Economics Working Paper 23

Adaptive Learning, Evolutionary Dynamics and Equilibrium Selection in Games

Ramon Marimon*

September 1992

Keywords: Learning, Evolution, Equilibrium selection, Robust equilibrium.
Journal of Economic Literature classification: D81, D83, D84.

---

* Universitat Pompeu Fabra and University of Minnesota.

1 Introduction

We discuss three basic properties characterizing adaptive learning: *adaptation, experimentation* and *inertia*. These properties are shared by suitable modifications of most learning models and have their counterparts in evolutionary dynamics (*reproduction, mutation* and *conservation*). Furthermore, in the context of strategic form games, we also characterize the asymptotic behavior of the class of learning algorithms defined by these properties. The asymptotic behavior is characterized by strategy sets which are *robust* to individual perturbations. The corresponding equilibrium concept, that of *robust equilibrium*, provides a characterization of limit points of the learning process. The extension of this concept to a set theoretical formulation allows us to characterize the asymptotic behavior of all sequences of play, not only the convergent ones. This paper follows the theoretical treatment of Marimon (1992) which has been further explored in Marimon and McGrattan (1992).

1.1 From specific algorithms to a more general theory: the problem in defining optimal adaptive learning.

In standard decision theory (*i.e.*, Savage) the set of possible actions and the corresponding payoffs are well defined. Under standard assumptions, optimal decisions exist. Nevertheless, the problem of *finding an optimal action* may not be trivial. The computational *problem* of a rational decision maker can be viewed as a learning problem. If the decision maker has an exhaustive list of all possible actions and their consequences, then rationality implies optimality and the computational *problem* does not exist. Unfortunately, this is seldom the case.

In developing a theory of learning, we would like first, to isolate and characterize basic features defining a broad class of algorithms (Kreps (1990) forcefully makes this point), and second to have a measure of success for such a class of algorithms. This is not an obvious task. A similar problem arises in artificial Intelligence. Consider, for example, an *Expert System* designed for a very specific task. Success can be relatively easily measured. We may, for example: 1) compute the average number of times that the system performs

the task within some acceptance limits; 2) estimate the time and computer capacity required to perform every task, and 4) count the initial starting costs of setting up the system, and its possible maintenance costs.

In the context of not tailored designed algorithms, however, it is more difficult to define appropriate measures of success or to characterize optimal learning. One must take into account not just a specific algorithm but a more general class, not just a specific problem but a more general range of problems. There exist a *trade off* between the generality of the class of algorithms and the class of problems on one hand, and the efficiency requirements imposed on learning rules on the other. Up to which point can we have strict efficiency criteria and generality of rules and problems at the same time? Furthermore, learning algorithms, in contrast with a computational algorithms, have some underlying behavioral assumptions and part of a "measure of success" may be if they capture actual learning by human subjects.

We do not attempt to define an explicit metric that takes all these dimensions into account, but we confine ourselves to a broad class of learning algorithms that has some interesting asymptotic properties within a general class of multi-agent decision problems; *i.e.,* strategic form games. This class includes algorithms with minimal information requirements. We do not consider questions of speed or computational requirements which are linked to specific algorithms. Similarly, we do not consider the optimality of a learning algorithm along the path using, for example, some discounted expected utility criteria.

1.2 Learning by a single agent

Consider first a single agent decision problem. There is a finite set of actions, C, and payoffs are given by $\pi(c,\omega)$, where ω is an exogeneous stochastic shock with distribution p on the finite set Ω. The agent can choose a mixed strategy $\sigma \in \Delta(C)$ with a corresponding expected payoff $E_{(\sigma,p)}\pi(c,\omega)$. If C, p and $\pi(\cdot,\cdot)$ are known, then the maximization problem is trivial. We want to define a class of learning algorithms including algorithms using as initial inputs C (or a enumeration of C) and as recurrent inputs the realized payoffs π_n; where $\pi_n = \pi(c_n,\omega_n)$. Genetic algorithms are of this type. A min-

imal -asymptotic- requirement is that learning algorithm should not choose -infinitely often- dominated actions, *i.e.,* an action $\tilde{c}$ such that, for all ω, and some σ, $\pi(\tilde{c},\omega) < E_{(\sigma)}\pi(c,\omega)$. Of course, a better requirement is that asymptotically only optimal strategies are chosen. That is, if c^* is chosen infinitely often then $c^* \in B(p)$, where $B(\cdot)$ is the *best response* map; *i.e.,*
$B(p)=\{c \in C : E_{(p)}\pi_i(c_i,\omega) \geq$
$E_{(p)}\pi_i(c_i,\omega),\ \forall \tilde{c} \in C\}$.

In this single agent context, an *adaptive rule* defines an action or strategy -possibly a mixed strategy- for every history of actions, shocks and realized payoffs, $\{(\sigma_n,\ c_n,\ \omega_n,\ \pi(c_n,\omega_n));(C,\ \Omega,\ p,\ \pi(\cdot,\cdot))\}$, given the information available to the decison maker at every point in time. For example, if only the set of actions, past strategies and realized payoffs are observed, then a learning algorithm defines an action at period t only using this information, *i.e.,* $\{\{(\sigma_n,\pi_n)\}_{n=0}^{t-1};\ C\}$. $\mathcal{F}_t$ represents the information known by the agent at the beginning of period t. The following two properties characterize *adaptive learning* in single-agent contexts. Furthermore, it can be shown that, when these two properties are satisfied, only optimal strategies are played asymptotically.

Experimentation If the agent does not know the consequences of his actions then it is important that he experiments with all his options *infinitely often*, otherwise he may not be able to make the right inference about the value of his different options. Formally,

- There exists $\{\epsilon_t\}$, $\epsilon_t \in (0,1), \sum_{t=0}^{\infty}\epsilon_t = +\infty$ such that, for every t, $\sigma_t(c) \geq \epsilon_t,\ \forall c \in C$.

Experimentation alone is a powerful element in order to obtain asymptotic results. Consider, for example, that the agent observes the pure actions that he plays and satisfies the *experimentation* hypothesis. The exogeneous stochastic process is *i.i.d.*, (identically, independently distributed), therefore if the agent's beliefs about the consequences of his actions are based on computing the average payoffs obtained from each strategy he will eventually -with probability one- discriminate which strategy has the highest expected payoff. This

result can be derived applying the *Ergodic* theorem (Marimon (1992)). Notice, however, that this only defines how the agent plays asymptotically.

Adaptation This property requires that choices of strategies should move in the direction of the best response map, according to the beliefs based on observed frequencies. It does not assume that the agent observes the exogeneous shock. Formally[2],

- For every t there exists m such that, if

$$\frac{1}{\sum_{n=t}^{n=t+m}\chi_{\bar{c}}(c_n)}\sum_{n=t}^{n=t+m}\chi_{\bar{c}}\pi(c_n,\omega_n) \;<\; \frac{1}{\sum_{n=t}^{n=t+m}\chi_{\hat{c}}(c_n)}\sum_{n=t}^{n=t+m}\chi_{\hat{c}}\pi(c_n,\omega_n),$$

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

$$\text{whenever } \sigma_{t+m}(\bar{c}) \;>\; \epsilon_{t+m}$$

Most learning rules in the literature satisfy these two conditions if they are appropriately modified to include experimentation (see, Marimon and McGrattan (1992)). Therefore, we have characterized a broad class of adaptive learning algorithms which for a relatively large class of individual decision problems has the property that optimal outcomes are chosen asymptotically -with probability one- even when the agent has very limited information. Milgrom and Roberts (1991) have a somewhat weaker condition to characterize adaptive learning. With their definition only dominated actions are eliminated and asymptotic outcomes may not be optimal.

2 Adaptation, experimentation and inertia

In multiagent decision theory, the problem of evaluating a decision or learning procedure is more complex, since an agent's optimal action is a function of the other agents' actions and these may be changing over time. Following the analogy with individual decision theory, one can choose as a minimal requirement that a learning algorithm should not choose -infinitely often- strategies which are not a best response to the strategies played by other players. If this is true for all players, then an iterative process will end up in agents choosing among *rationalizable* strategies. This is the result obtained by Milgrom

and Roberts (1991). Unfortunately, some rationalizable strategies are far from being optimal decisions.

In fact, a rationalizable strategy profile can be a non-optimal response for all players. The well known example of *the battle of the sexes* illustrates this point (see Γ_1, where $x \in (0,1]$). The pair of actions (a_1, b_2) are *rationalizable*, but no player is using a *best response* strategy against the other player's strategy. As in the case of individual decision maker, this requirement is too weak and we can still define a broad class of adaptive learning rules having stronger asymptotic properties. Before we proceed we want to show that, in order to obtain stronger asymptotic results, it is not enough to satisfy the two hypothesis of *experimentation* and *adaptation*. Again, Γ_1, provides a good example.

Γ_1	a_2	b_2
a_1	x,1	0,0
b_1	0,0	1,x

Suppose that both players follow a rule that instructs them to play "best response to the previous play"; *i.e.*, the "Cournot rule". This may result in lack of coordination or cycles. In the battle of the sexes, if 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, Young (1991)). Further, this lack of coordination is robust to perturbations or experimentation 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. If there is no inertia, *i.e.*, where $\rho_t = 1$, the four strategy profiles occur with equal asymptotic probability and both players get an average payoff of $(a+1)/4$. 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 introduces a crucial degree of *stationarity* in the sequence of play. In contrast with the single agent framework, where an agent's action does not change the exogeneous stochastic process, multiagent decisions can create correlations through the play. Even if players act myopically, in the sense that they do not take into account the social effect of their actions, this effect is present and may distort the information gathered through the process of experimentation. As in the *indeterminacy* principle in physics, a player's experimentation can trigger a reaction from other players. 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 a positive constant, γ_i, such that for every t and history of play, up to period t, and for every $c_{-i,t}$

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

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.

Adaptation can be appropriately defined in this context. For example, for player i it is enough to change the exogeneous stochastic process, $\{\omega\}$, by

the endogeneous stochastic process $\{c_{-i,t}\}$ generated by his opponents mixed strategies. What appears as a trivial notational change is however a non trivial mathematical change: from an environment characterized by an *i.i.d.* process we move to one where, in general, the stochastic process is not stationary. Simple ergodic theorems do not apply, but using the additional *inertia* property it is still possible to obtain strong asymptotic results (see Section 5). A very large class of adaptive learning algorithms satisfy the three properties of: *Experimentation, Adaptation and Inertia* (see Marimon and McGrattan (1992) for a description of these algorithms).

3 Reproduction, mutation and conservation

Evolutionary models have a similar structure. A standard evolutionary model consist of I types of agents, corresponding to the I players of the above learning model. Agents are of the same type if they share the same set of actions and payoffs. That is, type i is characterized by (C_i, π_i). In the learning model $\sigma_{i,t}(c_i)$ denotes the probability that player i assigns to strategy c_i at t; in an evolutionary model, $\mu_{i,t}(c_i)$ denotes the fraction of agents of type i playing the pure strategy c_i at t. With this translation we can map learning environments with evolutionary environments; as long as $I > 1$. An additional element defining the evolutionary environment is how a particular agent is matched. More precisely, whether an agent plays against the population distribution or against some random sample, and whether this agent's action may have an effect on the distribution. As long as there is a finite number of agents of each type, there will be a *feed-back* effect similar to the one described above. Agents of type i gain collective experience by being matched and playing against the other types of agents. A finite m sample from the population of type i's opponents, $\mu_{i,t}$, in an evolutionary environment, plays the same role than a m finite realization of a player i's opponents play, $(\sigma_{-i,t-m+1}, \ldots, \sigma_{-i,t})$ in the learning environment; provided that there is enough inertia.

In addition to this map between environments, we can also see that the three basic properties that characterize *adaptive learning* also have their translation in three basic features of evolutionary models. *Experimentation* takes

the form of *mutation* (and *crossover* in genetic algorithms) guaranteeing that every population's type experiments with all its possible alternatives. *Adaptation* is satisfied as long as some form of the Darwinian *replicator* equation describes the evolution of the fraction of agents of a given type using a specific strategy. More precisely, Darwinian dynamics specifies that $\mu_i(\bar{c}_i)$ has a positive growth rate as long as the expected payoff to strategy $\bar{c}_i$ is greater than average. Using specific formulations of this rule, one can see that, in general, the corresponding evolution of μ_t satisfies the *adaptation* hypotesis. Finally, *inertia* takes different forms that we can label *conservation*. Nature, as human learners, can overreact, but some degree of conservation is needed to evalute which characteristics have higher than average payoff. Evolutionary models capture this feature by, for example, having only reproduction of a fraction of the population at any point in time.

It should be noticed that a fairly general feature of learning and evolutionary models is their *path dependence*. The specific experience of players and species can have a long lasting effect. Nevertheless, it is common in the literature on evolution to postulate *stationary markovian* models (see, for example, Kandori *et al.* (1992) and Young (1991)). This, in general, is achieved by imposing strong assumptions on how agents are matched, and on the form that *experimentation* and *mutation* takes place. For example, constant experimentation rates and constant inertia allow for more stationarity in the model. On the other hand, they may preclude convergence.

4 Robust equilibria and asymptotic behavior

In this Section we briefly describe the type of asymptotic results that are achieved in multi-agent problems when our behavioral assumptions are satisfied. There are two main types of results: *i)* to characterize the limit point of converging sequences of play, and *ii)* to characterize the set of outcomes that are asymptotically attained. Previously known versions of these two statements are: *i')* If $\sigma_t \to \sigma^*$ then σ^* is a *Nash equilibrium*, and *ii')* Asymptotically, only rationalizable strategies are played. *i')* is a *Folk Theorem* in the learning literature, *ii')* is due to Milgrom and Roberts(1991). Marimon (1992)

obtains further results. The following concept of *robust equilibrium* provides a characterization of limit points.

Definition The strategy profile σ^* is a **robust equilibrium** if, for every player i, there exists an open set $\mathcal{N}_i$ of perturbations, such that for every $\tilde{\sigma}_i \in \mathcal{N}_i$, and for every player $j \neq i$, $\sigma_j^* \in B_j(\tilde{\sigma}_i, \sigma_{-ij}^*)$.

That is, a *robust equilibrium* requires that player's best replies remain best replies even when single players perturb their strategies. In a two players game a pure strategy perfect equilibrium is a robust equilibrium, while games with a unique mixed equilibrium (*i.e.*, perfect, by definition), such as the *matching pennies* game, do not have a robust equilibrium. However, robust equilibrium is not a refinement of perfect equilibrium since in games with more than two players there may be robust equilibria which are not perfect (see Marimon and McGrattan (1992)). The concept of *robust equilibrium* can be extended to a set theoretical concept providing a more general solution concept. The basic idea is that the conditional best response map should be robust to individual perturbations. In other words, given a random mechanism that correlates players' actions, it is required that incentive constraints are satisfied with strict inequality within a given set of actions. Minimal sets with this property are called *robust-recurrent* sets. For example, in the *battle of the sexes* game, Γ_1, the only robust-recurrent sets are the two pure strategy Nash equilibrium and in the *matching pennies* game the set of all the pure strategies defines a robust recurrent set.

These robustness concepts provide a characterization of the asymptotic behavior of the class of adaptive learning algorithms (or evolutionary algorithms) that satisfies the above properties of *adaptation, experimentation* and *inertia* (*reproduction, mutation* and *conservation*). In particular we have: *i") if play converges to a strategy profile, then the strategy profile is* a robust equilibrium.; ii") *asymptotically, play converges (with probability one) to a set of strategy profiles which is a* robust-recurrent set, *and* iii) 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 results are obtained in Marimon (1992) and further discussed in Marimon and McGrattan (1992).

References

Kandori, M., G.J. Mailath and R. Rob, , 1992, Learning, mutation and long run equilibria in games, CARESS Working Paper #91–01R, University of Pennsylvania.

Kreps, David, 1990, Game theory and economic modeling (Clarendon Press, Oxford).

Marimon, Ramon, 1992, Correlated rationalizability and adaptive learning, unpublished paper, Universitat Pompeu Fabra, Barcelona.

Marimon, Ramon and Ellen McGrattan, 1992, On adaptive learning in strategic games, in A. Kirman and M. Salmon *eds.*, Learning and rationality in economics (Basil Blackwell (forthcoming)).

Milgrom, Paul and John Roberts, 1991, Adaptive and sophisticated learning in normal form games, Games and Economic Behavior 3, 82–100.

Young, H. P., 1991, The evolution of conventions, working waper 91–10–043, Santa Fe Institute.

Endnotes

[1] This paper is part of a research project with Ellen McGrattan and some of the ideas here presented are more fully developed in Marimon and McGrattan (1992). I also want to thank Giorgia Giovannetti for her comments and the NSF for financial support.

[2] The indicator function of action $\bar{c}$ is denoted $\chi_{\bar{c}}(\cdot)$.

[3] We use standard notation for games: I is the set of players; C_i is the finite set of pure strategies for player i; $\sigma_{i,t}$ denotes the mixed strategy of plyer i at t, and $\sigma_{-i,t}$ the profile of mixed strategies of the opponents of i.

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)

14. Juan Pablo Nicolini
Ruling out Speculative Hyperinflations: a Game Theoretic Approach. (April 1992)

15. Albert Marcet and Thomas J. Sargent
Speed of Convergence of Recursive Least Squares Learning with ARMA Perceptions. (May 1992)

16. Albert Satorra
Multi-Sample Analysis of Moment-Structures: Asymptotic Validity of Inferences Based on Second-Order Moments. (June 1992)

Special issue Vernon L. Smith
Experimental Methods in Economics. (June 1992)

17. Albert Marcet and David A. Marshall
Convergence of Approximate Model Solutions to Rational Expectation Equilibria Using the Method of Parameterized Expectations.

18. M. Antònia Monés, Rafael Salas and Eva Ventura
Consumption, Real after Tax Interest Rates and Income Innovations. A Panel Data Analysis. (December 1992)

19. Hugo A. Hopenhayn and Ingrid M. Werner
Information, Liquidity and Asset Trading in a Random Matching Game. (February 1993)

20. Daniel Serra
The Coherent Covering Location Problem. (February 1993)

21. Ramon Marimon, Stephen E. Spear and Shyam Sunder
Expectationally-driven Market Volatility: An Experimental Study. (March 1993)

22. Giorgia Giovannetti, Albert Marcet and Ramon Marimon
Growth, Capital Flows and Enforcement Constaints: The Case of Africa. (March 1993)

23. Ramon Marimon
Adaptive Learning, Evolutionary Dynamics and Equilibrium Selection in Games. (March 1993)