-\rho_t)\epsilon_t$ $+\rho_t^2\epsilon_t^2$	$\rho_t(1-\rho_t)(1-\epsilon_t)$ $+\rho_t^2\epsilon_t(1-\epsilon_t)$
c^{22}	$\rho_t^2\epsilon_t^2$	$\rho_t(1-\rho_t)\epsilon_t$ $+\rho_t^2\epsilon_t(1-\epsilon_t)$	$\rho_t(1-\rho_t)\epsilon_t$ $+\rho_t^2\epsilon_t(1-\epsilon_t)$	$(1-\rho_t)^2$ $+2\rho_t(1-\rho_t)(1-\epsilon_t)$ $+\rho_t^2(1-\epsilon_t)^2$
$p(c^{ij}\|\cdot)$	$(1-\rho_t)^2+\rho_t^2$ $+4\rho_t(1-\rho_t)(1-\epsilon_t)$	$(1-\rho_t)^2+\rho_t^2$ $+4\rho_t(1-\rho_t)\epsilon_t$	$(1-\rho_t)^2+\rho_t^2$ $+4\rho_t(1-\rho_t)\epsilon_t$	$(1-\rho_t)^2+\rho_t^2$ $+4\rho_t(1-\rho_t)(1-\epsilon_t)$

Notice that in the case where there is no inertia, *i.e.*, where $\rho_t = 1$, the four states occur with equal asymptotic probability and both players get an average payoff of $(a+1)/4$. While this result is better than the one obtained without experimentation, the players are still not learning to fully coordinate their actions. This result arises because both players are revising their strategies simultaneously and, as a result they are being systematically misled. However, when there is some degree of inertia *i.e.*, where $\rho_t < 1$, the play will converge to one of the two pure strategy Nash equilibria (a_1, b_1) or (a_2, b_2) with probability one.

Inertia is a particular form of what we call "imperfectly correlated learning." This feature of learning algorithms results in a weak form of stationarity in the environment. If the players constantly change their strategies, no player can devise "tests" to find a strategy that performs well. Inertia introduces stationarity by guaranteeing that with positive probability, the (mixed) strategies of player i's opponents remain fixed in the period immediately following a revision of i's strategy with positive probability.

There are other forms in which stationarity can be introduced in the model. Young [95], for example, assumes that players learn on the basis of a *sample* of previous plays. That is, players take a sample of length k of the last m plays $(k < m)$. A low enough k/m ratio introduces inertia in Young's algorithms.

In the learning environment that generates the transition probabilities described in Table 1, experimentation plays only a minor role in preventing the coordination problem. In general, however, the choice of the sequence $\{\epsilon_t\}$ is an important determinant of both the learning dynamics and the resulting strategy profiles. Experimentation guarantees that all possible strategies are tried and may enable the players to find the global maximum between multiple local maxima. Experimentation is especially important when players do not know their payoffs at the outset or when they do not observe their opponents' play.

In summary, the battle of the sexes highlights the three properties that are needed in this class of learning algorithms to ensure convergence to one (strict) Nash equilibrium in these simple coordination games. First, the learning rules must be adaptive in the sense that they conform to the best reply map. Second, there must be some experimentation. Third, the players' learning processes must be imperfectly correlated in order to allow for some degree of stationarity. In the next section we formalize these properties and define our general class of learning algorithms.

3 Adaptive learning algorithms

In this section we characterize a general class of learning algorithms that have the three stated properties: *adaptation, experimentation*, and *imperfect correlation*. We then discuss three specific subclasses which encompass most learning algorithms discussed in the existing literature. These subclasses have different formulations of adaptation, experimentation and imperfect correlation and different assumptions about agents' information of the past history of play.

Before stating formal definitions, we introduce some notation. A strategic form game is denoted by $\Gamma = \{(C_i, \pi_i), i \in I\}$, where C_i is the set of pure strategies of player i and π_i is his payoff function. Player i receives a payoff of $\pi_i(c_i, c_{-i})$ when he plays c_i and his opponents play c_{-i}. Unless we state otherwise, I is a finite set of players. Let $n_i = \#|C_i|$ be the number of pure strategies for player i; $C = C_1 \times \ldots \times C_I$ is the set of joint strategies; $\Delta(C_i)$ is the set of mixed strategies for i; and $\Delta(C)$ is the set of correlated strategies. Then, $E_\sigma \pi_i$ is the expected payoff for i if the correlated strategy σ is played. If player i's opponents play σ_{-i}, then his best response is $B_i(\sigma_{-i})$, where

$$B_i(\sigma_{-i}) = \{\sigma_i \in \Delta(C_i) : E_{(\sigma_i, \sigma_{-i})}\pi_i(c_i, c_{-i}) \geq E_{(\tilde{\sigma}_i, \sigma_{-i})}\pi_i(c_i, c_{-i}),\ \forall \tilde{\sigma}_i \in \Delta(C_i)\}.$$

We consider infinitely played games, Γ_∞. Let $\mathcal{F}_t$ be the $\sigma-$field generated by all possible sequences of play of Γ_∞ up to period t. In particular, if h^t denotes a sequence of play up to period t, then $h^t = (\Gamma, \sigma_0, c_0, \ldots, \sigma_{t-1}, c_{t-1})$. That is, h^t describes all the information that may be available up to period t. In general, players have limited information about the game being played and $\mathcal{F}_{i,t}$ represents the information known by player i at the beginning of period t. We assume perfect recall; $\mathcal{F}_{i,t}$ includes at least player i's past history of play, although, as we will see, specific algorithms may only use finite memory. We are also interested in computing finite frequencies of play. Let $f_{i,m}^{\bar{c}_{-i}}(c_{-i,0}, \ldots, c_{-i,m-1})$ be the frequency of the opponents' joint strategy profile, $\bar{c}_{-i}$, in the last m plays of the game. That is, $f_{i,m}^{\bar{c}_{-i}}(c_{-i,0}, \ldots, c_{-i,m-1})$ $= (1/m)\sum_{n=0}^{n=m-1} \chi_{\bar{c}_{-i}}(c_{-i,n})$, where $\chi_{\bar{c}_{-i}}(\cdot)$ is the indicator function of the strategy profile $\bar{c}_{-i}$. The vector of frequencies for player i over m plays is, therefore, $f_{i,m} : C_{-i}^m \mapsto \Delta(C_{-i})$.

An *adaptive rule* is defined by a sequence of behavioral strategies which are measurable with respect to a player's available information.

Definition. A process $\mathcal{A}_i = \{A_{i,t}\}_{t=0}^{\infty}$ defines an **adaptive rule** if:

i. $A_{i,t} : \mathcal{F}_\infty \mapsto \Delta(C_i)$, and

ii. $A_{i,t}$ is measurable with respect to $\mathcal{F}_{i,t}$

We can now provide a formal characterization of adaptive *learning* rules.

Definition. An adaptive rule $\mathcal{A}_i$ is said to be consistent with **adaptive learning with experimentation** if,

i. (*Experimentation*) there exists $\{\epsilon_{i,t}\}$, $\epsilon_{i,t} \in (0,1), \sum_{t=0}^{\infty} \epsilon_{i,t} = +\infty$ such that, for every t, $\sigma_{i,t}(c_i) \geq \epsilon_{i,t}$, $\forall c_i \in C_i$, and

ii. (*Adaptation*) for every h^t there exists m such that, if $\bar{c}_i \notin B_i(f_{i,m}(c_{-i,t}, \ldots, c_{-i,t+m-1}))$ and $\hat{c}_i \in B_i(f_{i,m}(c_{-i,t}, \ldots, c_{-i,t+m-1}))$ then

$$E\Big[\frac{\sigma_{i,t+m+1}(\bar{c}_i)}{\sigma_{i,t+m+1}(\hat{c}_i)}\Big|\mathcal{F}_{i,t+m}\Big] < \frac{\sigma_{i,t+m}(\bar{c}_i)}{\sigma_{i,t+m}(\hat{c}_i)}$$

whenever $\sigma_{i,t+m}(\bar{c}_i) > \epsilon_{i,t+m}$

This definition implies that the adaptive rule must move in the direction of the best response map, given past frequencies of play. It does not require the players to choose the best response. As we will see, this general definition defines a wide class of learning algorithms, including most of those studied in the existing literature. The function $f_{i,m}$ in the definition of adaptive learning with experimentation need not be the frequency map [6].

As we argued in Section 2, the above class may be too large. For example, it includes the possibility that too much coordination in the learning process may lead to too little coordination of actions and beliefs. If the starting point is one in which players make decisions in a fairly uncoordinated fashion, it is not a strong restriction to assume that learning rules must be imperfectly correlated. The following condition imposes the necessary stationarity in the environment.

Stationarity Given a player i, there exist positive numbers $\{\eta_{i,t}\}$, satisfying $\sum_{t=0}^{\infty} \eta_{i,t} = \infty$, such that for every t and history of play, up to period t, and for every $c_{i,t}$

$$Prob\{\sigma_{-i,t+1} = \sigma_{-i,t}\} \geq \eta_{i,t}$$

We can, alternatively, define a condition in terms of inertia of a single agent learning process[3]. Formally,

Inertia Given a player i, there exist positive constants, $\gamma_{i,t}$, $\sum_{t=0}^{\infty} \gamma_{i,t}^{I} = +\infty$, such that for every t and history of play, up to period t, and for every $c_{-i,t}$

$$Prob\{\sigma_{i,t+1} = \sigma_{i,t}\} \geq \gamma_{i,t}$$

Notice that *i) inertia* implies *stationarity*, and that *ii) experimentation* does not imply *inertia*. Experimentation only requires that every pure strategy, and in particular the last period strategy, must be played with probability at least $\epsilon_{i,t}$. In contrast, inertia bounds the probability that the player changes his mixed strategy.

[6] We could alternatively consider some arbitrary strictly monotone order statistic, *e.g.* a map $f_{i,m} : C_{-i}^{m} \mapsto \Delta(C_{-i})$ such that $f_{i,m}^{\bar{c}_{-i}}$ is strictly increasing with respect to $\chi_{\bar{c}_{-i}}$. For the algorithms studied in this paper, there is no loss of generality in considering the frequency map.

Definition. An adaptive rule $\mathcal{A}_i$ is said to be consistent with **adaptive learning with inertia and experimentation** if it is consistent with adaptive learning with experimentation and satisfies the inertia condition.

We now consider three subclasses of adaptive learning rules and in Section 4 we show how most evolutionary learning algorithms can be mapped into adaptive learning algorithms. The difference between subclasses is the specification of information set of a player i at t, $\mathcal{F}_{i,t}$, and the process by which this player chooses $\sigma_{i,t}$. A particular learning algorithm is defined by a specification of $(\mathcal{F}_{i,t}, \epsilon_{i,t}, \rho_{i,t}, \sigma_{i,t})$. As we will see, most of the learning algorithms studied in the literature are specific elements of these subclasses.

3.1 Best-Reply Learning

This subclass includes variants of Cournot's rule, fictitious play, and Bayesian learning. The following information structure $(\mathcal{F}_{i,t})$ is assumed. Agents know their own payoffs but not necessarily the payoffs of their opponents. They observe the realized actions of their opponents. Given that the players observe their opponents' past history of play, they can use these frequencies of play as a basis to form their expectations.

Consider first the choice of strategies for player i. For a sequence of numbers, $\{\alpha_{i,t}\}$, $\alpha_{i,t} \in [0,1]$, and a mixed strategy $\bar{x}_i \in \Delta(C_{-i})$, define the sequence of vectors $x_{i,t} \in \Delta(C_{-i})$ recursively by $x^{\bar{c}}_{i,-1} = \bar{x}^{\bar{c}}_i$,

$$x^{\bar{c}}_{i,t} = x^{\bar{c}}_{i,t-1} + \alpha_{i,t}(\chi_{\bar{c}}(c_{-i,t}) - x^{\bar{c}}_{i,t-1})$$

where $x^{\bar{c}}_{i,t}$ is the element of $x_{i,t}$ associated with strategy $\bar{c}$ and $\chi_{\bar{c}}$ is an indicator function such that $\chi_{\bar{c}}(\bar{c}) = 1$. The vector $x_{i,t}$ can be thought of as player i's beliefs about his opponents' mixed strategies in time t. The class of best-reply learning algorithms is defined as follows. Player i chooses his behavioral strategy at t according to the following rule:

- with probability $\rho_{i,t}$, choose $\tilde{\sigma}_{i,t} \in B_i(x_{i,t-1})$ and let

$$\sigma_{i,t}(c_i) = \begin{cases} \epsilon_{i,t} & \text{if } \tilde{\sigma}_{i,t}(c_i) \leq \epsilon_{i,t} \\ \frac{\tilde{\sigma}_{i,t}(c_i)}{\sum_{\{\tilde{\sigma}_{i,t}(\hat{c}_i) > \epsilon_{i,t}\}} \tilde{\sigma}_{i,t}(\hat{c}_i)} \cdot (1 - \tilde{\epsilon}_{i,t}) & \text{otherwise} \end{cases}$$

  where $\tilde{\epsilon}_{i,t} = \epsilon_{i,t} \times \#\{\tilde{\sigma}_{i,t}(\hat{c}_i) \leq \epsilon_{i,t}\}$ and of g_i and not its arguments. This class of problems includes many of the environments for which we have experimental evidence.[7]

In the previous subsection we have discussed the special case where $E = C_{-i}$ and g_i is the identity function. Therefore, we can define a class of learning algorithms for this class of games as in the previous subsection. Briefly, let $\mathcal{F}_{i,t}$ represent all the information available to player i up to period t, which now includes his own payoffs, his own history of play, and the history of the public outcome. Define a sequence of vectors $x_t \in \Delta(E)$, for a given sequence of numbers, $\{\alpha_{i,t}\}$, $\alpha_{i,t} \in [0,1]$, and a $\bar{x}_i \in \Delta(E)$, by $x^{\bar{e}}_{i,-1} = \bar{x}^{\bar{e}}_i$,

$$x^{\bar{e}}_{i,t} = x^{\bar{e}}_{i,t-1} + \alpha_{i,t}(\chi_{\bar{e}}(e_{i,t}) - x^{\bar{e}}_{i,t-1})$$

where $x^{\bar{e}}_{i,t}$ is the element of $x_{i,t}$ associated with the public outcome $\bar{e}$ and $\chi_{\bar{e}}$ is an indicator function such that $\chi_{\bar{e}}(\bar{e}) = 1$.

In this case, player i forms beliefs about the outcome $e_{i,t}$ in period t. The definition of competitive learning assumes the same behavioral strategy as in best-reply learning with the best response map defined over the sequence of vectors $x_t \in \Delta(E)$ instead of $x_t \in \Delta(C_{-i})$. Lemma 1 has the following corollary.

Corollary. Assume $\sum_{t=0}^{\infty} \alpha_{i,t} = +\infty$, $\sum_{t=0}^{\infty} \epsilon_{i,t} = +\infty$, $\sum_{t=0}^{\infty}(1 - \rho_{i,t}) = +\infty$ and that i's payoffs can be represented by $\pi_i : C_i \times E \mapsto I\!R$ where g_i is some function, $g_i : C \mapsto E$ and E is a finite set. Competitive learning rules are consistent with adaptive learning with inertia and experimentation.

Example 5. *Coordinating Efforts.* Set $g_i(c_i, c_{-i}) = min(c_i, c_{-i})$, $\alpha_{i,t}$ and $x_{i,t}$ as in Examples 1-4, and let E be a finite set of possible production outputs.

Example 5 is an example studied by Bryant [13]. Each player chooses a level of effort in production of a public good. The payoffs to i depend on i's own level of effort and the amount of the good produced. In this case, the number of goods produced is equal to $min(c_i, c_{-i})$ where c_i is the input of i and c_{-i} are the inputs of i's opponents.

[7] See, for example, van Huyck *et al.* [91] [92], Cooper *et al.* [17] and Friedman [28].

3.3 Adaptive Evolutionary Learning

By *adaptive evolutionary learning*, we mean a class of learning algorithms that assumes players have minimal information about the evolution of the game. Here we assume that players know their own number of pure strategies and can observe their own realized payoffs. They do not observe the actions or payoffs of their opponents. These informational assumptions are captured by the sequence $\{\mathcal{F}_{i,t}\}_t$ for each player i.

Players assign a measure of performance to each strategy based on its past payoffs. We call this measure the "strength" of a strategy. The term strength has been used in describing optima found by applying genetic algorithms; genetic algorithms rely on the ideas of genetics to find global solutions to optimization problems. Let $S_{i,t}(c_i)$ be the strength or value assigned by player i to strategy c_i in period t. The strengths are updated using the average realized payoffs as follows:

$$S_{i,t}(c_i) = \begin{cases} S_{i,t-1}(c_i) - \frac{1}{\eta_{i,t-1}(c_i)}\Big(S_{i,t-1}(c_i) - \pi_i(c_i, c_{-i})\Big) & \text{if } i \text{ plays } c_i \\ S_{i,t-1}(c_i) & \text{otherwise.} \end{cases}$$

The function $\eta_{i,t}(c_i)$ is the number of times that strategy c_i was played between the period of the last revision of i's mixed strategy $(\sigma_{i,t})$ and period t. The "clock" is updated as follows:

$$\eta_{i,t}(c_i) = \begin{cases} \eta_{i,t-1}(c_i) + 1 & \text{if } i \text{ plays } c_i \text{ at } t \\ \eta_{i,t-1}(c_i) & \text{otherwise.} \end{cases}$$

The strengths are used by the agents to evaluate and revise their different strategies. In particular, we assume that mixed strategies evolve as follows:

$$\tilde{\sigma}_{i,t}(c_i) = \begin{cases} \sigma_{i,t-1}(c_i) \cdot \frac{S_{i,t-1}(c_i)}{\sum_{\hat{c}_i} \sigma_{i,t-1}(\hat{c}_i) S_{i,t-1}(\hat{c}_i)} & \text{with prob. } \rho_{i,t} \\ \sigma_{i,t-1}(c_i) & \text{with prob. } 1 - \rho_{i,t}. \end{cases}$$

The sequence $\{\rho_{i,t}\}$ determines whether or not player i revises, or, in the language of genetics, reproduces, his current strategy. If player i does revise his mixed strategy in period t, all of the clocks are reset, *i.e.* $\eta_{i,t}(c_i) = 1$, for all c_i.

There is some probability that player i experiments with different strategies. As in the previous algorithms, the sequence of $\{\epsilon_{i,t}\}$ governing experimentation is such that

$$\sigma_{i,t}(c_i) = \begin{cases} \epsilon_{i,t} & \text{if } \tilde{\sigma}_{i,t}(c_i) \leq \epsilon_{i,t} \\ \frac{\tilde{\sigma}_{i,t}(c_i)}{\sum_{\{\tilde{\sigma}_{i,t}(\hat{c}_i) > \epsilon_{i,t}\}} \tilde{\sigma}_{i,t}(\hat{c}_i)} \cdot (1 - \tilde{\epsilon}_{i,t}) & \text{otherwise} \end{cases}$$

where $\tilde{\epsilon}_{i,t} = \epsilon_{i,t} \times \#\{\tilde{\sigma}_{i,t}(\hat{c}_i) \leq \epsilon_{i,t}\}$, $\epsilon_{it} \in (0,1)$, and $\sum_t \epsilon_t = +\infty$.

Lemma 2. Assume $\sum_{t=0}^{\infty} \epsilon_{i,t} = +\infty, \sum_{t=0}^{\infty}(1 - \rho_{i,t}) = +\infty$. Adaptive evolutionary learning rules are consistent with adaptive learning with inertia and experimentation.

A proof of Lemma 2 follows from the observation that for every h^t there exists a m such that if $\bar{c}_i \notin B_i(f_{i,m}(c_{-i,t}, \ldots, c_{-i,t+m-1}))$ and $\hat{c}_i \in B_i(f_{i,m}(c_{-i,t}, \ldots, c_{-i,t+m-1}))$ then

$$\rho_{i,t+m+1} \frac{S_{i,t+m}(\bar{c}_i)}{S_{i,t+m}(\hat{c}_i)} \frac{\sigma_{i,t+m}(\bar{c}_i)}{\sigma_{i,t+m}(\hat{c}_i)} + (1 - \rho_{i,t+m+1}) \frac{\sigma_{i,t+m}(\bar{c}_i)}{\sigma_{i,t+m}(\hat{c}_i)} < \frac{\sigma_{i,t+m}(\bar{c}_i)}{\sigma_{i,t+m}(\hat{c}_i)}$$

or $S_{i,t+m}(\bar{c}_i) < S_{i,t+m}(\hat{c}_i)$. It is always possible to find an m satisfying this inequality. Notice that if player i revises his strategy in some period between t and $t+m$, the above inequality is satisfied.

We can now describe several examples of adaptive evolutionary rules. We first set some initial conditions. For example, let $(\sigma_{i,0}(c_i), \eta_{i,0}(c_i), S_{i,0}(c_i)) = (\frac{1}{n_i}, 1, 0)$ for all i, c_i.

Example 6. *Exogenous Reproduction and Experimentation.* Set $\rho_{i,t} = 1/t^{p_i}$, $\epsilon_{i,t} = 1/t^{q_i}$, $p_i < 1$, $q_i < 1$.

Example 7. *Exogenous Reproduction (m periods).* Fix an integer m. Set $\rho_{i,t} = \hat{\rho}$ for some constant $\hat{\rho} \in (0,1)$ if $t = n \cdot m$, for some integer n, and $\rho_{i,t} = 0$ otherwise.

In all of the above examples, the algorithms have important parameters, such as the probability of revising strategies, that are exogenously given. In this sense, the algorithms are a fairly crude approximation to human learning. It seems more reasonable to assume that learning parameters are endogenously given. For example, one might think that "we change our mind when our mind is at odds with reality." The following algorithm tries to capture this feature.

Example 8. *Endogenous Reproduction (or Entropy).* The relative entropy of ν with respect to υ is given by

$$H(\nu,\ \upsilon) \;=\; \sum_k \nu^k \cdot log\frac{\nu^k}{\upsilon^k} \;\geq\; 0.$$

Notice that, $H(\nu,\ \nu) \;=\; 0$. Define $\rho_{i,t}$ endogenously as follows:

$$\rho(\hat{\sigma}_{i,t+1},\ \sigma_{i,t}) \;=\; \frac{H(\hat{\sigma}_{i,t+1},\ \sigma_{i,t})}{1 + H(\hat{\sigma}_{i,t+1},\ \sigma_{i,t})}$$

where $\hat{\sigma}_{i,t+1}(c_i)$ is the new mixed strategy that would have had resulted if player i would have changed his strategy, *i.e.,*

$$\hat{\sigma}_{i,t+1}(c_i) = \frac{S_{i,t}(c_i)}{\sum_{\hat{c}_i} \sigma_{i,t}(\hat{c}_i) S_{i,t}(\hat{c}_i)} \cdot \sigma_{i,t}(c_i)$$

In other words, if the current play suggests that the strategy used by player i should be very different from the one being played, then it should be revised with very high probability. That is, the *status quo* strategy is revised with high probability when the current test results in a very different strategy.

4 Alternative Matching Environments and Evolutionary Dynamics

We are interested in relating and integrating models of adaptive learning and models of evolutionary dynamics for two main reasons. First, the Darwinian dynamic process known as the "survival of the fittest" can also be viewed as a process of social learning. Second, because both types of models have similar underlying features and dynamic properties, they can be analyzed with a common framework.

A standard evolutionary model consists of the following elements: *i*) a payoff matrix; *ii*) a set of players, often identified by their strategy spaces; *iii*) a matching technology that describes the interaction between players; *iv*) a specification for the "replicator dynamics" which determine the growth rate of the population that is using a given strategy, *i.e.* in evolutionary models, the fraction of agents using a particular strategy grows or decays depending on the performance of their strategy relative to that of the average strategy, and these changes are reflected in the replicator mapping; and v) some form of individual or social experimentation, such as mutation in genetic models.

Most evolutionary models complement these features with the assumptions that individual players are both atomistic and myopic. Players are assumed to have no influence on the social outcome and to take no account of the future evolution of the environment. That is, they do not take into account the strategic effects of their actions.

The competitive assumption that players are atomistic has two main components. The first component is behavioral or subjective and it strictly implies that individual agents do not take into account the strategic effects of their actions. The second component is environmental or objective and it implies that the actions of individual agents do not have aggregate effects.

It is precisely this environmental component of the competitive assumption that is sometimes used as a partial rationalization of the assumption of myopic behavior. If a player cannot influence the social outcome, then there is no strategic effect to take into account. In evolutionary and adaptive learning models, however, myopic behavior is also taken to mean that agents do not follow well-defined deductive processes in forming their expectations about the future course of the economy or the game. But in rational expectations competitive equilibrium models, agents do not behave myopically even though their individual actions have no effects on the social outcome.

In most evolutionary models, players are randomly matched before each play of the game. On the other hand, in most learning environments, and certainly in all environments considered in Section 2, this is not the case. In spite of this difference, adaptive learning and evolutionary dynamic models have many features in common. We highlight the following. First, both types of models assume that players behave myopically. While it can be argued that in learning models with a small number of agents, individual actions affect social outcomes and players have incentives to be more sophisticated, it should be noted that this is also the case in evolutionary models with finite populations. Second, in adaptive learning, agents tend to assign higher probabilities to strategies with better recent performance. In this sense, adaptive learning provides a foundation for replicator dynamics. Third, by introducing a matching technology, the learning models described in Section 3 become standard evolutionary models albeit with possibly finite populations. Finally, some matching technologies and some forms of social experimentation in evolutionary models are isomorphic to some changes in the inertia and experimentation parameters of learning models with a fixed and constantly matched set of players. In this section we discuss these issues briefly and in the process of doing so we comment on some recent work in evolutionary game theory.

Before we proceed we need some additional definitions and assumptions. In general, we consider games with I types of players. Two players are defined to be of the same type, say type i, if they have the same set of actions and the same payoff functions, (C_i, π_i)[8]. Unless otherwise stated, we only consider games with a finite number of types I and a finite number of players J. Let r_i be the fraction of players of type i, then $m_i = J \times r_i$ is the number of players of type I.

An additional assumption in the evolutionary context is the application of the standard expected utility hypothesis. An implication of this assumption is that the interdependence of the players' actions in their payoff functions can be aggregated by player-types to obtain a simple and convenient form. For example, the expected payoff for player j of type 1 playing $c_{j,1}$ when his opponents play $c_{-(j,1)}$ is $\sum_{c_2 \in C_2} \cdots \sum_{c_I \in C_I} \mu_2(c_2) \times \cdots \times \mu_I(c_I)\ \pi_1(c_{j,1}, c_{-(j,1)})$ where $\pi : C \mapsto R$ for $C = C_1 \times \ldots \times C_I$, and where $\mu_i(c_i)$ is the fraction of players of type i that play the pure strategy c_i. In this case, therefore, the fraction of agents playing a pure strategy c_i, $\mu_i(c_i)$, in the evolutionary formulation of the model, is isomorphic with the mixed strategy, $\sigma_i(c_i)$, played by agent i in the standard game theoretic formulation with I players.

Using this isomorphism between fractions of players in evolutionary models and mixed strategies in standard games it is easy to see how adaptive learning can provide a foundation for replicator dynamics. A standard primitive of evolutionary models is that the fraction of players using a particular strategy evolves according to a law of motion of the form

$$\frac{\mu_{i,t+1}(\bar{c}_i)}{\mu_{i,t}(\bar{c}_i)} = h\left(\frac{E_{(\bar{c}_i,\sigma_{-i,t})}\pi_i(\bar{c}_i, c_{-i})}{E_{(\sigma_{i,t},\sigma_{-i,t})}\pi_i(c_i, c_{-i})}\right)$$

where $h : R \mapsto R$ is an increasing differentiable function with $h' \geq 1$. The above equation implies that, $\mu_i(\bar{c}_i)$ has a positive growth rate as long as the

[8]It should be noted that in evolutionary models, a "type" or a "phenotype" usually refers to the fraction of players of a specific game theoretical type who play the same pure strategy. In this paper we use type in the game theoretic sense.

expected payoff to strategy $\bar{c}_i$ is greater than average. Moreover, if we take $h(x) = \beta x$, for some constant $\beta \geq 1$, then we obtain

$$\frac{\mu_{i,t+1}(\bar{c}_i)}{\mu_{i,t+1}(\hat{c}_i)} = \frac{E_{(\bar{c}_i,\sigma_{-i,t})}\pi_i(\bar{c}_i, c_{-i})}{E_{(\hat{c}_i,\sigma_{-i,t})}\pi_i(\hat{c}_i, c_{-i})} \frac{\mu_{i,t}(\bar{c}_i)}{\mu_{i,t}(\hat{c}_i)}$$

which satisfies our hypothesis of adaptive learning. Conversely, our hypothesis of adaptive learning can be shown to result in a type of replicator equation that determines a similar law of motion for the fraction of players.

In models with a continuum of players, where one player of each type is randomly selected to play the I-players game, the above isomorphism between fractions of players and mixed strategies is justified by the law of large numbers. In this case, the expected payoffs of the randomly selected players will be determined by the fractions of players of each type that would be playing a given strategy. It could be argued, however, that when agents are myopic they take into account current and past payoffs rather than expected payoffs. To justify the use of the expected payoff computation in this case, consider the following cases. First, consider a modification of the above model where players are randomly matched in samples of I distinct types and then compute the average payoff of each sample. In this case, we would be computing the expected payoff of a cross-section of myopic players. Second, use this same modification of the model and compute the expected payoff of the time-path of an individual that is constantly matched with random independent samples of the same population. In this case, we would be computing the expected payoff of the time series of an agent's payoffs. The first case corresponds to the standard analysis of evolutionary dynamics while the second case is a study of individual learning as described in the previous section (albeit simplified by the fact that the actions of any individual player have no effect on the actions of the others).

To illustrate the isomorphism between the adaptive learning model and the evolutionary model, consider two types of agents playing the battle of the sexes game (Γ_1). Suppose that we have a continuum of agents of each type and that, every period, every agent of one type is randomly matched with one player of the other type. Suppose further that every agent plays the best response to the strategy played by his opponent in the previous period. In this case, the evolution of the population strategies is determined by the transition probabilities of the learning algorithm described in

Table 1. The probabilities of inertia and experimentation in the adaptive model correspond, respectively, to the fraction of agents that do not change their strategies and to the fraction of agents that experiment or "mutate" in the evolutionary model.

In general, however, player populations are finite. Consider the case where there are m players of each type, and where I players, one of each type, are successively and randomly matched m times before they revise their strategies. This model of matching and learning can be shown to correspond to either one of Examples 2 and 7 of Section 3 depending on how players choose to revise their strategies. The only difference between the examples of Section 3 and this case is that, now, the strategy of an opponent is not a unique mixed strategy but, rather, the average mixed strategy of the m players. Since this average is also a mixed strategy, it follows that the dynamic analysis of this case is the same as in that of the adaptive learning examples.

The competitive models of Section 2, are yet another example of models where evolutionary and learning dynamics can be easily integrated. Indeed, taking population fractions as given is a very simple form of dependence of the individual payoff functions on an aggregate statistic. Furthermore, macroeconomic applications of coordination problems often have the formulation of random pairing of agents taking a particular action. By introducing this random pairing in Example 5, for instance, we obtain a version of Diamond's search model.

While an evolutionary model with I types can be easily mapped into a game with I players, many evolutionary models have only one population or type. In this case the mapping requires symmetry of the payoff matrix. Consider, for example, the "Hawks versus Doves" example. (See Maynard Smith [66].)

Example 9. *Hawks versus Doves.* Consider the game, Γ_2, with payoff matrix given by

Γ_2	a_2	b_2
a_1	1,1	4,2
b_1	2,4	3,3

Hawks play a while doves play b. In this case, $I = 1$ and the expected utility for i playing $\bar{c}_i$ is $\sum_{c_{-i} \in C_{-i}} \mu(c_{-i}) \pi_i(\bar{c}_i, c_{-i})$ where $\mu(c_{-i})$ is the distribution over strategies of i's opponents. If there are two populations of players repeatedly matched, then $\mu(c_{-i})$ is the distribution for the population of i's opponents. If there is one population, μ is the distribution for the single population. Therefore, in the case of a single population, the action of a player affects his own distribution. As we show below, the nature of the matching technology can play an important role in the outcome of the game.

Assume that players are randomly matched. In this case, the dynamics of the model depend crucially on whether there are two populations repeatedly matched against each other or a single population from which pairs are extracted and repeatedly matched. We attain convergence to one of the two strict Nash Equilibria regardless of whether we have two fixed players repeatedly playing against each other or whether we have random matchings of players extracted from two different and fixed populations. In the case where there is only one population, there is convergence to the mixed strategy equilibrium which is the unique evolutionary stable equilibrium of the game. Figure 1, plots the frequency of play for strategy (b_1, a_2) for simulations of the two alternative models. In the first model, there is only one population of agents from which random matches are made and the game is then played. The series denoted "one population" reports the frequency of play in this case. The second model assumes that matches are made between agents extracted from two different populations. The series denoted "two populations" shows that this economy converges to the Nash Equilibria where player one becomes a dove (*i.e.*, ends up only using the b strategy) and player two becomes a hawk[9]. Note that although we are comparing two environments which use a matching technology, we can also mimic the results of the two populations case with two fixed players who use the learning algorithms described in Examples 2 and 7 of Section 3.

There are, however, some evolutionary models, that are not easily mapped into the general learning model of Section 3. But, as we will argue, this result arises from some of the special assumptions made in those models. For example, the special assumption that an evolutionary process is a stationary Markovian process is very convenient because one can use the theory of sta-

[9]In our simulation, the single population has 50 agents. Experimentation and inertia are described by $\epsilon_{i,t} = .1/t^{\frac{1}{5}}$ and $\rho_{i,t} = .99$ for $t = 50n$, $n = 1, 2, \ldots$ and $\rho_{i,t} = 0$, otherwise. The initial conditions are $(\sigma_{i,0}(c_i), \eta_{i,0}(c_i), S_{i,0}(c_i)) = (\frac{1}{2}, 1, 0)$ for all i, c_i. In the case of the two populations, each of them has 50 agents of one type and set $\epsilon_{i,t}$, $\rho_{i,t}$, and initial conditions are the same as in the single population case.

tionary Markovian chains in the analysis of finite games. Unfortunately, this assumption severely restricts the types of environments that can be studied. We have already discussed a version of Markovian dynamics in our example of the battle of the sexes. To make the environment stationary, it is enough to assume that the probabilities of inertia and experimentation remain time invariant. Unfortunately it is not so obvious why inexperienced players at the beginning of playing the repeated game should play in the same way as when they gain experience. To justify this form of stationary experimentation Canning [15] and Fudenberg and Levine [31], for example, include a fraction of newly born and inexperienced players who enter the game at every stage.

The dynamics for learning models and evolutionary models with finite populations are typically path-dependent. With a finite number of players, even when there is random matching, the experience of every player depends on the pure strategies played by each of his opponents. The particular sample or realization that a player experiences affects his behavior and, therefore, the social outcome. In this way, play becomes naturally correlated. To make a model Markovian, one must somehow eliminate this source of path dependence.

Having a very large but finite number of agents does not, in general, eliminate path-dependence[10]. To make their model stationary Markovian, Kandori *et al.* [50], [51] assume, in addition to constant inertia and experimentation rates, a complete matching of the players. In other words, after m periods, every player has played against all the players each of whom uses a pure strategy that remains unchanged throughout the m periods. On the other hand, Young [95] assumes that players behave as statisticians taking samples from their finite past histories. This randomization device, which as we have seen in Section 2 introduces inertia into the model, can also suppress the sources of path dependence and reduce the model to a stationary Markovian model. Other ingenious devices, such as population renewal, have also been tried. But, as is usually the case, the gain in analytical tractability is obtained at the expense of restricting the environments.

The study general adaptive learning models that allow for path depen-

[10] As Boylan has pointed out in the context of evolutionary models, one must be careful not to abuse the law of large numbers even in cases when there is a countable number of players[9].

dence is not as much an issue of elegance and generality. On the contrary, it should be noted that the dynamics of learning and evolution can be crucially sensitive to the choice of modeling technique. As we will see in the following section, the predictions of a stationary Markovian model with experimentation that is constant through time ($\epsilon_t = \epsilon$ for all t) do not correspond to the predictions of the general adaptive model with time-dependent experimentation and with ϵ_t approaching to zero.

5 Learning and Equilibrium Selection

There are at least two reasons why we are interested in studying learning in strategic form games. First, even we would like to determine the asymptotic properties of learning in a larger class of economic environments, the simple structure of strategic form games provides a good starting point. Second, in games with multiple equilibria, we would like to be able to characterize those equilibria that can be selected by the process of learning. In this section, we study the asymptotic properties of the class of adaptive learning algorithms studied in Section 2; for the most part, we follow Marimon [60]. We assume here that the learning rules are consistent with adaptive learning with inertia and experimentation. We also assume that experimentation decreases over time, *i.e.* $\epsilon_{i,t}$ approaches 0.

We want to complete and make precise the following statements.

i. If play converges to a strategy profile, then the strategy profile is ____.

ii. If a strategy profile is a ____, then the profile is asymptotically stable.

iii. With probability one, play converges asymptotically to a set of strategy profiles which is ____.

iv. If the game has ____, then play converges with probability one to ____ profile.

Different learning processes will imply different results to be filled into the above statements. Notice the difference between these statements. The first characterizes the outcome that can emerge from learning. The second says that a particular profile has the following property: for some neighborhood of this profile (in $\Delta(C)$), learning processes with play in this neighborhood converge to that profile. Statements *i* and *ii* are local convergence results. The third only guarantees that asymptotically play will remain in a certain set (in C). The fourth statement provides a global convergence result for learning by restricting the class of games under consideration. There are known weak forms of these statements.

i. If play converges to a strategy profile, then the strategy profile is *a Nash equilibrium.*

ii. If a strategy profile is a *strict Nash equilibrium* then, it is asymptotically stable.

iii. With probability one, play converges asymptotically to a set of strategy profiles, which is *the set of rationalizable strategies.*

iv. If the game is *acyclic*, has *a unique Nash equilibrium* and it is *strict*, then play converges with probability one to *the equilibrium* profile.

The first result has been labeled the *folk theorem* of the learning literature. It appears, for example, in Milgrom and Roberts [68], and in Fudenberg and Kreps [30]. For Bayesian learning, there are versions of statement *i* in Jordan [42], Kalai and Lehrer [47], and Nyarko [74]. It is also satisfied for our class of adaptive learning algorithms. Notice that by the definition of adaptive learning with experimentation, if $\sigma_t \rightarrow \sigma*$ then it must be that for all i, strategy c_i (or a mixed strategy with all weight on c_i) is a best response (and in $B_i(\sigma^*_{-i})$) whenever $\sigma^*_i(c_i) > 0$. Otherwise there is a better response against the opponents' play and, given that other strategies are tried infinitely often, the player will deviate to his best response. However, as Fudenberg and Levine [31] have shown, this result does not have to be true in extensive form games with more than two players, since some parts of the tree may not be searched even when there is experimentation.

The resultin *ii* is based on the fact that a strict equilibrium σ^* has the property that for all i, $B_i(\sigma^*_{-i})$ is a singleton.[11] This fact not only precludes mixed strategy equilibria (*i.e.*, for some $c^* \in C$, $\sigma^*(c^*) = 1$), but also says

[11] A version of this statement appears, for example, in Fudenberg and Kreps[30]).

that small perturbations around this point are still mapped into this point (*i.e.*, $c_i^* \in B_i(\tilde{\sigma}_{-i}^*)$, where $\tilde{\sigma}^*$ is an arbitrary small perturbation of σ^*). From this characterization it should be clear why asymptotic stability holds around a strict equilibrium. Let the learning process start in the pre-defined small neighborhood of the equilibrium and make the experimentation rates sufficiently small, then adaptive learning guarantees convergence to the stricrt equilibrium.

The third result is that of Milgrom and Roberts [68] discussed in the first two sections. The fourth result is less obvious. It is based on the fact that with enough experimentation the strategy profile of the players must visit a neighborhood of the strict equilibrium and once in the neighborhood the second result applies. This "visiting result" is based on the fact that play can not get stuck somewhere else, for example in some cycle; and this uses the fact that the equilibrium is unique[12]. There are two alternative versions of the fourth statement.

iv b. If the game has *two players and two strategies*, then the beliefs generated by *fictitious play without experimentation* converge to *the beliefs of a Nash equilibrium* profile.

iv c. If the game has *strategic complementarities and diminishing returns*, then *players' beliefs* converge to *the beliefs of a Nash equilibrium* profile.

It should be noted that both results are about convergence of beliefs. As Jordan [43] has shown, we can have convergence of beliefs to a mixed strategy equilibrium without having convergence of the strategies played to the corresponding mixed strategy.

As we will see, generically, there is no convergence of play to mixed strategies (see also [19]). That is, if one considers a two-by-two game satisfying the properties of (*iv c*) with a unique Nash equilibrium which is a mixed strategy equilibrium and players use algorithms in our class of adaptive algorithms with experimentation, play does not converge to the mixed strategy equilibrium even if beliefs may converge. The result in (*iv b*), which appears in Miyasawa [70], generalizes the same result for zero-sum games due to Robinson [76]. A famous example of Shapley[84] shows that this result does not

[12] A more precise definition of cycles is provided below.

generalize to games with more than two strategies.[13] The result in (*iv c*) is due to Krishna [58]. His assumptions require that strategies can be naturally ordered in a way that each player's best response is increasing in the other players' strategies and that his own strategies have diminishing returns.

In this section we discuss versions of the above statements and we also show by means of examples why further sharpening of the results requires us to restrict the class of learning models.

Example 10. *Risk Dominance and Optimality in Coordination Games.*

Consider the game, Γ_3, with payoff matrix given by

Γ_3	a_2	b_2
a_1	2,2	0,0
b_1	0,0	1,1

As in the battle of the sexes (Γ_1), all of the strategies in Γ_3 are rationalizable. Thus, statement *ii* above has no predictive power. The game satisfies the conditions of (*iv c*), but it also has the property that deviations from the equilibrium strategy are more costly at the Pareto optimal equilibrium since the Pareto optimal Nash equilibrium *risk dominates* the suboptimal equilibrium. For this class of games a stronger version of statement four is available.

iv d. If the game has *two players, two strategies, and two pure strategy equilibria ranked by risk dominance* then *the limiting distribution of evolutionary dynamics* converges to a *risk-dominant Nash equilibrium* profile.

By the limiting distribution of evolutionary dynamics it is usually understood the limit distribution that arises in a stationary Markovian process as the noise due to experimentation converges to zero. That is, assume the process governing experimentation is constant (*i.e.* $\epsilon_t = \epsilon > 0$ for all t) and that the matching technology is such that the evolutionary process can be characterized as a stationary Markovian process. Then we get a unique limit ergodic distribution, which is a function of ϵ. The limiting distribution is the

[13] We return to Shapley's example below.

limit of these ergodic distributions as ϵ converges to zero. In a the class of games characterized in *ivd*, the distributions concentrate most of their mass in the two pure strategy equilibria and there is a positive probability to move from one equilibrium to the other. However both transitions (from the good equilibrium to the bad and from the bad to the good) are not equally likely; they depend on the costs from deviations. Therefore, the ϵ ergodic measure is more concentrated in the risk dominant equilibrium and as ϵ approaches zero the limit ergodic distribution concentrates all the mass in this equilibrium (see Kandori, Mailath, and Rob [50] [51], Young [95], and Samuelson [81].) The results of Kandori *et al.* also generalize to other pure coordination games where the selected equilibrium is the efficient one [51]. In their case,

This selection result of Young [95] and Kandori *et al.*[50] is remarkable. Nevertheless, as we have discussed in Section 4, relies on the particular structure of the evolutionary model. [14] In fact, Fudenberg and Harris [32] have shown, in a continuous time model with a continuum of players, that if one makes different assumptions on how noise is introduced into the model, then the above result of a unique selection it may not be satisfied for certain parameter specifications.

We can also see why the result of Kandori *et al.*[50] is too strong to be satisfied for our class of adaptive learning algorithms. Consider adaptive evolutionary rules applied to Γ_3 in Example 10 with $\epsilon_{i,t} = .1/t^{\frac{1}{5}}$, $\rho_{i,t} = .5/t^{\frac{1}{5}}$, and initial conditions $(\sigma_{i,0}(c_i), \eta_{i,0}(c_i), S_{i,0}(c_i)) = (\frac{1}{2}, 1, 0)$ for all i, c_i, t. For one realization of the processes governing experimentation and inertia, we found the following mixed strategies being applied:

Period	$\sigma_1(a)$	$\sigma_1(b)$	$\sigma_2(a)$	$\sigma_2(b)$
0	.500	.500	.500	.500
1	.333	.667	.500	.500
10	.484	.516	.727	.273
25	.053	.947	.055	.945
100	.040	.960	.040	.960
1000	.025	.975	.025	.975

The results for period t, $t > 1000$, also show that player i chooses b_i with probability $1 - \epsilon_{i,t}$. Furthermore, follows from statement *ii* above that for

[14]For risk dominance, van Damme [90] shows how an inductive process will also select the risk dominant equilibrium in 2×2 games. The inductive process is similar to the evolutionary dynamics argument concerning the cost from deviations.

a set of initial conditions, play will converge with high probability to the inefficient outcome. As we noted in Section 4, the difference between our result and that of Kandori *et al.*[50] is that we let ϵ_t approach zero.

Example 10 can also be used to study the role that different parameters play in our class of learning algorithms. For example, if reproduction is low enough and ϵ_t converges to zero at a slow enough rate, then one can obtain convergence to the Pareto optimal equilibrium. Like *simulated annealing algorithms* for optimization, the learning algorithms can be parameterized in such a way that they achieve the global maximum.[15] In some sense, we would be mimicking the results of Kandori *et al.* Of course, if one is willing to impose a uniform prior and low enough reproduction, then it is easy to show that there is convergence to the Pareto optimal outcome. In this case, the uniform distribution is in the basin of attraction of the Pareto-optimal equilibrium for a large set of parameters.

Before we proceed with our examples it is convenient to properly define different equilibrium concepts which differ in the types of perturbations that are allowed. The following are well known concepts[16]:

Definition. σ^* is a **perfect equilibrium** if $\exists\{\sigma^n\}$, $\sigma_i^n(c_i) > 0\, \forall c_i \in C_i$ such that $\sigma^n \rightarrow \sigma^*$ and $\forall i\, \sigma_i^* \in B_i(\sigma_{-i}^n)$.

Definition. σ^* is a **strict equilibrium** if $\exists \epsilon > 0$ such that $\forall \tilde{\sigma}$ with $|\tilde{\sigma} - \sigma^*| < \epsilon$, $\forall i$, $\sigma_i^* \in B_i(\tilde{\sigma}_{-i})$.

Definition. $\bar{\sigma}$ is an ϵ-**proper equilibrium** if $\bar{\sigma}_i(c_i) > 0\, \forall c_i \in C_i$ and if for every pair of pure strategies $\hat{c}_i$, $\tilde{c}_i$ if $E_{(\hat{c}_i, \bar{\sigma}_{-i})}\pi_i(\hat{c}_i, c_{-i}) < E_{(\tilde{c}_i, \bar{\sigma}_{-i})}\pi_i(\tilde{c}_i, c_{-i})$ then $\bar{\sigma}_i(\hat{c}_i) \leq \epsilon \bar{\sigma}_i(\tilde{c}_i)$.

Definition. σ^* is a **proper equilibrium** if $\exists\{\epsilon^n\}, \{\bar{\sigma}^n\}$ such that $\epsilon^n \rightarrow 0$, $\bar{\sigma}^n \rightarrow \sigma^*$ and for every $n, \bar{\sigma}^n$ is a ϵ^n proper equilibrium.

Definition. A closed set of Nash equilibria Θ is **K-M stable** if it is minimal with respect to the following property. $\forall \epsilon, \exists \bar{\epsilon}$ such that $\forall \epsilon_i \in (0, \bar{\epsilon})$ and $\forall i$ and $\lambda_i \in \Delta(C_i)$, with $\lambda_i(c_i) > 0\, \forall c_i \in C_i$, the perturbed game $\Gamma_{\epsilon,\lambda}$ in which σ of Γ is replaced by $(1 - \epsilon_i)\sigma_i + \epsilon_i \lambda_i$ has a Nash equilibrium which is ϵ close to Θ.

[15] For a very readable introduction to simulated annealing, see Kirkpatrick[53].

[16] See, for example, Myerson[71].

Notice that a strict equilibrium is robust to all possible perturbations. For this reason, games with a unique Nash equilibrium which is a mixed strategy equilibrium, such as matching pennies, have no strict equilibria. A perfect equilibrium is robust to an open set of perturbations if the equilibrium is a pure strategy profile. However, if it is a mixed strategy profile, then perfection imposes no restrictions. This guarantees the existence of a perfect equilibrium but does not guarantee that every equilibrium profile is robust to an open set of perturbations. In games, such as matching pennies (Γ_6), it is not. In a learning context, perturbations are present due to experimentation. However, not all perturbations are necessarily present; one must take into account open sets of perturbations. But, the corresponding restriction may be too strong. If players experiment independently, then their experimentation does not necessarily cover an open set of perturbations in $\Delta(C)$. That is, with experimentation players may test all of their own available options, but these individual experimentations may not amount to considering all possible collective (correlated) options. The robustness concepts introduced in Marimon [60] are aimed at capturing these subtleties.

Definition. Let $M \subseteq I$ be a subset of players. A set $\mathcal{N} \subset \Delta(C)$ is an M-open neighborhood of $\bar{\sigma} \in \Delta(C)$ if $\forall \sigma \in \mathcal{N}$, $\forall \epsilon > 0$, $\exists \epsilon\prime$, $0 < \epsilon\prime < \epsilon$, such that $(\tilde{\sigma}_M, \bar{\sigma}_{-M}) \in \mathcal{N}$ whenever $|\tilde{\sigma}_M - \bar{\sigma}_M| < \epsilon\prime$.

Let $M_m = \{M \subseteq I : \#|M| \leq m\}$ be the set of subets of n players, where $n \leq m$.

Definition. Let $\sigma^* \in \Delta(C_1) \times \cdots \times \Delta(C_I)$. σ^* is a **m-robust equilibrium** if $\forall M \in M_m$, $\exists$ an M-open neighborhood of σ^*, $\mathcal{N}$, such that $\forall \tilde{\sigma} \in \mathcal{N}$ with $\tilde{\sigma}_i \in \Delta(C_i)$, and $\forall i \in I$, $\sigma_i^* \in B_i(\tilde{\sigma}_{-i})$.

Below we use the term **robust equilibrium** to mean a 1-robust equilibrium. A **strongly robust equilibrium** is the term used for $(I-1)$-robust equilibrium.[17]

Notice that in a two-players game an equilibrium is robust if and only if it is strongly robust and if and only if it is a perfect pure strategy equilibrium. However, a mixed strategy equilibrium, such as the equilibrium of

[17]It seems that Okeda used the term *robust* to refer to a point which is robust to all possible perturbations not just an open set of I-perturbations. However, here we use the term as defined in Marimon [60].

the matching pennies game, can be perfect and not robust. This also shows that a robust equilibrium may not exist. A pure strategy perfect equilibrium is robust and a strong robust equilibrium is robust, but, as the following example shows, in games with more than two players, a robust equilibrium may not be perfect.

Example 11. *Perfect Equilibrium.* Consider the game, Γ_4, with payoff matrix given by

Γ_4	a_2	b_2
a_1	1,1,1	0,0,0
b_1	0,0,0	0,0,0

a_3

a_1	0,0,0	0,0,0
b_1	0,0,0	1,1,0

b_3

In Γ_4[18] there is a perfect pure (proper) and efficient equilibrium, namely (a_1, a_2, a_3). Considered as a coordination game, this is the good equilibrium. There is a set of Nash equilibria consisting of $\{(b_1, b_2, \sigma_3(a_3)); \sigma_3(a_3) \in [0, 1]\}$ of which only (b_1, b_2, a_3) is perfect (and proper). All of these equilibria are robust. To see this, consider the equilibrium (b_1, b_2, b_3). It is not perfect since for player 3 strategy a_3 dominates strategy b_3, but it is only strictly better when both players 1 and 2 jointly experiment and move from (b_1, b_2) to (a_1, a_2). Robustness only considers individual experimentations and not perturbations that require jointly correlated actions.

We now illustrate with a few examples the relation between the above equilibrium concepts and the asymptotic dynamics of adaptive learning.

Example 12. *Stable Equilibrium.* Consider the games, Γ_5 and Γ_6, with payoff matrices given by

[18]Game Γ_4 has been discussed by Kalai and Samet [48].

Γ_5	a_2	b_2
a_1	2,2	2,2
b_1	3,0	0,1

and

Γ_6	a_2	b_2
a_1	1,-1	-1,1
b_1	-1,1	1,-1

Swinkels[87] shows that one can strengthen a modified version of statement *ii* by replacing "strict" with "Kholberg – Mertens stable". It cannot be the same statment *ii* with a relabeling of terms since not all K-M stable equilibria are asymptotically stable. For example, in the matching pennies game, Γ_6, the unique mixed strategy equilibrium is K-M stable but it is not asymptotically stable. In our context, Swinkels' result says that if there is a set which is asymptotically stable under adaptive learning then it contains a K-M stable subset (provided that the set has enough convexity). In the example, Γ_5, the stable component has player 1 using his top strategy and player two playing right with at least 1/3 probability. Notice that every element of the stable component is a *robust equilibrium.* In Figure 2, we see how player two's strategy can remain a mixed strategy of the stable component.

Example 13. *Robust Equilibrium.*

Consider the game, Γ_7, with payoff matrix given by

Γ_7	a_2	b_2	c_2	d_2
a_1	2,2	1,1	0,0	0,0
b_1	1,1	1,1	2,0	2,0
c_1	0,0	0,2	5,0	0,5
d_1	0,0	0,2	0,5	5,0

Example 13 shows that stability is a too strong requirement; even for pure strategy equilibria. The game has a strict equilibrium, (a_1, a_2), and a robust and perfect (but not K-M stable) pure strategy equilibrium, (b_1, b_2). In

Figure 3, we show that there can be convergence to the (b_1, b_2) equilibrium.[19] We plot the frequency of play for strategies (a_1, a_2) and (b_1, b_2); the strategy being used after 25000 periods is (b_1, b_2). In this case, the experimentation of strategies c and d reinforces convergence to strategy b.

As we have discussed before, in games with more than two players, robust equilibria may not be perfect. Consider again Example 11. The following realizations suggest that there can be convergence to any robust equilibrium. For example, if players 1 and 2 assign most weight to their b strategies, then it will require joint experimentation to test (a_1, a_2). Even if $\sum_{t=0}^{\infty} \epsilon_{i,t} = +\infty$, $\forall i$, we might have $\sum_{t=0}^{\infty} \epsilon_{1,t}\epsilon_{2,t} < \infty$. Thus, (a_1, a_2) is not tested infinitely often, *i.e.*, there is a positive probability that play will remain among the set of robust equilibria which are not perfect. Furthermore, notice that even if (a_1, a_2) is tested infinitely often (*i.e.*, $\sum_{t=0}^{\infty} \epsilon_{1,t}\epsilon_{2,t} = +\infty$), this does not guarantee that play will converge to the efficient perfect equilibrium. It is enough to guarantee that player 3 will assign most weight to strategy a_3. But if player 3's probability of experimentation is high enough, then the experiments of player 3 may reinforce players 1 and 2 to play their b strategies. For example, with an adaptive evolutionary algorithm in which players do not observe the point (a_1, a_2, a_3), play may converge to the strongly robust equilibrium (b_1, b_2, a_3). For example, player 1's strengths after a long phase without reproduction, $S_{1,t}(a_1) \simeq \epsilon_{2,t}(1 - \epsilon_{3,t})$ and $S_{1,t}(b_1) \simeq \epsilon_{3,t}(1 - \epsilon_{2,t})$, and if $\epsilon_{2,t} < \epsilon_{3,t}$, then player 1 will tend to assign more weight to strategy b_1. Of course, one could simplify the analysis by assuming $\epsilon_{i,t} = \epsilon, \forall i$ and t, as it is often done in the literature of evolutionary learning, but this is not an assumption that follows from adaptive individual learning.

Finally, we look at some examples where play does not converge, but follows cyclical patterns.

Example 14. *Correlated Equilibria and Cycling.*

Consider the games, Γ_8 and Γ_9, with payoff matrices given by

Γ_8	a_2	b_2	c_2
a_1	0,0	4,5	5,4
b_1	5,4	0,0	4,5
c_1	4,5	5,4	0,0

[19] In this case, two populations of 50 agents each are randomly matched each period. In period t, we set $\epsilon_{i,t} = .1(e^{\frac{-t}{10000}} + 1/t^{\frac{1}{8}})$ and $\rho_{i,t}$ as in Example 8 for each i.

and

Γ_9	a_2	b_2	c_2	d_2
a_1	0,0,0	1,9,1	9,1,0	1,0,9
b_1	9,1,1	0,0,0	1,9,1	0,0,0
c_1	1,9,1	9,1,1	0,0,0	0,0,0

a_3

a_1	0,0,0	1,1,0	0,1,9	9,0,0
b_1	1,1,0	$\frac{6}{5},\frac{6}{5},\frac{6}{5}$	1,1,0	0,0,0
c_1	1,1,0	1,1,0	9,0,1	0,9,1

b_3

Game Γ_8 is a famous example that Shapley [84] used to show that statement *ivb* on convergence of beliefs in 2×2 games does not generalize to games with more than two strategies. In Γ_8 there is a unique mixed strategy equilibrium with $\sigma_i(x_i) = 1/3$ and expected payoff 3. As in Shapley's example with fictitious play, algorithms from our general class follow a cyclical pattern with most of the frequency of play outside the diagonal. Play cycles around as follows: (a_1, b_2) to (c_1, b_2) to (c_1, a_2) to (b_1, a_2) to (b_1, c_2) to (a_1, c_2) and back to (a_1, b_2). Thus, we find an average payoff greater then 3.

Cyclical behavior can be fairly complex, and it does not occur because there are no pure strategy equilibria in a game. Γ_9, for example, has a strict equilibrium (b_1, b_2, b_3) and yet a cyclical pattern involving twelve strategy profiles may emerge. Figures 4a – 4d show the frequency of play of a repeated play of Γ_9.[20] Because of the complexity of these patterns, we illustrate the 12-stages of the cycle in four plots. Notice that the unique pure and strict strategy equilibrium, (b_1, b_2, b_3), is not played. This type of asymptotic pattern emerges from a large set of initial conditions.

Cyclical behavior has been studied in the context of evolutionary dynamics by Gilboa and Matsui [35]. However, their concept of social stability imposes conditions on the adjustment process of the replicator dynamics which

[20]In this case, we set $\rho_{i,t}$ as in Example 8 and $\epsilon_{i,t} = .1(e^{\frac{-t}{10000}} + 1/t^{1/5})$. Initial conditions are $(\sigma_{i,0}(c_i), \eta_{i,0}(c_i), S_{i,0}(c_i)) = (\frac{1}{n_i}, 1, 0)$ for all i, c_i.

do not have to be satisfied for the general algorithms that we consider here. Marimon extends the concept of robust equilibrium to encompass cycling sets of strategies. Cycling behavior can emerge when, through the learning process, players correlate their strategies. This phenomena suggests a refinement of correlated equilibrium. In fact, the robustness concepts provide a natural refinement of correlated equilibrium.

We first recall the definition of a correlated equilibrium.

The set of correlated strategies with full support in $D \in C$ is defined by $\mathcal{S}(D) = \{\sigma \in \Delta(C) : \text{supp}\{\sigma\} \subset D\}$. Given $\sigma \in \Delta(C)$ and $c_i \in \text{supp}\{\sigma\}$, the conditional correlated strategy is defined by

$$(\sigma|c_i)(\cdot) = \frac{\sigma(c_i, \cdot)}{\sum_{c_{-i}} \sigma(c_i, c_{-i})} \quad \in \Delta(C_{-i})$$

Definition. The **conditional best response** map $B(\cdot|\cdot)$ is defined, for all $c \in$supp$\{\sigma\}$, by

$$B(\sigma|c) = \left\{\tilde{c} \in C : \quad \forall i, \forall \hat{c}_i \in C_i, E_{(\sigma|c_i)}\pi_i(\tilde{c}_i, c_{-i}) \geq E_{(\sigma|c_i)}\pi_i(\hat{c}_i, c_{-i})\right\}$$

Definition. σ^* is a **correlated equilibrium**, denoted $\sigma^* \in CE(\Gamma)$, iff for all $c \in \text{supp}\{\sigma^*\}$, $\quad c \in B(\sigma^*|c)$.

Definition. σ^* is a **m-robust correlated equilibrium** if $\forall M \in M_m$, $\exists$ an M-open neighborhood of σ^*, $\mathcal{N}$, such that $\forall \tilde{\sigma} \in \mathcal{N}$ and $\forall i \in I$ and $c_i \in C_i$ with $\sigma^*(c_i) > 0$, $c_i \in B_i(\tilde{\sigma}|c_i)$.

We define a **robust correlated equilibrium** as a 2-robust correlated equilibrium and a **strongly robust correlated equilibrium** as a I-robust correlated equilibrium.

As with robust equilibria, robust correlated equilibria may not exist and therefore may be too strong of a requirement to characterize asymptotic behavior. Furthermore, as Shapley's example [84] shows, a cyclical pattern does not imply that the frequency of play converges to a well defined ergodic distribution. These facts suggest that we should not be looking at the correlated distributions but rather at their supports. The following definitions relax the requirements of robust correlated equilibrium.

Definition. Let $D \subset C$. The set of **m-admissible correlated strategies** for D is defined by,

$$\mathcal{T}_m(D) = \Big\{\sigma \in \Delta(C) : \forall c \in D, \sigma(c) > 0, \text{and } \forall M \in M_m, \\ \exists \text{ an } M\text{-open neighborhood of} \sigma, \mathcal{N}, \text{ s.t.} \\ \forall \tilde{\sigma} \in \mathcal{N},\ B(\tilde{\sigma}|c) \cap D \neq \emptyset\Big\}$$

Definition. A set $E \supset D$ is said to provide a **robust justification** for D if $\mathcal{T}_1(D) \cap \mathcal{S}(E) \neq \emptyset$.

Definition. Let $RJ(D)$ be the collection of *minimal* sets that provide a robust justification for D

Definition. A set D is said to be **robust self-justified** if $RJ(D) = \{D\}$.

Definition. A set $D \subseteq C$ is said to be a **robust-recurrent set** of Γ, denoted $RR(\Gamma)$, if it is a *minimal* robust self-justified set.

In [60] it is shown that a robust-recurrent set exists in any strategic form game[21]. Furthermore, robust-recurrent sets are either points or define a cyclical pattern of the conditional best response map. For example, in example Γ_7 the set of robust-recurrent sets, denoted $RR(\Gamma)$, consists of the two points (a_1, a_2) and (b_1, b_2) and the cycle $\{(c_1, c_2), (c_1, d_2), (d_2, d_1), (c_2, d_1)\}$. In [60] it is also shown that the first statement can be strengthened to include robust equilibria and that if $c \notin RR(\Gamma)$ then $\sigma_t(c) \to 0$ with probability one.

We are now in a position to make the final statements relating adaptive learning and equilibrium selection.

> *ib.* If play converges to a strategy profile, then the strategy profile is *a robust equilibrium.*

[21]Kalai and Samet define a ***persistent set*** to be a set of mixed strategies which is minimal with respect to the property of being robust to all possible perturbations; where these perturbations are defined as in the definition of strict equilibrium . They define, however, a ***persistent equilibrium*** to be any Nash equilibrium in a persistent set. A pure strategy persistent equilibrium is a strict equilibrium. The concept of ***Robust- Recurrent*** set is a weakening of the concept of persistent set in a similar way that the concept of robust equilibrium is a weakening of the concept of strict equilibrium (see also, [6]).

iib. If a strategy profile is a *strict Nash equilibrium* then, it is asymptotically stable.[22]

iiib. Asymptotically, play converges (with probability one) to a set of strategy profiles which is a *robust-recurrent set.*

ive. If the game has the property that all the *robust-recurrent sets are singletons* then play converges with probability one to a *robust equilibrium.*

These four statements provide a fairly complete characterization of the asymptotic properties of adaptive learning. They also show the sense in which a learning process may select among equilibria.

6 Concluding Remarks

In this paper we provide a fairly thorough overview of adaptive learning in repeatedly played strategic form games. In doing so, we cover a fair amount of new ground. First, we define a general class of adaptive learning algorithms which include most of the rules studied in the existing literature on learning. Second, we show how evolutionary theory and adaptive learning can be integrated in a natural way. Third, we discuss how different parameterizations of our class of learning algorithms affect the equilibrium outcomes. Finally, we analyze the asymptotic properties of learning algorithms and we relate them to known and new refinements of Nash equilibria.

There is an aspect of learning behavior that we do not directly explore, but it is present in our discussion. There is tradeoff between conservative and reactive behavior that characterizes most learning processes. This tradeoff is a tradeoff between two goals that are contradictory. In order to gain experience, a player must be conservative and change his strategies only infrequently. On the other hand, in the rapidly changing environment that is typical in the context of games, a player must be reactive in order to take advantage of the new options and opportunities. In our algorithms the tradeoff between conservative and reactive behavior is typically captured by the different probabilities of reproduction and experimentation. In many of the examples of Section 5, we show how the outcomes of the games respond

[22]This statement is true for certain specifications of inertia and experimentation.

to different assumptions about both the conservative or reactive behavior of players.

As is usually the case in the study of dynamic systems, the understanding of the asymptotic behavior often sheds little light on the transitional dynamics. Unfortunately, in learning environments, players spend most of their time in transition. We hope that the development of a general framework to study adaptive learning will eventually allow us to address this issue. In Section 4, for example, we describe how some differences in matching technologies can be isomorphic to some variations in the parameters of learning models and we show how some of these differences affect the final outcome and also the process of convergence.

Economic experiments performed with evolutionary games tend to confirm most of the general theoretical results discussed in this paper. The findings of these experiments, however, highlight the importance of assumptions such as the type of matching environment or the number of players. They suggest that we should direct our attention not only to the underlying game, but also to the interaction of players and the learning process that takes place across different games. Van Huyck et al., for example, report important group size effects in their experiments with the Stag Hunt game (Example 5 of Section 3). Friedman [26] has detected runs in which play fluctuates around an unstable mixed strategy equilibrium, albeit only after the same subjects have been playing a game with a unique stable mixed strategy equilibrium. Path dependence is also an important aspect of the findings of Marimon, Spear and Sunder in their experiments with monetary economies with multiple equilibria[23].

In spite of the increasing interest in learning and evolution among game theorists and economists, we are still laying down the basic elements of a theory. In a sense, the asymptotic theory is close to being complete. If we are only interested in a few asymptotic properties and equilibrium selection, then all of the recent work in the area provides a fairly complete account of these issues. But, if we want to achieve a better understanding of how agents learn and interact, of the role that economic institutions or social norms play, and of the impact that different learning environments have on achieving alternative social outcomes, then the theory is still in an early stage.

---

[23]See references [91], [92], [18], [28], [62], [63], [64].

References

[1] W.B. Arthur, On Learning and Adaptation in the Economy, unpublished paper, Stanford University, April 1992.

[2] W.B. Arthur, Designing Economic Agents that Act Like Human Agents: A Behavioral Approach to Bounded Rationality, *Learning and Adaptive Econ. Behav.* (May 1991), **81:2**, 353–359.

[3] W.B. Arthur, A Learning Algorithm that Mimics Human Learning, unpublished paper, Stanford University, July 1990.

[4] R. Aumann, Correlated Equilibrium as an Expression of Bayesian Rationality, *Econometrica* (1987) **55**, 1-18.

[5] R. Aumann and A. Brandenburger, Epistemic Conditions for Nash Equilibrium, Working Paper 91–042, Harvard Business School, 1991.

[6] D. Balkenborg, Properties of Persistent Retracts and Their Relations to Other Solution Concepts, unpublished paper, University of Bonn, August 1991.

[7] L. Blume, The Statistical Mechanics of Strategic Interaction, unpublished paper, Cornell University, January 1992.

[8] L. Blume and D. Easley, Wealth Dynamics and the Market Selection Hypothesis, Mimeo, Cornell University, 1991.

[9] R. Boylan, Evolutionary Equilibria Resistant to Mutation, mimeo Olin Business School, Washington University at St. Louis, 1990.

[10] A. Brandenburger, Knowledge and Equilibrium in Games, Discussion Paper 91-032, Harvard Business School, 1991.

[11] W. Brock, A Possible Unification of Some Results on Learning Dynamics in Games, Discussion Paper 9019, University of Wisconsin, Madison, August 1990.

[12] G. Brown, Iterated Solution of Games by Fictitious Play, in Activity Analysis of Production and Allocation, Wiley: New York, 1951.

[13] J. Bryant, A Simple Rational Expectations Keynes-Type Model, *Quarterly Journal of Economics*, (1983), **98**, 525-528.

[14] A. Cabrales and J. Sobel, On the Limit Points of Discrete Selection Dynamics, Discussion Paper 91-28, University of California, San Diego, September 1991.

[15] D. Canning, Convergence to Equilibrium in a Sequence of Games with Learning, STICERD Discussion Paper TE/89/190, LSE, 1989.

[16] I–K. Cho and A. Matsui, Learning the Ramsey Policy, unpublished paper, University of Chicago, March 1992.

[17] R. Cooper, D. DeJong, R. Forsythe and T. Ross, Selection Criteria in Coordination Games: Some Experimental Results, *American Economic Review,* (1990), **80**, 218-233.

[18] V.P. Crawford, Adaptive Dynamics in Coordination Games, Discussion Paper 92–02, University of California, San Diego, January 1992.

[19] V. Crawford, Learning and Mixed–Strategy Equilibria in Evolutionary Games, *J. Theor. Biol.* (1989), **140**, 537–550.

[20] V. Crawford, Nash Equilibrium and Evolutionary Stability in Large– and Finite–Population "Playing the Field" Models, *J. Theor. Biol.* (1990), **145**, 83–94.

[21] V. Crawford, An "Evolutionary" Interpretation of Van Huyck, Battalio, and Beil's Experimental Results on Coordination, *Games Econ. Behav.* (1991), **3** 25–59.

[22] E. Dekel and S. Scotchmer, On the Evolution of Optimizing Behavior, unpublished paper, University of California, Berkeley, 1991.

[23] G. Ellison and D. Fudenberg, Rules of Thumb for Social Learning, unpublished manuscript, MIT and IDEI, March 1992.

[24] D. Foster and H. Young, Cooperation in the Short and in the Long Run, *Games Econ. Behav.* (1991), **3**, 145–156.

[25] D. Foster and P. Young, Stochastic Evolutionary Game Dynamics, *Theor. Pop. Biol.* (1990), **38**, 219–232.

[26] D. Friedman, Evolutionary Games: Some Experimental Results, University of California, Santa Cruz, unpublished paper, February 1992.

[27] D. Friedman, Evolutionary Games in Economics, *Econometrica* (1991), **59**, 637–666.

[28] D. Friedman, Evolutionary Games: Some Experimental Results, unpublished paper, University of California, Santa Cruz, February 1992.

[29] J.W. Friedman, Low information Nash Equilibria for Oligopolistic Markets, *Information Economics and Policy* (1983), **1**, 37–53. North Holland.

[30] D. Fudenberg and D. Kreps, Learning and Equilibrium in Games, Draft. 1991.

[31] D. Fudenberg and D. Levine, Steady State Learning and Nash Equilibrium, Mimeo. 1991.

[32] D. Fudenberg and C. Harris, Evolutionary Dynamics with Aggregate Shocks, unpublished manuscript, MIT, February 1992.

[33] I. Gilboa and D. Schmeidler, Infinite Histories and Steady Orbits in Repeated Games, unpublished paper, Northwestern University, May 1992.

[34] I. Gilboa and D. Samet, Absorbent Stable Sets, unpublished paper, Northwestern University, April 1991.

[35] I. Gilboa and A. Matsui, Social Stability and Equilibrium, *Econometrica* (1991), **59**, 859–867.

[36] F. Gul, Rationality and Coherent Theories of Strategic Behavior, Research Paper #1990, Stanford Business School, 1991.

[37] W. Hines, Three Characterizations of Population Strategy Stability, *J. Appl. Probab.* (1980), **17**, 333–340.

[38] W. Hines, Evolutionary Stable Strategies: A Review of Basic Theory, *Theor. Pop. Biol.* (1987), **31**, 195–272.

[39] F. Hofbauer and K. Sigmund, The Theory of Evolution and Dynamical Systems, Cambridge University Press, Cambridge, 1988.

[40] J.S. Jordan, Convergence to Rational Expectations in a Stationary Linear Game, *Review of Econ. Studies* (1992), **59**, 109–123.

[41] J.S. Jordan, Bayesian Learning in Games: A Non-Bayesian Perspective, University of Minnesota, unpublished paper, June 1992.

[42] J.S. Jordan, Bayesian Learning in Normal Form Games, *Games Econ. Behav.* (1991), **3**, 60–81.

[43] J.S. Jordan, Three Problems in Learning Mixed–Strategy Nash Equilibria, University of Minnesota, unpublished paper, March 1992.

[44] B. Jovanovic, Selection and the Evolution of Industry, *Econometrica* (1982), **50:3**, 649–670.

[45] E. Kalai and E. Lehrer, Weak and Strong Merging of Opinions, unpublished paper, Northwestern University, February 1992.

[46] E. Kalai and E. Lehrer, Subjective Equilibrium in Repeated Games, unpublished paper, Northwestern University, February 1992.

[47] E. Kalai and E. Lehrer, Rational Learning Leads to Nash Equilibrium. Center for Mathematical Studies in Economics and Management Science Discussion Paper No. 925, Northwestern University, 1991.

[48] E. Kalai, and D. Samet, Persistent Equilibria in Strategic Games, *International Journal of Game Theory,* (1984), **13**, 129-144.

[49] M. Kandori, Independent Mutations and Long Run Equilibria in Games: A Continuous Time Formulation, Mimeo, Princeton University, 1991.

[50] M. Kandori, G.J. Mailath and R. Rob, Learning, Mutation, and Long Run Equilibria in Games, CARESS Working Paper #91–01R, University of Pennsylvania, 1992.

[51] M. Kandori and R. Rob, Evolution of Equilibria in the Long Run: A General Theory and Applications, CARESS Working Paper #92–06, University of Pennsylvania, 1992.

[52] Y–G Kim and J. Sobel, An Evolutionary Approach to Pre–Play Communication, Discussion Paper 92–01, University of California, San Diego, 1992.

[53] S. Kirkpatrick, C. Gelatt, and M. Vecchi, Optimization by Simulated Annealing, *Science*, (1983), **220**, 671-680.

[54] E. Kohlberg and J–F. Mertens, On the Strategic Stability of Equilibria, *Econometrica* (1986), **54**, 1003–1037.

[55] K. Kollman, J.H. Miller, S.E. Page, Adaptive Behavior and Political Theory, unpublished paper, Northwestern University, March 1992.

[56] K. Kollman, J.H. Miller, and S.E. Page, Adaptive Parties in Spatial Elections, unpublished paper, Northwestern University, February 1992.

[57] D. Kreps, Game Theory and Economic Modeling, Clarendon Press, Oxford, 1990.

[58] V. Krishna, Learning in Games with Strategic Complementarities, Harvard Business School Working Paper 92–073, March 1992.

[59] G.J. Mailath, JET Symposium on Evolutionary Game Theory: Introduction, CARESS Working Paper #92–08, April 1992.

[60] R. Marimon, Correlated Rationalizability and Adaptive Learning, Unpublished paper, Universitat Pompeu Fabra, 1992.

[61] R. Marimon, E. McGrattan and T. Sargent, Money as a Medium of Exchange in an Economy with Artificially Intelligent Agents, *J. Economic Dynamics and Control* (1990), **47**, 282-366.

[62] R. Marimon and S. Sunder, Indeterminacy of Equilibria in an Hyperinflationary World, unpublished paper, Carnegie Mellon University. 1990.

[63] R. Marimon and S. Sunder, Some Puzzles from Experimental Monetary Economies, unpublished paper, Carnegie Mellon University. 1991.

[64] R. Marimon, S. Spear and S. Sunder, Expectationally-driven Market Volatility: An Experimental Study, *Journal of Economic Theory* (forthcoming).

[65] A. Matsui and R. Rob, Evolution, Rationality and Equilibrium Selection in Societal Games, CARESS Working Paper #91–31, University of Pennsylvania, 1991.

[66] J. Maynard Smith, Evolution and the Theory of Games, Cambridge University Press, Cambridge, 1982.

[67] K. McCabe, S. Rassenti, V. Smith, An Experimental Study of Learning in Extensive Form Games with Private Information, unpublished paper.

[68] P. Milgrom and J. Roberts, Adaptive and Sophisticated Learning in Normal Form Games, *Games and Econ. Behav.* (1991), **3**, 82–100.

[69] J. Miller, Artificial Adaptive Agents in Game Theory, unpublished paper, Carnegie Mellon University, 1992.

[70] K. Miyasawa, On the Convergence of the Learning Process in a 2×2 Non-zero-sum Game, Econometric Research Program, Research Memorandum No 33, Princeton University, 1961.

[71] R. Myerson, Game Theory: Analysis of Conflict, Harvard University Press, Cambridge, MA, 1991.

[72] J. Nachbar, "Evolutionary" Selection Dynamics in Games: Convergence and Limit Properties, *Int. J. Game Theory* (1990), **19**, 59–89.

[73] R. Nelson and S. Winter, An Evolutionary Theory of Economic Change, Harvard University Press, Cambridge, MA, 1982.

[74] Y. Nyarko, Bayesian Learning Without Common Priors and Convergence To Nash Equilibria, unpublished paper, New York University, Fall 1991.

[75] Y. Nyarko, Bayesian Learning in Repeated Games Leads to Correlated Equilibria, unpublished paper, New York University, Spring 1992.

[76] J. Robinson, An Iterative Method of Solving a Game, *Annals of Mathematics*, (1951), **54**, 296-301.

[77] A. Robson, Efficiency in Evolutionary Games: Darwin, Nash and the Secret Handshake, *J. Theor. Biol.* (1990), **144**, 379–396.

[78] A. Robson, Evolutionary Game Theory, in Creedy J, Borland J, Eichberger J. (ed.) *Recent Developments in Game Theory*, Edward Elgar Publishing, Hants, England (1992).

[79] L. Samuelson and J. Zhang, Evolutionary Stability in Asymmetric Games, Working Paper #9024, University of Wisconsin–Madison, Sept. 1991.

[80] L. Samuelson, Limit Evolutionarily Stable Strategies in Two–Player, Normal Form Games, *Games Econ. Behav.* (1991), **3**, 110–128.

[81] L. Samuelson, How to Tremble if You Must, SSRI Working Paper 9122, University of Wisconsin, September 1991.

[82] R. Selten, Evolutionary Stability in Extensive Two–Person Games, *Math. Social Sciences* (1983), **5**, 269–363.

[83] R. Selten, Evolution, Learning, and Economic Behavior, *Games Econ. Behav.* (1991), **3**, 3–24.

[84] L. Shapley, Some Topics in Two–Person Games, *In Advances in Game Theory, Annals of Mathematical Studies* (1964), **5**, 1–28.

[85] J. Swinkels, Evolutionary Stability with Equilibrium Entrants, unpublished paper, Stanford University.

[86] J. Swinkels, Evolution and Strategic Stability: from Maynard Smith to Kohlberg–Mertens, unpublished paper, Stanford University.

[87] J. Swinkels, Adjustment Dynamics and Rational Play in Games, unpublished manuscript, Stanford University, June 1992.

[88] P. Taylor and L. Jonker, Evolutionarily Stable Strategies and Game Dynamics, *Math. Biosci.* (1978), **40**, 145–156.

[89] E. van Damme, Stability and Perfection of Nash Equilibria, Springer–Verlag, Berlin, 1987.

[90] E. van Damme, Equilibrium Selection in 2×2 Games, *Revista Española de Economia,* (1991).

[91] J. van Huyck, R. Battalio and R. Beil, Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure, *American Economic Review*, (1990), **80**, 234-248.

[92] J. van Huyck, R. Battalio and R. Beil, Strategic Uncertainty, Equilibrium Selection Principles and Coordination Failure in Average Opinion Games, *Quarterly Journal of Economics*, (1991), **106**, 885-910.

[93] K. Wärneryd, Cheap Talk, Coordination, and Evolutionary Stability, Stockholm School of Economics, 1990.

[94] H.P. Young, An Evolutionary Model of Bargaining, unpublished paper, University of Maryland, June 1992.

[95] H. Young, The Evolution of Conventions, Working Paper 91–10–043, Santa Fe Institute, 1991.

[96] E. Zeeman, Dynamics of the Evolution of Animal Conflicts, *J. Theor. Biol.* (1981), **89**, 249–270.

Figure 1. Frequency of (b1,a2) (during last 100 plays).

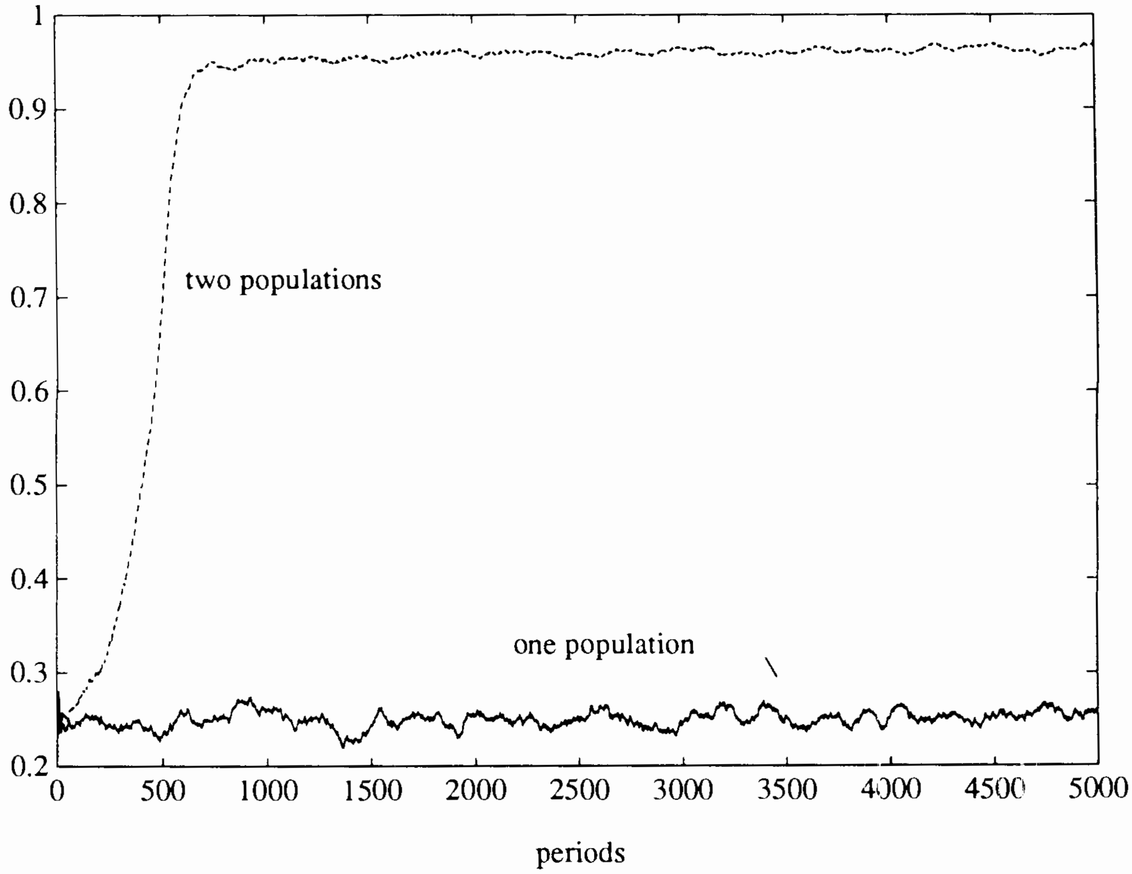


Figure 2. Frequency of play (during last 100 plays).

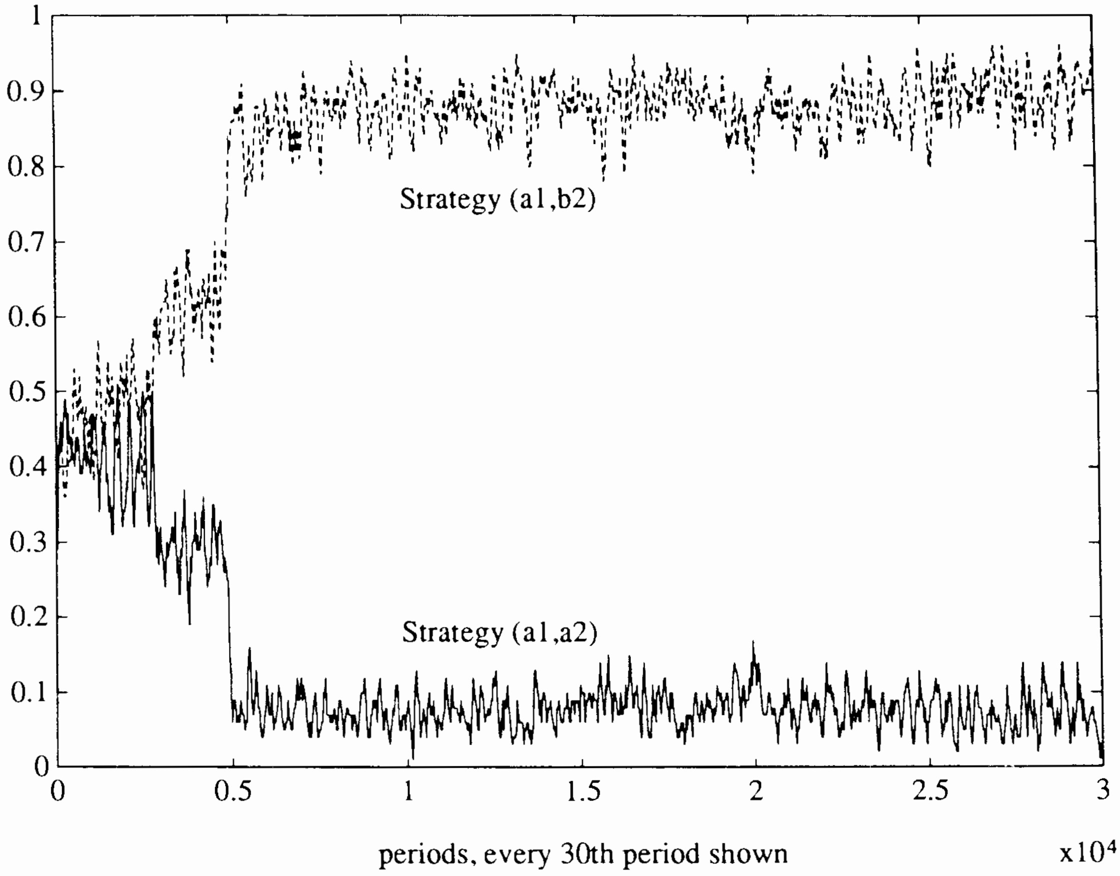


Figure 3. Frequency of play (during last 100 plays).

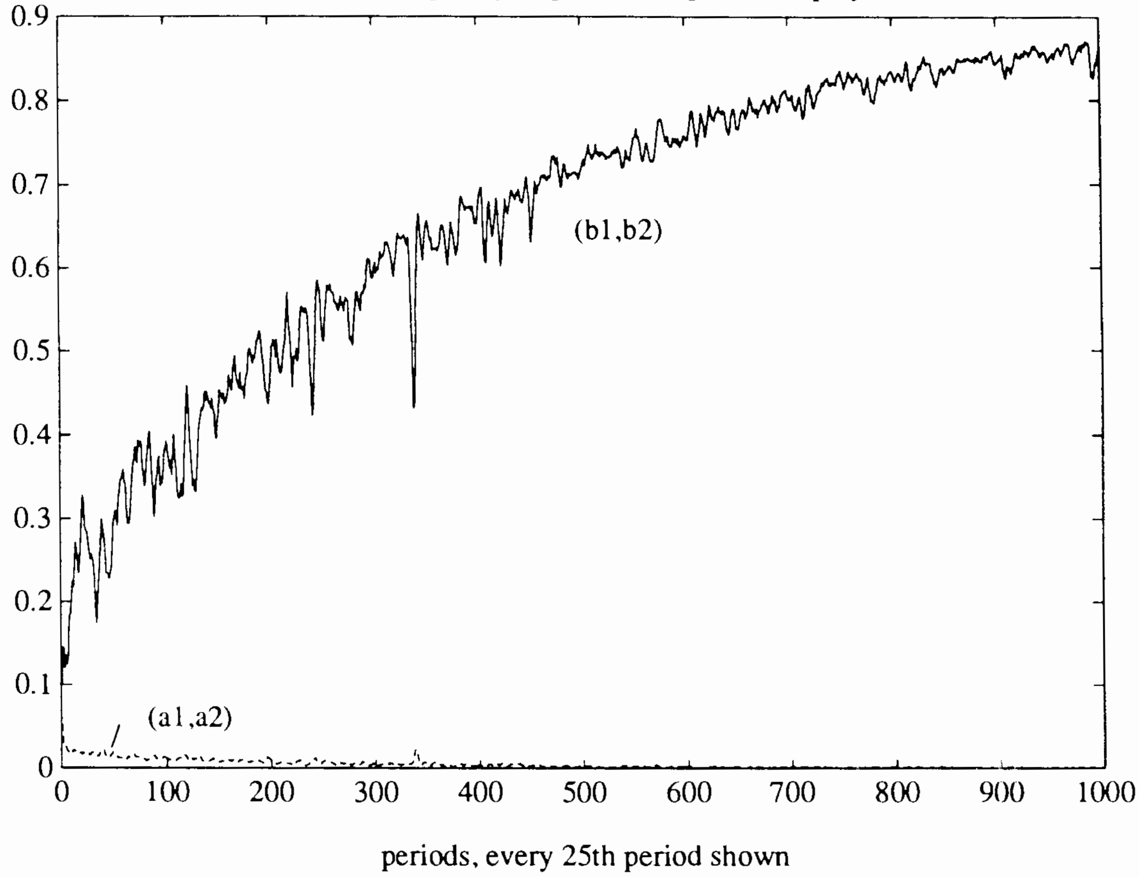


Figure 4a. Frequency of play (during last 100 plays).

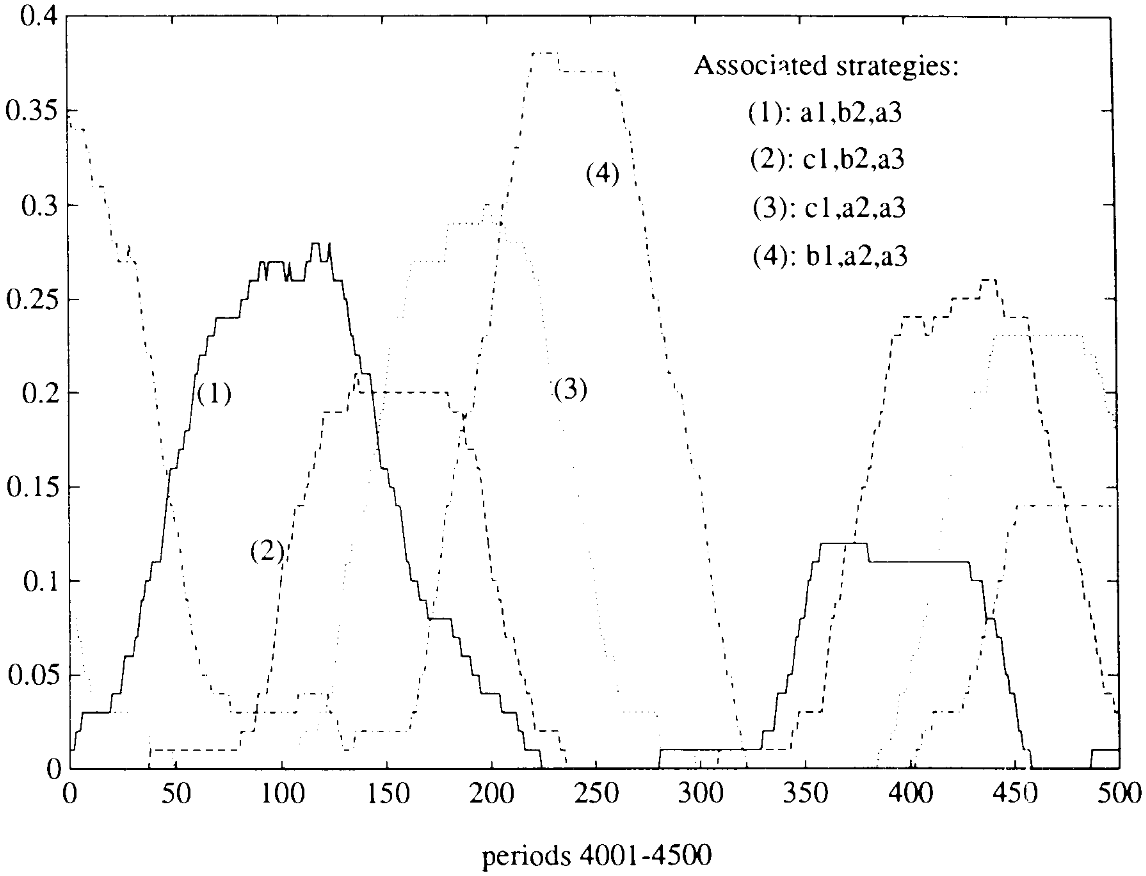


Figure 4b. Frequency of play (during last 100 plays).

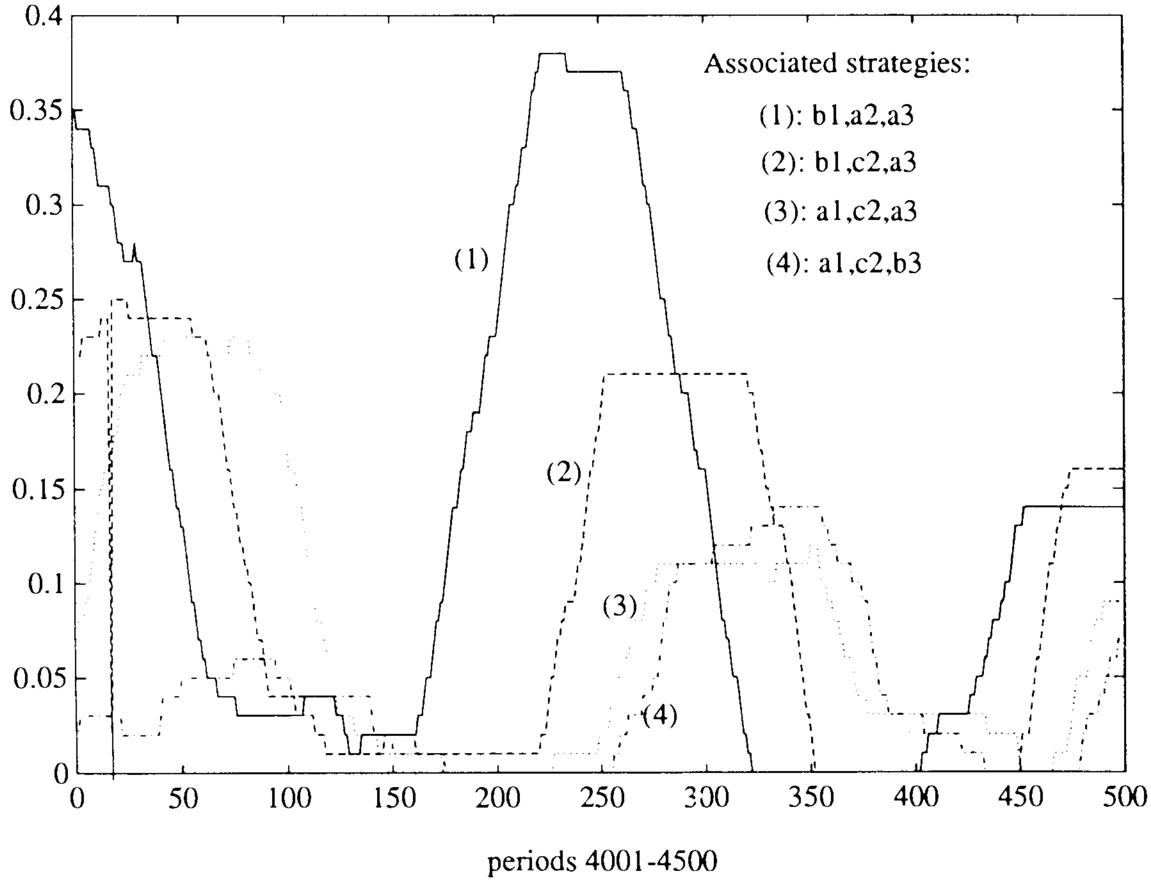


Figure 4c. Frequency of play (during last 100 plays).

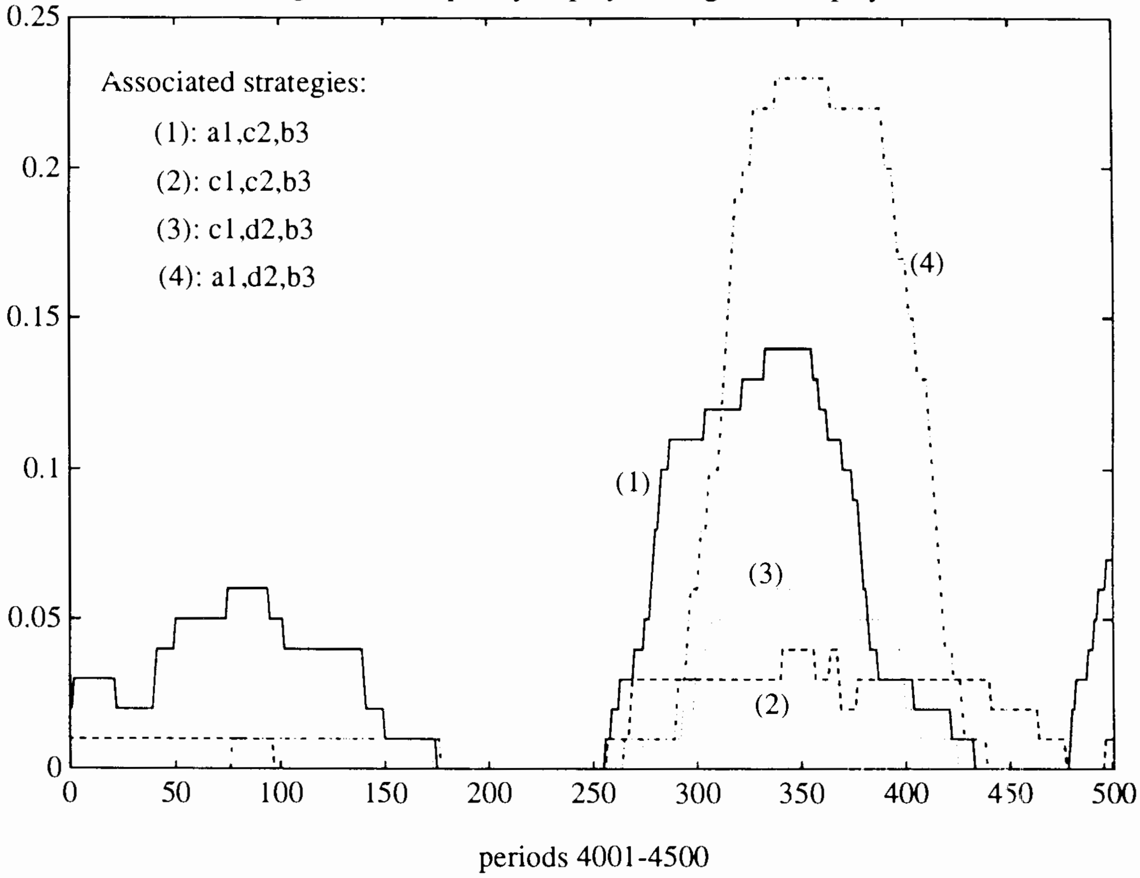
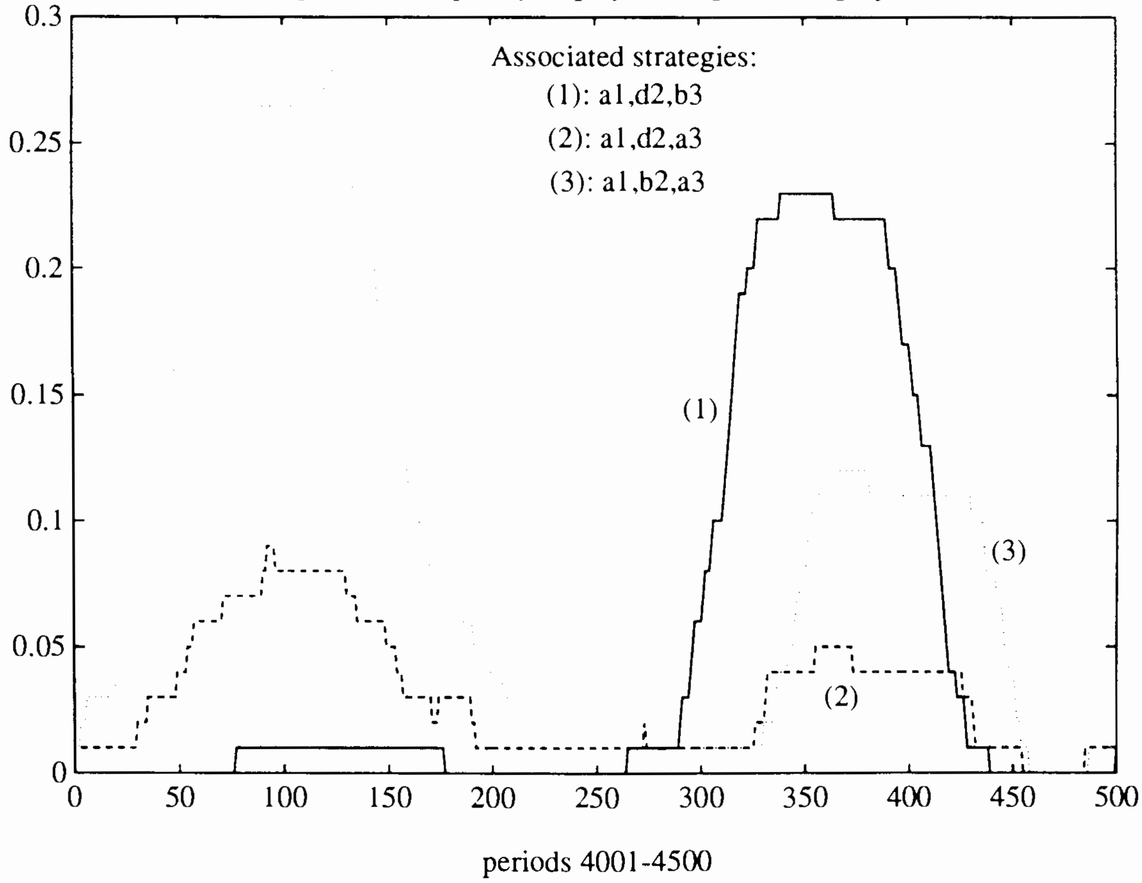


Figure 4d. Frequency of play (during last 100 plays).



RECENT WORKING PAPERS

1. Albert Marcet and Ramon Marimon
Communication, Commitment and Growth. (June 1991)

2. Antoni Bosch
Economies of Scale, Location, Age and Sex Discrimination in Household Demand.(June 1991)

3. Albert Satorra
Asymptotic Robust Inferences in the Analysis of Mean and Covariance Structures. (June 1991)

4. Javier Andrés and Jaume Garcia
Wage Determination in the Spanish Industry. (June 1991)

5. Albert Marcet
Solving Non-Linear Stochastic Models by Parameterizing Expectations: An Application to Asset Pricing with Production. (July 1991)

6. Albert Marcet
Simulation Analysis of Dynamic Stochastic Models: Applications to Theory and Estimation. (November 1991)

7. Xavier Calsamiglia and Alan Kirman
A Unique Informationally Efficient and Decentralized Mechanism with Fair Outcomes. (November 1991)

8. Albert Satorra
The Variance Matrix of Sample Second-order Moments in Multivariate Linear Relations. (January 1992)

9. Teresa Garcia-Milà and Therese J. McGuire
Industrial Mix as a Factor in the Growth and Variability of States'Economies. (Jannuary 1992)

10. Walter Garcia-Fontes and Hugo Hopenhayn
Entry Restrictions and the Determination of Quality. (February 1992)

11. Guillem López and Adam Robert Wagstaff
Indicadores de Eficiencia en el Sector Hopitalario. (March 1992)

12. Daniel Serra and Charles ReVelle
The PQ-Median Problem: Location and Districting of Hierarchical Facilities. Part I (April 1992)

13. Daniel Serra and Charles ReVelle
The PQ-Median Problem: Location and Districting of Hierarchical Facilities. Part II: Heuristic Solution Methods. (April 1992)