Proceedings of Symposia in Applied Mathematics
Volume **69**, 2011

On Some Global and Unilateral Adaptive Dynamics

Sylvain Sorin

Abstract. The purpose of this chapter is to present some adaptive dynamics arising in strategic interactive situations. We will deal with discrete time and continuous time procedures and compare their asymptotical properties. We will also consider global or unilateral frameworks and describe the wide range of applications covered by this approach. The study starts with the discrete time fictitious play procedure and its continuous time counterpart which is the best reply dynamics. Its smooth unilateral version presents interesting consistency properties. We then analyze its connection with the time average replicator dynamics. Several results rely on the theory of stochastic approximation and basic tools are briefly presented in a last section.

1. Fictitious Play and Best Reply Dynamics

Fictitious play is one of the oldest and most famous dynamical processes introduced in game theory. It has been widely studied and is a good introduction to the field of adaptive dynamics. This procedure is due to Brown (1949, 1951) and corresponds to an interactive adjustment process with (increasing and unbounded) memory.

1.1. Discrete fictitious play.

Consider a game in strategic form with a finite set of players $i \in I$, each having a finite pure strategy set S^i. For each $i \in I$, the mixed strategy set $X^i = \Delta(S^i)$ corresponds to the simplex on S^i. $F^i : S = \prod_{j \in I} S^j \to \mathbb{R}$ is the payoff of player i and we define $F^i(y) = \mathsf{E}_y F(s)$ for every $y \in \Delta(S)$, where E stands for the expectation.
The game is played repeatedly in discrete time. Given an n-stage history, which is the sequence of profiles of past moves of the players, $h_n = (x_1 = \{x_1^i\}_{i=1,\dots,I}, x_2, \dots, x_n) \in S^n$, the fictitious play procedure requires the move x_{n+1}^i of each player i at stage $n+1$ to be a best reply to the "time average moves" of her opponents.
There are two variants, that coincide in the case of two-player games :
- **independent FP**: for each i, let

$$\overline{x}_n^i = \frac{1}{n} \sum\nolimits_{m=1}^{n} x_m^i$$

1991 *Mathematics Subject Classification.* Primary 91A22.
I want to thank Bill Sandholm for useful comments.
This research was partially supported by grant ANR-08-BLAN-0294-01 (France).

and $\overline{x}_n^{-i} = \{\overline{x}_n^j\}_{j\neq i}$. Player i computes, at each stage n and for each of her opponents $j \in I$, the empirical distribution of her past moves and considers the product distribution. Then, her next move at stage $n+1$ satisfies:

$$x_{n+1}^i \in BR^i(\overline{x}_n^{-i}) \tag{1.1}$$

where BR^i denotes the best reply correspondence of player i, from $\Delta(S^{-i})$ to X^i, with $S^{-i} = \prod_{j\neq i} S^i$: $BR^i(y^{-i}) = \{x^i \in X^i; F^i(x^i, y^{-i}) = \max_{z^i\in X^i} F^i(z^i, y^{-i})\}$.
- **correlated FP**: one defines a point $\tilde{x}_n^{-i}$ in $\Delta(S^{-i})$ by :

$$\tilde{x}_n^{-i} = \frac{1}{n}\sum\nolimits_{m=1}^{n} x_m^{-i}$$

which is the empirical distribution of the joint moves of the opponents $-i$ of player i. Here the discrete time process satisfies:

$$x_{n+1}^i \in BR^i(\tilde{x}_n^{-i}). \tag{1.2}$$

Since one deals with time averages one has

$$\overline{x}_{n+1}^i = \frac{n\overline{x}_n^i + x_{n+1}^i}{n+1}$$

hence the stage difference is expressed as

$$\overline{x}_{n+1}^i - \overline{x}_n^i = \frac{x_{n+1}^i - \overline{x}_n^i}{n+1}$$

so that (1.1) can also be written as :

$$\overline{x}_{n+1}^i - \overline{x}_n^i \in \frac{1}{(n+1)}[BR^i(\overline{x}_n^{-i}) - \overline{x}_n^i]. \tag{1.3}$$

Definition. A sequence $\{x_n\}$ of moves in S satisfies **discrete fictitious play (DFP)** if (1.3) holds.
Remarks.

x_n^i does not appear explicitly any more in (1.3): the natural state variable of the process is $\overline{x}_n$ which is the product of the marginal empirical averages $\overline{x}_n^j \in X^j$.

One can define a procedure based, for each player, on her past vector payoffs $g_n^i = \{F^i(s^i, x_n^{-i})\}_{s^i\in S^i} \in \mathbb{R}^{S^i}$, rather than on the past moves of all players, as follows: $x_{n+1}^i \in \mathbf{br}^i(\bar{g}_n^i)$ with $\mathbf{br}^i(U) = \mathrm{argmax}_{X^i}\langle x, U\rangle$ and $\bar{g}_n^i = \frac{1}{n}\sum_{m=1}^n g_m^i$. Due to the linearity of the payoffs, this corresponds to the correlated fictitious play procedure. Note that $\overline{x}_n$ is no longer the common state variable but rather the correlated empirical distribution of moves $\tilde{x}_n$ which satisfies:

$$\tilde{x}_{n+1} = \frac{n\tilde{x}_n + x_{n+1}}{n+1}$$

and has the same marginal on each factor space X^i. The joint process (1.2) is defined by:

$$\tilde{x}_{n+1} - \tilde{x}_n \in \frac{1}{(n+1)}[\prod_i BR^i(\tilde{x}_n) - \tilde{x}_n]. \tag{1.4}$$

1.2. Continuous fictitious play and best reply dynamics.

The continuous time (formal) counterpart of the above difference inclusion (1.3) is the differential inclusion, called **continuous fictitious play (CFP)**:

$$\dot{X}^i_t \in \frac{1}{t}[BR^i(X^{-i}_t) - X^i_t]. \tag{1.5}$$

The change of time $Z_s = X_{e^s}$ leads to

$$\dot{Z}^i_s \in [BR^i(Z^{-i}_s) - Z^i_s] \tag{1.6}$$

which is the **(continuous time) best reply dynamics (CBR)** introduced by Gilboa and Matsui (1991), see Section 12 in K. Sigmund's chapter.
Note that the asymptotic properties of (CFP) or (CBR) are the same, since the differential inclusions differ only by their time scales.
The interpretation of (CBR) in evolutionary game theory is as follows: at each stage n a randomly selected fraction ε of the current population Z_n dies and is replaced by newborns Y_{n+1} selected according to their abilities to adjust to the current population. The discrete time process is thus

$$Z_{n+1} = \varepsilon Y_{n+1} + (1-\varepsilon)Z_n$$

with $Y_{n+1} \in BR(Z_n)$ leading to the difference inclusion

$$Z_{n+1} - Z_n \in \varepsilon[BR(Z_n) - Z_n].$$

Note that it is delicate in his framework to justify the fact that the step size ε (which is induced by the choice of the time unit) should go to 0. However numerous asymptotic results are available for small step sizes.

Comments. Recall that a solution of a differential inclusion of the form

$$\dot{z}_t \in \Psi(z_t) \tag{1.7}$$

where Ψ is a correspondence defined on a subset of $\mathbb{R}^n$ with values in $\mathbb{R}^n$, is an absolutely continuous function z from $\mathbb{R}$ to $\mathbb{R}^n$ that satisfies (1.7) almost everywhere. Let Z be a compact convex subset of $\mathbb{R}^n$ and $\Phi : Z \rightrightarrows Z$ a correspondence from Z to itself, upper semi continuous and with non empty convex values. Consider the differential inclusion

$$\dot{z}_t \in \Phi(z_t) - z_t. \tag{1.8}$$

LEMMA 1.1. *For every $z(0) \in Z$, (1.8) has a solution with $z_t \in Z$ and $z_0 = z(0)$.*

See e.g. Aubin and Cellina (1984).
In particular this applies to (CBR) where $Z = \prod X^i$ is the product of the sets of mixed strategies.
Note also that rest points of (1.8) coincide with fixed points of Φ.

1.3. General properties.

We recall briefly here basic properties of (DFP) or (CFP), in particular the link to Nash equilibrium.
Definition. A process z_n (discrete) or z_t (continuous) converges to a subset Z of some metric space if $d(z_n, Z)$ or $d(z_t, Z)$ goes to 0 as n or $t \to \infty$.

PROPOSITION 1.1. *If (DFP) or (CFP) converges to a point x, x is a Nash equilibrium.*

PROOF. If x is not a Nash equilibrium then $d(x, BR(x)) = \delta > 0$. Hence by upper semicontinuity of the best reply correspondence $d(y, BR(z)) \geq \delta/2 > 0$ for each y and z in a neighborhood of x which prevents convergence of the discrete time or continuous time processes. □

The dual property is clear:

PROPOSITION 1.2. *If x is a Nash equilibrium, it is a rest point of (CFP).*

Comments.
(i) (DFP) is "predictable": in the game with payoffs

$\sqrt{2}$	0
0	1

if player 1 follows (DFP) her move is always pure, since the past frequency of Left, say y, is a rational number so that $y\sqrt{2} = 1 - y$ is impossible; hence player 1 is guaranteed only 0. It follows that the unilateral (DFP) process has bad properties, see Section 2.
(ii) Note also the difference between convergence of the marginal distribution and convergence of the product distribution of the moves and in particular the consequences in terms of payoffs. In the next game

	L	R
T	1	0
B	0	1

a sequence of $TR, BL, TR, \ldots$ induces asymptotical average marginal distributions $(1/2, 1/2)$ for both players (hence optimal strategies) but the average payoff is 0 while an alternative sequence $TL, BR, \ldots$ would have the same average marginal distributions and payoff 1.

We analyze now (DFP) and (CFP) in some classes of games. We will deduce properties of the initial discrete time process from the analysis of the continuous time counterpart.

1.4. Zero-sum games.

This is the framework in which (DFP) was initialy introduced in order to generate optimal strategies. The continuous time model is mathematically easier to analyze.

1.4.1. *Continuous time.*
We first consider the finite case.
1) *Finite case* : Harris (1998); Hofbauer (1995); Hofbauer and Sandholm (2009).
The game is defined by a bilinear map $F = F^1 = -F^2$ on a product of simplexes $X \times Y$.
Introduce $a(y) = \max_{x \in X} F(x, y)$ and $b(x) = \min_{y \in Y} F(x, y)$ that correspond to the best reply levels, then the duality gap at (x, y) is $W(x, y) = a(y) - b(x) \geq 0$. Moreover (x^*, y^*) belongs to the set of optimal strategies, $X_F \times Y_F$, iff $W(x^*, y^*) = 0$, see Section 3 of K. Sigmund's chapter. Consider the evaluation of the duality gap $W(x_t, y_t)$ along a trajectory of (1.5).

PROPOSITION 1.3. *The "duality gap" criteria converges to 0 at a speed of $1/t$ in (CFP).*

PROOF. Let (x_t, y_t) be a solution of (CBR) (1.6) and introduce

$$\alpha_t = x_t + \dot{x}_t \in BR^1(y_t)$$
$$\beta_t = y_t + \dot{y}_t \in BR^2(x_t).$$

The duality gap along the trajectory is given by $w_t = W(x_t, y_t)$. Note that $a(y_t) = F(\alpha_t, y_t)$ hence taking derivative with respect to the time

$$\frac{d}{dt}a(y_t) = D_1F(\alpha_t, y_t)\dot{\alpha}_t + D_2F(\alpha_t, y_t)\dot{y}_t$$

but the first term is 0 (envelope theorem). As for the second one

$$D_2F(\alpha_t, y_t)\dot{y}_t = F(\alpha_t, \dot{y}_t)$$

by linearity. Thus:

$$\dot{w}_t = F(\alpha_t, \dot{y}_t) - F(\dot{x}_t, \beta_t) = F(x_t, \dot{y}_t) - F(\dot{x}_t, y_t)$$
$$= F(x_t, \beta_t) - F(\alpha_t, y_t) = b(x_t) - a(y_t) = -w_t.$$

It follows that exponential convergence holds for (CBR)

$$w_t = e^{-t}w_0$$

hence convergence at a rate $1/t$ in the original (CFP). □

This proof in particular implies the minmax theorem and is reminiscent of the analysis due to Brown and von Neumann (1950).

The analysis extends to the framework of continuous strategy space as follows.
2) *Saddle case* : Hofbauer and Sorin (2006)
Define the condition (H) : F is a continuous, concave/convex real function defined on a product $X \times Y$ of two compact convex subsets of an euclidean space.

PROPOSITION 1.4. *Under (H), any solution w_t of (CBR) satisfies*

$$\dot{w}_t \le -w_t \ a.e.$$

The proof, while much more involved, is in the spirit of Proposition 1.3 and the main application is (see Section 5 for the definitions):

COROLLARY 1.1. *For (CBR)*
i) $X_F \times Y_F$ is a global attractor .
ii) $X_F \times Y_F$ is a maximal invariant subset.

PROOF. From the previous Proposition 1.4 one deduces the following property: $\forall \varepsilon > 0$, $\exists T$ such that for all (x_0, y_0), $t \ge T$ implies

$$w_t \le \varepsilon$$

hence in particular the value v_F of the game F exists and for $t \ge T$

$$b(x_t) \ge v_F - \varepsilon.$$

Continuity of F (and hence of the function b) and compactness of X imply that for any $\delta > 0$, there exists T' such that $d(x_t, X_F) \le \delta$ as soon as $t \ge T'$. This shows that $X_F \times Y_F$ is a global attractor.
Now consider any invariant trajectory. By Proposition 1.4 at each point w one can write, for any t, $w = w_t \le e^{-t}w_0$, but the duality gap w_0 is bounded, hence w equal to 0 which gives ii). □

To deduce properties of the discrete time process we introduce a general procedure.

1.4.2. *Discrete deterministic approximation.*
Consider again the framework of (1.8).
Let α_n a sequence of positive real numbers with $\sum \alpha_n = +\infty$.
Given $a_0 \in Z$, define inductively a_n through the following difference inclusion:

$$a_{n+1} - a_n \in \alpha_{n+1}[\Phi(a_n) - a_n]. \tag{1.9}$$

The interpretation is that the evolution of the process satisfies $a_{n+1} = \alpha_{n+1}\tilde{a}_{n+1} + (1-\alpha_{n+1})a_n$ with some $\tilde{a}_{n+1} \in \Phi(a_n)$, and where α_{n+1} is the step size at stage $n+1$.

Definition. A sequence $\{a_n\} \in Z$ following (1.9) is a **discrete deterministic approximation** (DDA) of (1.8).
The associated continuous time trajectory $\mathbf{A} : \mathbb{R}^+ \to Z$ is constructed in two stages. First define inductively a sequence of times $\{\tau_n\}$ by: $\tau_0 = 0$, $\tau_{n+1} = \tau_n + \alpha_{n+1}$; then let $A_{\tau_n} = a_n$ and extend the trajectory by linear interpolation on each interval $[\tau_n, \tau_{n+1}]$:

$$A_t = a_n + \frac{(t - \tau_n)}{(\tau_{n+1} - \tau_n)}(a_{n+1} - a_n).$$

Since $\sum \alpha_n = +\infty$ the trajectory is defined on $\mathbb{R}^+$.

To compare $\mathbf{A}$ to a solution of (1.8) we will need the approximation property corresponding to the next proposition: it states that two differential inclusions defined by correspondences having graphs close one to the other will also have sets of solutions close one to each other, on a given compact time interval.

Notations. Let $\mathcal{A}(\Phi, T, z) = \{\mathbf{z}; \mathbf{z}$ is a solution of (1.8) on $[0, T]$ with $z_0 = z\}$, $D_T(\mathbf{y}, \mathbf{z}) = \sup_{0 \le t \le T} \|y_t - z_t\|$. G_Φ is the graph of Φ and G_Φ^ε is an ε-neighborhood of G_Φ.

PROPOSITION 1.5. $\forall T \ge 0, \forall \varepsilon > 0, \exists \delta > 0$ *such that*

$$\inf\{D_T(\mathbf{y}, \mathbf{z}); \mathbf{z} \in \mathcal{A}(\Phi, T, z)\} \le \varepsilon$$

for any solution $\mathbf{y}$ *of*

$$\dot{y}_t \in \tilde{\Phi}(y_t) - y_t$$

with $y_0 = z$ *and* $d(G_\Phi, G_{\tilde{\Phi}}) \le \delta$.

See e.g. Aubin and Cellina (1984), Chapter 2.

Let us now compare the two dynamics defined by $\{a_n\}$ and $\mathbf{A}$.

Case 1 Assume α_n decreasing to 0.
In this case the set $L(\{a_n\})$ of accumulation points of the sequence $\{a_n\}$ coincides with the limit set of the trajectory: $L(\mathbf{A}) = \cap_{t \ge 0} \overline{A_{[t,+\infty)}}$.

PROPOSITION 1.6.
i) If Z_0 is a global attractor for (1.8), it is also a global attractor for (1.9).
ii) If Z_0 is a maximal invariant subset for (1.8), then $L(\{a_n\}) \subset Z_0$.

PROOF. i) Given $\varepsilon > 0$, let T_1 be such that any trajectory $\mathbf{z}$ of (1.8) is within ε of Z_0 after time T_1. Given T_1 and ε, let $\delta > 0$ be defined by Proposition 1.5. Since α_n decreases to 0, given $\delta > 0$, for $n \ge N$ large enough for a_n, hence $t \ge T_2$ large enough for A_t, one can write :

$$\dot{A}_t \in \Psi(A_t) \quad \text{with} \quad G_\Psi \subset G^\delta_{\Phi - Id}.$$

Consider now A_t for some $t \geq T_1 + T_2$. Starting from any position A_{t-T_1} the continuous time process $\mathbf{z}$ defined by (1.8) approaches within ε of Z_0 at time t. Since $t - T_1 \geq T_2$, the interpolated process A_s remains within ε of the former z_s on the interval $[t - T_1, t]$, hence is within 2ε of Z_0 at time t. In particular this shows: $\forall \varepsilon, \exists N_0$ such that $n \geq N_0$ implies

$$d(a_n, Z_0) \leq 2\varepsilon.$$

ii) The result follows from the fact that $L(\mathbf{A})$ is invariant.
In fact consider $a \in L(\mathbf{A})$, hence let $t_n \to +\infty$ and $A_{t_n} \to a$. Given $T > 0$ let $\mathbf{B}^n$ denote the translated solution A_{t-t_n} defined on $[t_n - T, t_n + T]$. The sequence $\{\mathbf{B}^n\}$ of trajectories is equicontinuous and has an accumulation point $\mathbf{B}$ satisfying $B_0 = a$ and B_t is a solution of (1.8) on $[-T, +T]$. This being true for any T the result follows. □

Case 2 α_n small not vanishing.

Proposition 1.7. *If Z_0 is a global attractor for (1.8), then for any $\varepsilon > 0$ there exists α such that if $\limsup_{n\to\infty} \alpha_n \leq \alpha$, there exists N with $d(a_n, Z_0) \leq \varepsilon$ for $n \geq N$. Hence a neighborhood of Z_0 is still a global attractor for (1.9).*

Proof. The proof of Proposition 1.6 implies easily the result. □

We are now in position to study the initial discrete time fictitious play procedure.

1.4.3. *Discrete time.*
Recall that $X_F \times Y_F$ denote the product of the sets of optimal strategies in the zero-sum game with payoff F.

Proposition 1.8. *(DFP) converges to $X_F \times Y_F$ in the continuous saddle zero-sum case.*

Proof. The result follows from 1) the properties of the continuous time process, Corollary 1.1, 2) the approximation result, Proposition 1.6 and 3) the fact that the discrete time process (DFP) is a DDA of the continuous time one (CFP). □

The initial convergence result in the finite case is due to Robinson (1951). Her proof is quite involved and explicitly uses the finiteness of the strategy sets.
In this framework one has also the next result on the payoffs which is not implied by the convergence of the marginal empirical plays. In fact the distribution of the moves at each stage need not converge.

Proposition 1.9. *(Rivière, 1997)*
The average of the realized payoffs along (DFP) converges to the value in the finite zero-sum case.

Proof. Write $X = \Delta(I), Y = \Delta(J)$ and let $U_n = \sum_{p=1}^{n} F(., j_p)$ be the sum of the columns played by player 2. Consider the sum of the realized payoffs

$$R_n = \sum_{p=1}^{n} F(i_p, j_p) = \sum_{p=1}^{n} (U_p^{i_p} - U_{p-1}^{i_p})$$

Thus

$$R_n = \sum_{p=1}^{n} U_p^{i_p} - \sum_{p=1}^{n-1} U_p^{i_{p+1}} = U_n^{i_n} + \sum_{p=1}^{n-1} (U_p^{i_p} - U_p^{i_{p+1}})$$

but the fictitious property implies, since i_{p+1} is a best reply to $\bar{U}_p$, that

$$U_p^{i_p} - U_p^{i_{p+1}} \leq 0.$$

Thus $\limsup \frac{R_n}{n} \leq \limsup \max_i \frac{U_n^i}{n} \leq v$ by the previous Proposition 1.8 and the dual property implies the result. □

To summarize, in zero sum games the average empirical marginal distribution of moves are close to optimal strategies and the average payoff close to the value when the number of repetitions is large enough and both players follow (DFP).

We turn now to general I player games.

1.5. Potential games.

For a general presentation of this class, see the chapter by W. Sandholm. Since we are dealing with best-reply based processes, we can assume that the players share the same payoff function.
Hence the game is defined by a continuous payoff function F from X to $\mathbb{R}$ where each $X^i, i \in I$ is a compact convex subset of an euclidean space. Let $NE(F)$ be the set of Nash equilibria of the game defined by F.

1.5.1. *Discrete time.*
We study here the finite case and we follow Monderer and Shapley (1996).
Recall that x_n converges to $NE(F)$ if $d(x_n, NE(F))$ goes to 0. Since F is continuous and X is compact, an equivalent property is to require that for any $\varepsilon > 0$, for any n large enough x_n is an ε-equilibrium in the sense that:

$$F(x_n) + \varepsilon \geq F(x^i, x_n^{-i})$$

for all $x^i \in X^i$ and all $i \in I$.

Proposition 1.10. *(DFP) converges to $NE(F)$.*

Proof. Since F is multilinear and bounded, one has:

$$F(\bar{x}_{n+1}) - F(\bar{x}_n) = F(\bar{x}_n + \frac{1}{n+1}(x_{n+1} - \bar{x}_n)) - F(\bar{x}_n)$$

hence, by a Taylor approximation

$$F(\bar{x}_{n+1}) - F(\bar{x}_n) \geq \sum_i \frac{1}{n+1}[F(x_{n+1}^i, \bar{x}_n^{-i}) - F(\bar{x}_n)] - \frac{K_1}{(n+1)^2}$$

for some constant K_1 independent of n. Let $a_{n+1} = \sum_i [F(x_{n+1}^i, \bar{x}_n^{-i}) - F(\bar{x}_n)]$, which is ≥ 0 by definition of (DFP). Adding the previous inequality implies

$$F(\bar{x}_{n+1}) \geq \sum_{m=1}^{n+1} \frac{a_m}{m} - K_2$$

for some constant K_2. Since $a_m \geq 0$ and F is bounded, $\sum_{m=1}^{n+1} \frac{a_m}{m}$ converges. This property in turn implies

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n \leq N} a_n = 0, \tag{1.10}$$

Now a consequence of (1.10) is that, for any $\varepsilon > 0$,

$$\frac{\#\{n \leq N; \bar{x}_n \notin NE^\varepsilon(F)\}}{N} \to 0, \qquad \text{as } N \to \infty. \tag{1.11}$$

In fact, there exists $\delta > 0$ such that $\bar{x}_n \notin NE^{\varepsilon}(F)$ forces $a_{n+1} \geq \delta$. Inequality (1.11) in turns implies that $\bar{x}_n$ belongs to $NE^{2\varepsilon}(F)$ for n large enough. Otherwise $\bar{x}_m \notin NE^{\varepsilon}(F)$ for all m in a neighborhood of n of non negligible relative size of the order $O(\varepsilon)$. (This is a general property of Cesaro mean of Cesaro means). □

1.5.2. *Continuous time.*
The finite case was studied in Harris (1998), the compact case in Benaim, Hofbauer and Sorin (2005).
Let (H') be the following hypothesis: F is defined on a product X of compact convex subsets X^i of a euclidean space, $\mathcal{C}^1$ and concave in each variable.

PROPOSITION 1.11. *Under (H'), (CBR) converges to NE(F).*

PROOF. Let $W(x) = \sum_i [G^i(x) - F(x)]$ where $G^i(x) = \max_{s \in X^i} F(s, x^{-i})$. Thus x is a Nash equilibrium iff $W(x) = 0$. Let x_t be a solution of (CBR) and consider $f_t = F(x_t)$. Then $\dot{f}_t = \sum_i D_i F(x_t)\dot{x}_t^i$. By concavity one obtains:

$$F(x_t^i, x_t^{-i}) + D_i F(x_t^i, x_t^{-i})\dot{x}_t^i \geq F(x_t^i + \dot{x}_t^i, x_t^{-i})$$

which implies

$$\dot{f}_t \geq \sum_i [F(x_t^i + \dot{x}_t^i, x_t^{-i}) - F(x_t)] = W(x_t) \geq 0$$

hence f is increasing but bounded. f is thus constant on the limit set $L(\mathbf{x})$. By the previous inequality, for any accumulation point x^* one has $W(x^*) = 0$ and x^* is a Nash equilibrium. □

In this framework also, one can deduce the convergence of the discrete time process from the properties of the continuous time analog, however $NE(F)$ is not a global attractor and the proof is much more involved (Benaim, Hofbauer and Sorin, 2005).

PROPOSITION 1.12. *Assume $F(X_F)$ with non empty interior. Then (DFP) converges to $NE(F)$.*

PROOF. Contrary to the zero-sum case where $X_F \times Y_F$ was a global attractor the proof uses here the tools of stochastic approximation, see Section 5, Proposition 5.3, with $-F$ as Lyapounov function and $NE(F)$ as critical set and Theorem 5.3. □

Remarks. Note that one cannot expect uniform convergence. See the standard symmetric coordination game:

$(1,1)$	$(0,0)$
$(0,0)$	$(1,1)$

The only attractor that contains $NE(F)$ is the diagonal. In particular convergence of (CFP) does not imply directly convergence of (DFP). Note that the equilibrium $(1/2, 1/2)$ is unstable but the time to go from $(1/2^+, 1/2^-)$ to $(1,0)$ is not bounded.

1.6. Complements.

We assume here the payoff to be multilinear and we state several properties of (DFP) and (CFP).

1.6.1. *General properties.*
Strict Nash are asymptotically stable and stricly dominated strategies are eliminated.

1.6.2. *Anticipated and realized payoff.*
Monderer, Samet and Sela (1997) introduce a comparison between the anticipated payoff at stage n ($E^i_n = F^i(x^i_n, \bar{x}^{-i}_{n-1})$) and the average payoff up to stage n (exclusive) ($A^i_n = \frac{1}{n-1}\sum_{p=1}^{n-1} F^i(x_p)$).

PROPOSITION 1.13. *Assume (DFP) for player i (with 2 players or correlated (DFP)), then*

$$E^i_n \geq A^i_n. \tag{1.12}$$

PROOF. In fact, by definition of (DFP) and by linearity:

$$\sum_{m\leq n-1} F^i(x^i_n, x^{-i}_m) \geq \sum_{m\leq n-1} F^i(s, x^{-i}_m), \quad \forall s \in X^i. \tag{1.13}$$

Write $(n-1)E^i_n = b_n = \sum_{m\leq n-1} a(n,m)$ for the left hand side. By choosing $s = x^i_{n-1}$ one obtains

$$b_n \geq a(n-1, n-1) + b_{n-1}$$

hence by induction

$$E^i_n \geq A^i_n = \sum_{m\leq n-1} a(m,m)/(n-1).$$

□

Remark
This is a unilateral property: no hypothesis is made on the behavior of player $-i$.

COROLLARY 1.2. *The average payoffs converge to the value for (DFP) in the zero-sum case.*

PROOF. Recall that in this case E^1_n (resp. E^2_n) converges to v (resp. $-v$), since $\bar{x}^{-i}_n$ converges to the set of optimal strategies of $-i$. □

The corresponding result in the continuous time setting is

PROPOSITION 1.14. *Assume (CFP) for player i in a two-person game, then*

$$\lim_{t\to+\infty}(E^i_t - A^i_t) = 0.$$

PROOF. Denote by α_s the move at time s so that:

$$tx_t = \int_0^t \alpha_s ds.$$

and $\alpha_t \in BR^1(y_t)$. One has

$$t\dot{x}_t + x_t = \alpha_t$$

which is

$$\dot{x}_t \in \frac{1}{t}[BR^1(y_t) - x_t].$$

Hence the anticipated payoff for player 1 is

$$E^1_t = F^1(\alpha_t, y_t)$$

and the past average payoff satisfies

$$tA^1_t = \int_0^t F^1(\alpha_s, \beta_s)ds.$$

Taking derivatives one obtains

$$\frac{d}{dt}[tA^1_t] = F^1(x_t + t\dot{x}_t, y_t + t\dot{y}_t) = F^1(\alpha_t, \beta_t)$$

$$\frac{d}{dt}[tE^1_t] = E^1_t + t\frac{d}{dt}E^1_t.$$

But $D_1F^1(\alpha, y) = 0$ (envelope theorem) and $D_2F^1(\alpha, y)\dot{y} = F^1(\alpha, \dot{y})$ by linearity. Using again linearity one obtains

$$\frac{d}{dt}[tE^1_t] = F^1(x_t + t\dot{x}_t, y_t) + F^1(x_t + t\dot{x}_t, t\dot{y}_t) = \frac{d}{dt}[tA^1_t]$$

hence there exists C such that

$$E_t - A_t = \frac{C}{t}.$$

□

COROLLARY 1.3. *Convergence of the average payoffs to the value holds for (CFP) in the zero-sum case.*

PROOF. Since y_t converges to Y_F, E^1_t and the average payoff converges to the value. □

1.6.3. *Improvement principle.*
An interesting property is due to Monderer and Sela (1993). Note that it is not expressed in the usual state variable $(\bar{x}_n)$ but is related to **Myopic Adjustment Dynamics** satisfying: $F(\dot{x}, x) \geq 0$.

PROPOSITION 1.15. *Assume (DFP) for player i with 2 players; then*

$$F^i(x^i_n, x^{-i}_{n-1}) \geq F^i(x_{n-1}). \tag{1.14}$$

PROOF. In fact the (DFP) property implies

$$F^i(x^i_{n-1}, \overline{x}^{-i}_{n-2}) \geq F^i(x^i_n, \overline{x}^{-i}_{n-2}) \tag{1.15}$$

and

$$F^i(x^i_n, \overline{x}^{-i}_{n-1}) \geq F^i(x^i_{n-1}, \overline{x}^{-i}_{n-1}). \tag{1.16}$$

Hence if equation (1.14) is not satisfied adding it to (1.15) and using the linearity of the payoff would contradict (1.16). □

These properties will be useful in proving non convergence.

1.7. Shapley's example.
Consider the next two player game, due to Shapley (1964):

G =

$(0,0)$	(a,b)	(b,a)
(b,a)	$(0,0)$	(a,b)
(a,b)	(b,a)	$(0,0)$

with $a > b > 0$. Note that the only equilibrium is $(1/3, 1/3, 1/3)$.

PROPOSITION 1.16. *(DFP) does not always converge.*

PROOF.
Proof 1. Starting from a Pareto entry the improvement principle (1.14) implies that (DFP) will stay on Pareto entries. Hence the sum of the stage payoffs will always be $(a+b)$. If (DFP) converges then it converges to $(1/3, 1/3, 1/3)$ so that the anticipated payoff converges to the Nash payoff $\frac{a+b}{3}$ which contradicts inequality (1.12).

Proof 2. Add a line to the Shapley matrix G defining a new matrix

G' =

$(0,0)$	(a,b)	(b,a)
(b,a)	$(0,0)$	(a,b)
(a,b)	(b,a)	$(0,0)$
$(c,0)$	$(c,0)$	$(c,0)$

with $2a > b > c > \frac{a+b}{3}$.
By the improvement principle (1.14), starting from a Pareto entry one will stay on the Pareto set, hence line 4 will not be played so that (DFP) in G' is also (DFP) in G. If there were convergence it would be to a Nash equilibrium hence to $(1/3, 1/3, 1/3)$ in G, thus to $[(1/3, 1/3, 1/3, 0); (1/3, 1/3, 1/3)]$ in G'. But a best reply for player 1 to $(1/3, 1/3, 1/3)$ in G' is the fourth line, contradiction.
Proof 3. Following Shapley (1964) let us study explicitly the (DFP) trajectory. Starting from (12), there is a cycle : 12, 13, 23, 21, 31, 32, 12,... Let $r(ij)$ be the duration of the corresponding entry and α the vector of cumulative payoffs of player 1 at the beginning of the cycle i.e. if it occurs at stage $n+1$, given by:

$$\alpha_i = \sum_{m=1}^{n} A_{i_m j_m}$$

which is proportional to the payoff of move i against the empirical average $\bar{y}_n$. Thus, after $r(12)$ stages of (12) and $r(13)$ stages of (13) the new vector α' satisfies

$$\alpha'_1 = \alpha_1 + r(12)a + r(13)b$$

$$\alpha'_2 = \alpha_2 + r(12)0 + r(13)a$$

and then player 1 switches to move 2, hence one has

$$\alpha'_2 \geq \alpha'_1$$

but also

$$\alpha_1 \geq \alpha_2$$

(because 1 was played) so that

$$\alpha'_2 - \alpha_2 \geq \alpha'_1 - \alpha_1$$

which gives

$$r(13)(a-b) \geq r(12)a$$

and by induction at the next round

$$r'(11) \geq [\frac{a}{(a-b)}]^6 r(11)$$

so that exponential growth occurs and the empirical distribution does not converge (compare with the Shapley triangle, see Gaunersdorfer and Hofbauer (1995) and the chapter by J. Hofbauer). □

1.8. Other classes.

1.8.1. *Coordination games.*
A coordination game is a two person (square) game where each diagonal entry defines a pure Nash equilibrium. There are robust examples of coordination games where (DFP) fails to converge, Foster and Young (1998). Note that it is possible to have convergence of (DFP) and convergence of the payoffs to a non Nash payoff - like always mismatching. Better processes allow to select among the memory: choose s dates among the last m ones or work with finite memory adding a perturbation, see the survey in Young (2004).

1.8.2. *Dominance solvable games.*
Convergence properties are obtained in Milgrom and Roberts (1991).

1.8.3. *Supermodular games.*
In this class, convergence results are proved in Milgrom and Roberts (1990). For the case of strategic complementarity and diminishing marginal returns see Krishna and Sjöstrom (1997,1998), Berger (2008).

2. Unilateral Smooth Best Replies and Consistency

We consider here an unilateral process that will exhibit robust properties and which is deeply related to (CFP).

2.1. Consistency.

2.1.1. *Model and definitions.*
Consider a discrete time process $\{U_n\}$ of vectors in $\mathcal{U} = [0,1]^K$.
At each stage n, a player having observed the past realizations $U_1, ..., U_{n-1}$, chooses a component k_n in K. Then U_n is announced and the outcome at that stage is $\omega_n = U_n^{k_n}$.
A strategy σ in this prediction problem is specified by $\sigma(h_{n-1}) \in \Delta(K)$ (the simplex of $\mathbb{R}^K$) which is the probability distribution of k_n given the past history $h_{n-1} = (U_1, k_1, ..., U_{n-1}, k_{n-1})$.

External regret
The regret given $k \in K$ and $U \in \mathbb{R}^K$ is defined by the vector $R(k,U) \in \mathbb{R}^K$ with $R(k;U)^\ell = U^\ell - U^k, \ell \in K$.
Hence the evaluation at stage n is $R_n = R(k_n, U_n)$ i.e. $R_n^k = U_n^k - \omega_n$.
Given a sequence $\{u_m\}$, we define as usual $\bar{u}_n = \frac{1}{n}\sum_{m=1}^n u_m$. Hence the average external regret vector at stage n is $\overline{R}_n$ with

$$\overline{R}_n^k = \overline{U}_n^k - \overline{\omega}_n$$

It compares the actual (average) payoff to the payoff corresponding to a constant choice of a component, see Foster and Vohra (1999), Fudenberg and Levine (1995).

DEFINITION 2.1. A strategy σ satisfies *external consistency* (EC) if, for every process $\{U_m\}$:

$$\max_{k\in K}[\overline{R}^k_n]^+ \longrightarrow 0 \text{ a.s., as } n \to +\infty$$

or, equivalently $\quad \sum_{m=1}^n (U^k_m - \omega_m) \leq o(n), \qquad \forall k \in K.$

Internal regret

The evaluation at stage n is given by a $K \times K$ matrix S_n defined by:

$$S^{k\ell}_n = \begin{cases} U^\ell_n - U^k_n & for\ k = k_n \\ 0 & otherwise. \end{cases}$$

Hence the average internal regret matrix is

$$\overline{S}^{k\ell}_n = \frac{1}{n} \sum_{m=1, k_m=k}^{n} (U^\ell_m - U^k_m).$$

This involves a comparison, for each component k, of the average payoff obtained on the dates where k was played, to the payoff that would have been induced by an alternative choice ℓ, see Foster and Vohra (1999), Fudenberg and Levine (1999). Note that we normalize by $\frac{1}{n}$ to ignore the scores of unfrequent moves.

DEFINITION 2.2. A strategy σ satisfies *internal consistency* (IC) if, for every process $\{U_m\}$ and every couple k, ℓ:

$$[\overline{S}^{k\ell}_n]^+ \longrightarrow 0 \text{ a.s., as } n \to +\infty$$

Note that no assumption is made on the process $\{U_n\}$ (like stationarity or the Markov property), moreover the player has no a priori beliefs on the law of $\{U_n\}$: we are not in a Bayesian framework and there is in general no learning, but adaptation.

2.1.2. *Application to games.*

Consider a finite game with $\#I$ players having action spaces $S^j, j \in I$. The game is repeated in discrete time and after each stage the previous profile of moves is announced. Each player i knows her payoff function $G^i : S = S^i \times S^{-i} \to \mathbb{R}$ and her observation is the vector of moves of her opponents, $s^{-i} \in S^{-i}$.

Fix i and let $K = S^i$. Player i knows in particular after stage n his stage payoff $\omega_n = G^i(k_n, s^{-i}_n)$ as well as his vector payoff $U_n = G^i(., s^{-i}_n) \in \mathbb{R}^K$. The previous process describes precisely the situation that a player faces in a repeated game (with complete information and standard monitoring). She first has to choose her action, then she discovers the profile played and can evaluate her regret.

Introduce $z_n = \frac{1}{n}\sum_{m=1}^n s_m \in \Delta(S)$ with $s_m = \{s^j_m\}, j \in I$ which is the empirical distribution of profile of moves up to stage n so that by linearity

$$\bar{R}_n = \{G^i(k, z^{-i}_n) - G^i(z_n); k \in K\}.$$

Then we can express the property on the payoffs as a property on the moves.

σ satisfies EC is equivalent to : $z_n \to H^i$ a.s. with

$$H^i = \{z \in \Delta(S); G^i(k, z^{-i}) - G^i(z) \leq 0, \forall k \in K\}.$$

H^i is the Hannan's set of player i, Hannan (1957).

Similarly $\bar{S}_n = S(z_n)$ with

$$S^{k,j}(z) = \sum_{\ell\in S^{-i}} [G^i(j,\ell) - G^i(k,\ell)]z(k,\ell)$$

and σ satisfies IC is equivalent to $z_n \to C^i$ a.s. with

$$C^i = \{z \in \Delta(S); S^{k,j}(z) \leq 0, \forall k, j \in K\}$$

This corresponds to the set of correlated distributions z where, for each move $k \in S^i$, k is a best reply of player i to the conditional distribution of z given k on S^{-i}.
Note that $\cap_i C^i$ is the set of **correlated equilibrium distributions**, Aumann (1974).
In particular the existence of internally consistent procedures will provide an alternative proof of existence of correlated equilibrium distributions: consider any accumulation point of a trajectory generated by players using IC procedures.

2.2. Smooth fictitious play.

We described here a procedure that will satisfies IC. There are two connections with the previous section. First we will deduce properties of the random discrete time process from properties of a deterministic continuous time counterpart. Second the strategy is based on a smooth version of (DFP). Note that this procedure relies only on the previous observations of the process $\{U_n\}$ and not on the moves of the predictor, hence the regret needs not to be known, see Fudenberg and Levine (1995).

DEFINITION 2.3. A *smooth perturbation* of the payoff U is a map

$$V^\varepsilon(x, U) = \langle x, U \rangle + \varepsilon \rho(x),$$

with $0 < \varepsilon < \varepsilon_0$, such that:
(i) $\rho : X \to \mathbb{R}$ is a $\mathcal{C}^1$ function with uniform norm $\|\rho\| \leq 1$,
(ii) $\mathsf{argmax}_{x \in X} V^\varepsilon(., U)$ reduces to one point and defines a continuous map $\mathbf{br}^\varepsilon : U \to X$
called a *smooth best reply function*,
(iii) $D_1 V^\varepsilon(\mathbf{br}^\varepsilon(U), U).D\mathbf{br}^\varepsilon(U) = 0$
(for example $D_1 U^\varepsilon(., U)$ is 0 at $\mathbf{br}^\varepsilon(U)$).

A typical example is obtained via the entropy function

$$\rho(x) = -\sum_k x_k \log x_k. \tag{2.1}$$

which leads to the smooth perturbed best reply function

$$[\mathbf{br}^\varepsilon(U)]^k = \frac{\exp(U^k/\varepsilon)}{\sum_{j \in K} \exp(U^j/\varepsilon)}. \tag{2.2}$$

Let

$$W^\varepsilon(U) = \max_x V^\varepsilon(x, U) = V^\varepsilon(\mathbf{br}^\varepsilon(U), U)$$

that is close to the largest component of U and will be the evaluation criteria. A useful property is the following:

LEMMA 2.1. *(Fudenberg and Levine (1999))*

$$DW^\varepsilon(U) = \mathbf{br}^\varepsilon(U).$$

Let us first consider external consistency.

DEFINITION 2.4. A smooth fictitious play strategy σ^ε associated to the smooth best response function $\mathbf{br}^\varepsilon$ (in short a $\mathsf{SFP}(\varepsilon)$ strategy) is defined by:

$$\sigma^\varepsilon(h_n) = \mathbf{br}^\varepsilon(\overline{U}_n).$$

The corresponding discrete dynamics written in the spaces of both vectors and outcomes is

$$\overline{U}_{n+1} - \overline{U}_n = \frac{1}{n+1}[U_{n+1} - \overline{U}_n]. \tag{2.3}$$

$$\overline{\omega}_{n+1} - \overline{\omega}_n = \frac{1}{n+1}[\omega_{n+1} - \overline{\omega}_n]. \tag{2.4}$$

with

$$\mathsf{E}(\omega_{n+1}|\mathcal{F}_n) = \langle \mathbf{br}^{\varepsilon}(\overline{U}_n), U_{n+1}\rangle \tag{2.5}$$

which express the fact that the choice of the component of the unknown vector U_{n+1} is done according to $\sigma^{\varepsilon}(h_n) = \mathbf{br}^{\varepsilon}(\overline{U}_n)$.

We now use the properties of Section 5 to obtain, following Benaïm, Hofbauer and Sorin (2006):

Lemma 2.2. *The process* $(\overline{U}_n, \overline{\omega}_n)$ *is a Discrete Stochastic Approximation of the differential inclusion with values in* $\mathbb{R}^K \times \mathbb{R}$

$$(\dot{\mathbf{u}}, \dot{\omega}) \in \{(U - \mathbf{u}, \langle \mathbf{br}^{\varepsilon}(\mathbf{u}), U\rangle - \omega); U \in \mathcal{U}\}. \tag{2.6}$$

The main property of the continuous dynamics is given by:

Theorem 2.1. *The set* $\{(u,\omega) \in \mathcal{U} \times \mathbb{R} : W^{\varepsilon}(u) - \omega \le \varepsilon\}$ *is a global attracting set for the continuous dynamics.*
In particular, for any $\eta > 0$*, there exists* $\bar{\varepsilon}$ *such that for* $\varepsilon \le \bar{\varepsilon}$, $\limsup_{t\to\infty} W^{\varepsilon}(\mathbf{u}(t)) - \omega(t) \le \eta$ *(i.e. continuous* $\mathsf{SFP}(\varepsilon)$ *satisfies* η*-consistency).*

Proof. Let $q(t) = W^{\varepsilon}(\mathbf{u}(t)) - \omega(t)$.
Taking time derivative one obtains, using the previous two Lemmas:

$$\begin{aligned}
\dot{q}(t) &= DW^{\varepsilon}(\mathbf{u}(t)).\dot{\mathbf{u}}(t) - \dot{\omega}(t)\\
&= \langle \mathbf{br}^{\varepsilon}(\mathbf{u}(t)), \dot{\mathbf{u}}(t)\rangle - \dot{\omega}(t)\\
&= \langle \mathbf{br}^{\varepsilon}(\mathbf{u}(t)), U - \mathbf{u}(t)\rangle - (\langle \mathbf{br}^{\varepsilon}(\mathbf{u}(t)), U\rangle - \omega(t))\\
&\le -q(t) + \varepsilon.
\end{aligned}$$

so that $q(t) \le \varepsilon + Me^{-t}$ for some constant M. □

In particular we deduce from Theorem 5.3 properties of the discrete time process:

Theorem 2.2. *For any* $\eta > 0$*, there exists* $\bar{\varepsilon}$ *such that for* $\varepsilon \le \bar{\varepsilon}$, $\mathsf{SFP}(\varepsilon)$ *is* η*-consistent.*

Let us now consider internal consistency.
Define $\bar{U}_n[k]$ as the average of U_m on the dates $1 \le m \le n$, where k was played. $\sigma(h_n)$ is now an invariant measure for the matrix defined by the columns

$$\{\mathbf{br}^{\varepsilon}(\bar{U}_n[k])\}_{k\in K}.$$

Properties similar to the above shows that σ satisfies IC, see Benaïm, Hofbauer and Sorin (2006).

For general properties of global smooth fictitious play procedures, see Hofbauer and Sandholm (2002).

Aternative consistent procedures can be found in Hart and Mas Colell (2000, 2003), see also Cesa-Bianchi and Lugosi (2006).

3. Best Reply and Average Replicator Dynamics

3.1. Presentation.

We follow here Hofbauer, Sorin and Viossat (2009).
Recall that in the framework of a symmetric 2 person game with $K \times K$ payoff matrix A played within a single population, the **replicator dynamics** is defined on the simplex Δ of $\mathbb{R}^K$ by

$$\dot{x}_t^k = x_t^k \left(e^k A x_t - x_t A x_t\right), \quad k \in K \qquad (RD) \tag{3.1}$$

where x_t^k denotes the frequency of strategy k at time t. It was introduced by Taylor and Jonker (1978) as the basic selection dynamics for the evolutionary games of Maynard Smith (1982).
In this framework the **best reply dynamics** is the differential inclusion on Δ

$$\dot{z}_t \in BR(z_t) - z_t, \qquad t \geq 0 \qquad (CBR) \tag{3.2}$$

which is the prototype of a population model of rational (but myopic) behaviour.

Despite the different interpretation and the different dynamic character there are amazing similarities in the long run behaviour of these two dynamics, that have been summarized in the following heuristic principle:
For many games, the long run behaviour ($t \to \infty$) of the time averages $X_t = \frac{1}{t}\int_0^t x_s ds$ of the trajectories x_t of the replicator equation is the same as for the BR trajectories.

We provide here a rigorous statement that largely explains this heuristic by showing that for any interior solution of (RD), for every $t \geq 0$, x_t is an approximate best reply against X_t and the approximation gets better as $t \to \infty$. This implies that X_t is an asymptotic pseudo trajectory of (CBR), see section 5, and hence the limit set of X_t has the same properties as a limit set of a true orbit of (CBR), i.e. it is invariant and internally chain transitive under (CBR).
The main tool to prove this is via the logit map which is a canonical smoothing of the best response correspondence. We show that x_t equals the logit approximation at X_t with error rate $\frac{1}{t}$.

3.2. Unilateral processes.

The model will be in the framework of an I-person game but we consider the dynamics for one player, without hypotheses on the behavior of the others. The framework is unilateral, as in the previous section, but now in continuous time. Hence, from the point of view of this player, she is facing a (measurable) vector outcome process $\mathcal{U} = \{U_t, t \geq 0\}$, with values in the cube $C = [-c, c]^K$ where K is her move set and c is some positive constant. U_t^k is the payoff at time t if k is the move at that time. The cumulative vector outcome up to stage t is $S_t = \int_0^t U_s ds$ and its time average is denoted $\bar{U}_t = \frac{1}{t} S_t$.
br denotes the (payoff based) best reply correspondence from C to Δ defined by

$$\mathbf{br}(U) = \{x \in \Delta; \langle x, U\rangle = \max_{y \in \Delta} \langle y, U\rangle\}.$$

The $\mathcal{U}$-best reply process (CBR) is defined on Δ by

$$\dot{X}_t \in [\mathbf{br}(\bar{U}_t) - X_t]. \tag{3.3}$$

The $\mathcal{U}$-replicator process (RP) is specified by the following equation on Δ:

$$\dot{x}_t^k = x_t^k[U_t^k - \langle x_t, U_t\rangle], \qquad k \in K. \tag{3.4}$$

Explicitly, in the framework of an I-player game with payoff for player 1 defined by a function G from $\prod_{i\in I} S^i$ to $\mathbb{R}$, with $X^i = \Delta(S^i)$, U is the vector payoff i.e. $U_t = G(., x_t^{-1})$.
If all the players follow a (payoff based) continuous time correlated fictitious play dynamics, each time average strategy satisfies (3.3).
If all the players follow the replicator dynamics then (3.4) is the replicator dynamics equation.

3.3. Logit rule and perturbed best reply.

Define the map L from $\mathbb{R}^K$ to Δ by

$$L^k(V) = \frac{\exp V^k}{\sum_j \exp V^j}\ . \tag{3.5}$$

Given $\eta > 0$, let $[\mathbf{br}]^\eta$ be the correspondence from C to Δ with graph being the η-neighborhood for the uniform norm of the graph of $\mathbf{br}$.
The L map and the $\mathbf{br}$ correspondence are related as follows:

Proposition 3.1. *For any $U \in C$ and $\varepsilon > 0$*

$$L(U/\varepsilon) \in [\mathbf{br}]^{\eta(\varepsilon)}(U)$$

with $\eta(\varepsilon) \to 0$ as $\varepsilon \to 0$.

Remarks. L is also given by

$$L(V) = \text{argmax}_{x\in\Delta}\{\langle x, V\rangle - \sum_k x^k \log x^k\}.$$

Hence introducing the (payoff based) perturbed best reply $\mathbf{br}^\varepsilon$ from C to Δ defined by

$$\mathbf{br}^\varepsilon(U) = \text{argmax}_{x\in\Delta}\{\langle x, U\rangle - \varepsilon \sum_k x^k \log x^k\}$$

one has $L(U/\varepsilon) = \mathbf{br}^\varepsilon(U)$.
The map $\mathbf{br}^\varepsilon$ is the logit approximation, see (2.2).

3.4. Explicit representation of the replicator process.

The following procedure has been introduced in discrete time in the framework of on-line algorithms under the name "multiplicative weight algorithm", Littlestone and Warmuth (1994). We use here the name (CEW) (continuous exponential weight) for the process defined, given $\mathcal{U}$, by

$$x_t = L(\int_0^t U_s ds).$$

The main property of (CEW) that will be used is that it provides an explicit solution of (RP).

Proposition 3.2. *(CEW) satisfies (RP).*

PROOF. Straightforward computations lead to

$$\dot{x}_t^k = x_t^k U_t^k - x_t^k \sum_j \frac{U_t^j \exp \int_0^t U_v^j dv}{\sum_j \exp \int_0^t U_v^j dv}$$

which is

$$\dot{x}_t^k = x_t^k [U_t^k - \langle x_t, U_t \rangle]$$

hence gives the previous (RP) equation (3.4). □

The link with the best reply correspondence is the following:

PROPOSITION 3.3. *(CEW) satisfies*

$$x_t \in [\mathbf{br}]^{\delta(t)}(\bar{U}_t)$$

with $\delta(t) \to 0$ *as* $t \to \infty$.

PROOF. Write

$$x_t = L(\int_0^t U_s ds) = L(t\, \bar{U}_t).$$

Then

$$x_t = L(U/\varepsilon) \in [\mathbf{br}]^{\eta(\varepsilon)}(U)$$

with $U = \bar{U}_t$ and $\varepsilon = 1/t$, by Proposition 3.1. Let then $\delta(t) = \eta(1/t)$. □

We describe here the consequences for the time average process. Define

$$X_t = \frac{1}{t} \int_0^t x_s ds.$$

PROPOSITION 3.4. *If* x_t *follows (CEW) then* X_t *satisfies*

$$\dot{X}_t \in \frac{1}{t}([\mathbf{br}]^{\delta(t)}(\bar{U}_t) - X_t). \tag{3.6}$$

with $\delta(t) \to 0$ *as* $t \to \infty$.

PROOF. One has, taking derivatives:

$$t\dot{X}_t + X_t = x_t$$

and the result follows from the properties of x_t. □

3.5. Consequences for games.

Consider a 2 person (bimatrix) game (A, B).
If the game is symmetric this gives rise to the single population replicator dynamics (RD) and best reply dynamics (BRD) as defined in section 1.
Otherwise, we consider the two population replicator dynamics

$$\begin{aligned} \dot{x}_t^k &= x_t^k \left(e^k A y_t - x_t A y_t \right), \quad k \in S^1 \\ \dot{y}_t^k &= y_t^k \left(x_t B e^k - x_t B y_t \right), \quad k \in S^2 \end{aligned} \tag{3.7}$$

and the corresponding BR dynamics as in (3).
Let M be the state space (a simplex Δ or a product of simplices $\Delta_1 \times \Delta_2$). We now use the previous results with the process $\mathcal{U}$ being defined by $U_t = Ay_t$ for player 1, hence $\bar{U}_t = AY_t$. Note that $\mathbf{br}(AY) = BR^1(Y)$.

PROPOSITION 3.5. *The limit set of every replicator time average process* X_t *starting from an initial point* $x_0 \in M$ *is a closed subset of* M *which is invariant and internally chain transitive under (CBR).*

PROOF. Equation (3.6) implies that X_t satisfies a perturbed version of (CFP) hence X_{e^t} is a perturbed solution to the differential inclusion (CBR), according to Section 5 and Theorems 5.1 and 5.2 apply. □

In particular this implies:

PROPOSITION 3.6. *Let $\mathcal{A}$ be the global attractor (i.e., the maximal invariant set) of (CBR). Then the limit set of every replicator time average process X_t is a subset of $\mathcal{A}$.*

3.6. External consistency.
The natural continuous time counterpart of the (discrete time) notion is the following: a procedure satisfies external consistency if for each process $\mathcal{U}$ taking values in $\mathbb{R}^K$, it produces a process $x_t \in \Delta$, such that for all k

$$\int_0^t [U_s^k - \langle x_s, U_s\rangle]ds \leq C_t = o(t)$$

where, using a martingale argument, we have replaced the actual random payoff at time s by its conditional expectation $\langle x_s, U_s\rangle$. This property says that the (expected) average payoff induced by x_t along the play is asymptotically not less than the payoff obtained by any fixed choice $k \in K$.

PROPOSITION 3.7. (RP) *satifies external consistency.*

PROOF. By integrating equation (3.4), one obtains, on the support of x_0:

$$\int_0^t [U_s^k - \langle x_s, U_s\rangle]ds = \int_0^t \frac{\dot{x}_s^k}{x_s^k} ds = \log(\frac{x_t^k}{x_0^k}) \leq -\log x_0^k.$$

□

This result is the unilateral analog of the fact that interior rest points of (RD) are equilibria. A myopic unilateral adjustment process provides asymptotic optimal properties in terms of no regret.
Back to a game framework this implies that if player 1 follows (RP) the set of accumulation points of the empirical correlated distribution process will belong to her reduced Hannan set:

$$\bar{H}^1 = \{\theta \in \Delta(S); G^1(k, \theta^{-1}) \leq G^1(\theta), \forall k \in S^1\}$$

with equality for at least one component.
The example due to Viossat (2007, 2008) of a game where the limit set for the replicator dynamics is disjoint from the unique correlated equilibrium shows that (RP) does not satisfy internal consistency.
This later property uses additional information that is not taken into account in the replicator dynamics. This topic deserves further study.

3.7. Comments.
We can now compare several processes in the spirit of (payoff based) fictitious play.
The original fictitious play process (I) is defined by

$$x_t \in \mathbf{br}(\bar{U}_t)$$

The corresponding time average satisfies (CFP).
With a smooth best reply process one has (II)

$$x_t = \mathbf{br}^{\varepsilon}(\bar{U}_t)$$

and the corresponding time average satisfies a smooth fictitious play process.
Finally the replicator process (III) satisfies

$$x_t = \mathbf{br}^{1/t}(\bar{U}_t)$$

and the time average follows a time dependent perturbation of the fictitious play process.

While in (I), the process x_t follows exactly the best reply correspondence, the induced average X_t does not have good unilateral properties.
One the other hand for (II), X_t satisfies a weak form of external consistency, with an error term $\alpha(\varepsilon)$ vanishing with ε.
In contrast, (III) satisfies exact external consistency due to a both smooth and time dependent approximation of **br**.

4. General Adaptive Dynamics

We consider here random processes corresponding to adaptive behavior in repeated interactions.
The analysis is done from the point of view of one player, having a finite set K of actions. Time is discrete and the behavior of the player depends upon a parameter $z \in Z$.
At stage n, the state is z_{n-1} and the process is defined by two functions:
a **decision map** σ from Z to $\Delta(K)$ (the simplex on K) defining the law π_n of the current action k_n as a function of the parameter:

$$\pi_n = \sigma(z_{n-1})$$

and given the observation ω_n of the player, after the play at stage n, an **updating rule** for the state variable, that depends upon the stage:

$$z_n = \Phi_n(z_{n-1}, \omega_n).$$

Remark
Note that the decision map is stationary but that the updating rule may depend upon the stage.
A typical assumption in game theory is that the player knows his payoff function $G : K \times L \to \mathbb{R}$ and that the observation ω is the vector of moves of his opponents, $\ell \in L$. In particular ω_n contains the stage payoff $g_n = G(k_n, \ell_n)$ as well as the vector payoff $U_n = G(., \ell_n) \in \mathbb{R}^K$.

Example 1: Fictious Play
The state space is usually the empirical distribution of actions of the opponents but one can as well take $\omega_n = U_n$, the vector payoff, then $z_n = \overline{U}_n$ is the average vector payoff thus satisfies:

$$z_n = \frac{(n-1)z_{n-1} + U_n}{n}$$

and

$$\sigma(z) \in BR(z) \qquad \text{or} \qquad \sigma(z) = BR^{\varepsilon}(z).$$

Example 2: Potential regret dynamics
Here

$$R_n = U_n - g_n \mathbf{1}$$

is the "regret vector" at stage n and the updating rule $z_n = \Phi_n(z_{n-1}, \omega_n)$ is simply

$$z_n = \overline{R}_n.$$

Choose P to be a "potential function" for the negative orthant $D = \mathbb{R}^K_-$ and for $z \notin D$ let $\sigma(z)$ be proportional to $\nabla P(z)$.

Example 3: Cumulative proportional reinforcement
The observation ω_n is only the stage payoff g_n (we assume all payoffs ≥ 1).
The updating rule is

$$z_n^k = z_{n-1}^k + g_n \, \mathbf{I}_{\{k_n = k\}}$$

and the decision map is $\sigma(z)$ proportional to the vector z.
There is an important literature on such reinforcement dynamics, see e.g. Beggs (2005), Börgers, Morales and Sarin (2004), Börgers and Sarin (1997), Hopkins (2002), Hopkins and Posch (2005), Laslier, Topol and Walliser (2001), Leslie and Collins (2005), Pemantle (2007), Posch (1997).

Note that these three procedures can be written as

$$z_n = \frac{(n-1)z_{n-1} + v_n}{n}$$

where v_n is a random variable depending on the action(s) ℓ of the opponent(s) and on the action k_n having distribution $\sigma(z_{n-1})$. Thus

$$z_n - z_{n-1} = \frac{1}{n}[v_n - z_{n-1}].$$

Write

$$v_n = E_{\pi_n}(v_n | z_1, ..., z_{n-1}) + [v_n - E_{\pi_n}(v_n | z_1, ..., z_{n-1})]$$

and define

$$S(z_{n-1}) = Co\{E_{\pi_n}(v_n | z_1, ..., z_{n-1}); \ell \in L\}$$

where Co stands for the convex hull. Thus

$$z_n - z_{n-1} \in \frac{1}{n}[S(z_{n-1}) - z_{n-1}].$$

The differential inclusion is

$$\dot{z} \in S(z) - z \qquad (4.1)$$

and the process z_n is a Discrete Stochastic Approximation of (4.1), see section 5.

For further results with explicit applications of this procedure see e.g. Hofbauer and Sanholm (2002), Benaïm, Hofbauer and Sorin (2006), Cominetti, Melo and Sorin (2010).

In conclusion, a large class of adaptive dynamics can be expressed in discrete time as a random difference equation with vanishing step size. Information on the asymptotic behavior can then be obtained by studing the continuous time deterministic analog obtained as above.

5. Stochastic Approximation for Differential Inclusions

We summarize here results from Benaïm, Hofbauer and Sorin (2005).

5.1. Differential inclusions.

Given a correspondence F from $\mathbb{R}^m$ to itself, consider the differential inclusion

$$\dot{\mathbf{x}} \in F(\mathbf{x}) \qquad (I)$$

It induces a set-valued dynamical system $\{\Phi_t\}_{t\in\mathbb{R}}$ defined by

$$\Phi_t(x) = \{\mathbf{x}(t) : \mathbf{x} \text{ is a solution to (I) with } \mathbf{x}(0) = x\}.$$

We also write $\mathbf{x}(t) = \phi_t(x)$.

Definition 5.1.

1) x is a *rest point* if $0 \in F(x)$.
2) A set C is *strongly forward invariant* (SFI) if $\Phi_t(C) \subset C$ for all $t \geq 0$.
3) C is *invariant* if for any $x \in C$ there exists a complete solution: $\phi_t(x) \in C$ for all $t \in \mathbb{R}$.
4) C is *Lyapounov stable* if: $\forall \varepsilon > 0, \exists \delta > 0$ such that $d(y, C) \leq \delta$ implies $d(\Phi_t(y), C) \leq \varepsilon$ for all $t \geq 0$, i.e.

$$\Phi_{[0,+\infty)}(C^\delta) \subset C^\varepsilon.$$

5) C is a *sink* if there exists $\delta > 0$ such that for any $y \in C^\delta$ and any ϕ:

$$d(\phi_t(y), C) \to 0 \text{ as } t \to \infty.$$

A neighborhood U of C having this property is called a *basin of attraction* of C.
6) C is *attracting* if it is compact and the previous property is uniform. Thus there exist $\delta > 0$, $\varepsilon_0 > 0$ and a map $T : (0, \varepsilon_0) \to \mathbb{R}^+$ such that: for any $y \in C^\delta$, any solution ϕ, $\phi_t(y) \in C^\varepsilon$ for all $t \geq T(\varepsilon)$, i.e.

$$\Phi_{[T(\varepsilon),+\infty)}(C^\delta) \subset C^\varepsilon, \quad \forall \varepsilon \in (0, \varepsilon_0).$$

A neighborhood U of C having this property is called a *uniform basin of attraction* of C and we will write $(C; U)$ for the couple.
7) C is an *attractor* if it is attracting and invariant.
8) C is *forward precompact* if there exists a compact K and a time T such that $\Phi_{[T,+\infty)}(C) \subset K$.
9) The *ω-limit set* of C is defined by

$$\omega_\Phi(C) = \cap_{s\geq 0}\overline{\cup_{y\in C} \cup_{t\geq s} \Phi_t(y)} = \cap_{s\geq 0}\overline{\Phi_{[s,+\infty)}(C)} \tag{5.1}$$

where $\overline{A}$ denotes the closure of the set A.

Definition 5.2.

i) Given a closed invariant set L, the induced set-valued dynamical system Φ^L is defined on L by

$$\Phi_t^L(x) = \{\mathbf{x}(t) : \mathbf{x} \text{ is a solution to (I) with } \mathbf{x}(0) = x \text{ and } \mathbf{x}(\mathbb{R}) \subset L\}.$$

Note that $L = \Phi_t^L(L)$ for all t.
ii) Let $A \subset L$ be an attractor for Φ^L. If $A \neq L$ and $A \neq \emptyset$, then A is a *proper attractor*.
An invariant set L is *attractor free* if Φ^L has no proper attractor.

5.2. Attractors.
The next notion is fondamental in the analysis.

DEFINITION 5.3.
C is *asymptotically stable* if it has the following properties
i) invariant
ii) Lyapounov stable
iii) sink.

PROPOSITION 5.1. *Assume C compact. Attractor is equivalent to asymptotically stable.*

PROPOSITION 5.2. *Let A be a compact set, U be a relatively compact neighborhood and V a function from $\overline{U}$ to $\mathbb{R}^+$. Consider the following properties*
i) U is (SFI)
ii) $V^{-1}(0) = A$
iii) V is continuous and strictly decreasing on trajectories on $U \setminus A$:

$$V(x) > V(y), \quad \forall x \in U \setminus A, \forall y \in \phi_t(x), \quad \forall t > 0$$

iv) V is upper semi continuous and strictly decreasing on trajectories on $\overline{U} \setminus A$.
a) Then under i), ii) and iii) A is Lyapounov stable and $(A; U)$ is attracting.
b) Under i), ii) and iv), $(B; U)$ is an attractor for some $B \subset A$.

DEFINITION 5.4.
A real continuous function V on U open in $\mathbb{R}^m$ is a *Lyapunov function for $A \subset U$* if : $V(y) < V(x)$ for all $x \in U \setminus A, y \in \phi_t(x), t > 0$; and $V(y) \leq V(x)$ for all $x \in A, y \in \phi_t(x)$ and $t \geq 0$.
Note that for each solution ϕ, V is constant along its *limit set*

$$L(\phi)(x) = \cap_{s \geq 0} \overline{\phi_{[s,+\infty)}(x)}.$$

PROPOSITION 5.3. *Suppose V is a Lyapunov function for A. Assume that $V(A)$ has empty interior. Let L be a non empty, compact, invariant and attractor free subset of U. Then L is contained in A and V is constant on L.*

5.3. Asymptotic pseudo-trajectories and internally chain transitive sets.

5.3.1. *Asymptotic pseudo-trajectories.*

DEFINITION 5.5. The *translation flow* $\Theta : C^0(\mathbb{R}, \mathbb{R}^m) \times \mathbb{R} \to C^0(\mathbb{R}, \mathbb{R}^m)$ is defined by

$$\Theta_t(\mathbf{x})(s) = \mathbf{x}(s + t).$$

A continuous function $\mathbf{z} : \mathbb{R}^+ \to \mathbb{R}^m$ is an *asymptotic pseudo-trajectory* (APT) for Φ if for all T

$$\lim_{t \to \infty} \inf_{\mathbf{x} \in S_{\mathbf{z}(t)}} \sup_{0 \leq s \leq T} \|\mathbf{z}(t + s) - \mathbf{x}(s)\| = 0. \tag{5.2}$$

where S_x denotes the set of all solutions of (I) starting from x at 0 and $S = \bigcup_{x \in \mathbb{R}^m} S_x$.

In other words, for each fixed T, the curve: $s \to \mathbf{z}(t + s)$ from $[0, T]$ to $\mathbb{R}^m$ shadows some trajectory for (I) of the point $\mathbf{z}(t)$ over the interval $[0, T]$ with arbitrary accuracy, for sufficiently large t. Hence $\mathbf{z}$ has a forward trajectory under Θ attracted by S. One extends $\mathbf{z}$ to $\mathbb{R}$ by letting $\mathbf{z}(t) = \mathbf{z}(0)$ for $t < 0$.

5.3.2. *Internally chain transitive sets.*
Given a set $A \subset \mathbb{R}^m$ and $x, y \in A$, we write $x \to_A y$ if for every $\varepsilon > 0$ and $T > 0$ there exists an integer $n \in I\!N$, solutions $\mathbf{x}_1, \ldots, \mathbf{x}_n$ to (I), and real numbers $t_1, t_2, \ldots, t_n$ greater than T such that
a) $\mathbf{x}_i(s) \in A$ for all $0 \le s \le t_i$ and for all $i = 1, \ldots, n$,
b) $\|\mathbf{x}_i(t_i) - \mathbf{x}_{i+1}(0)\| \le \varepsilon$ for all $i = 1, \ldots, n-1$,
c) $\|\mathbf{x}_1(0) - x\| \le \varepsilon$ and $\|\mathbf{x}_n(t_n) - y\| \le \varepsilon$.
The sequence $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is called an (ε, T) chain (in A from x to y) for (I).

Definition 5.6.
A set $A \subset \mathbb{R}^m$ is *internally chain transitive* (ICT) if it is compact and $x \to_A y$ for all $x, y \in A$.

Lemma 5.1. *An internally chain transitive set is invariant.*

Proposition 5.4. *Let L be internally chain transitive. Then L has no proper attracting set for Φ^L.*

This (ICT) notion of recurrence due to Conley (1978) for classical dynamical systems is well suited to the description of the asymptotic behavior of APT, as shown by the following theorem. Let

$$L(\mathbf{z}) = \bigcap_{t \ge 0} \overline{\{\mathbf{z}(s) : s \ge t\}}$$

be the limit set.

Theorem 5.1. *Let $\mathbf{z}$ be a bounded APT of (I). Then $L(\mathbf{z})$ is internally chain transitive.*

5.4. Perturbed solutions.

The purpose of this paragraph is to study trajectories which are obtained as (deterministic or random) perturbations of solutions of (I).
5.4.1. *Perturbed solutions.*

Definition 5.7.
A continuous function $\mathbf{y} : \mathbb{R}_+ = [0, \infty) \to \mathbb{R}^m$ is a *perturbed solution* to (I) if it satisfies the following set of conditions (II):
i) $\mathbf{y}$ is absolutely continuous.
ii) There exists a locally integrable function $t \mapsto U(t)$ such that

$$\lim_{t \to \infty} \sup_{0 \le v \le T} \left\| \int_t^{t+v} U(s)\, ds \right\| = 0$$

for all $T > 0$
iii)

$$\frac{d\mathbf{y}(t)}{dt} - U(t) \in F^{\delta(t)}(\mathbf{y}(t))$$

for almost every $t > 0$, for some function $\delta : [0, \infty) \to \mathbb{R}$ with $\delta(t) \to 0$ as $t \to \infty$.
Here $F^\delta(x) := \{y \in \mathbb{R}^m : \exists z : \|z - x\| < \delta,\ d(y, F(z)) < \delta\}$.

The aim is to investigate the long-term behavior of $\mathbf{y}$ and to describe its limit set $L(\mathbf{y})$ in terms of the dynamics induced by F.

Theorem 5.2. *Any bounded solution $\mathbf{y}$ of (II) is an APT of (I).*

5.4.2. *Discrete stochastic approximation.*
As will be shown here, a natural class of perturbed solutions to F arises from certain stochastic approximation processes.

DEFINITION 5.8.
A discrete time process $\{x_n\}_{n\in\mathbb{N}}$ with values in $\mathbb{R}^m$ is a solution for (III) if it verifies a recursion of the form

$$x_{n+1} - x_n - \gamma_{n+1}U_{n+1} \in \gamma_{n+1}F(x_n), \qquad (III)$$

where the characteristics γ and U satisfy
i) $\{\gamma_n\}_{n\geq 1}$ is a sequence of nonnegative numbers such that

$$\sum_n \gamma_n = \infty, \qquad \lim_{n\to\infty} \gamma_n = 0;$$

ii) $U_n \in \mathbb{R}^m$ are (deterministic or random) perturbations.

To such a process is naturally associated a continuous time process as follows.

DEFINITION 5.9.
Let $\tau_0 = 0$ and $\tau_n = \sum_{i=1}^n \gamma_i$ for $n \geq 1$, and define the continuous time affine interpolated process $\mathbf{w} : \mathbb{R}_+ \to \mathbb{R}^m$ by

$$\mathbf{w}(\tau_n + s) = x_n + s\frac{x_{n+1} - x_n}{\tau_{n+1} - \tau_n}, \quad s \in [0, \gamma_{n+1}). \qquad (IV)$$

5.5. From interpolated process to perturbed solutions.
The next result gives sufficient conditions on the characteristics of the discrete process (III) for its interpolation (IV) to be a perturbed solution (II).
If (U_i) are random variables, assumptions (i) and (ii) below hold with probability one.

PROPOSITION 5.5. *Assume that the following hold:*
(i) *For all* $T > 0$

$$\lim_{n\to\infty} \sup \left\{ \left\| \sum_{i=n}^{k-1} \gamma_{i+1}U_{i+1} \right\| : k = n+1, \dots, m(\tau_n + T) \right\} = 0,$$

where

$$m(t) = \sup\{k \geq 0 : t \geq \tau_k\}; \tag{5.3}$$

(ii) $\sup_n \|x_n\| = M < \infty$.
Then the interpolated process $\mathbf{w}$ *is a perturbed solution of* (I).

We describe now sufficient conditions.
Let $(\Omega, \mathbf{\Psi}, P)$ be a probability space and $\{\mathbf{\Psi}_n\}_{n\geq 0}$ a filtration of $\mathbf{\Psi}$ (i.e., a non-decreasing sequence of sub-σ-algebras of $\mathbf{\Psi}$). A stochastic process $\{x_n\}$ given by (III) satisfies the *Robbins–Monro condition* with martingale difference noise if its characteristics satisfy the following:
i) $\{\gamma_n\}$ is a deterministic sequence.
ii) $\{U_n\}$ is *adapted* to $\{\mathbf{\Psi}_n\}$, which means that U_n is measurable with respect to $\mathbf{\Psi}_n$ for each $n \geq 0$.
iii) $\mathbf{E}(U_{n+1} \mid \mathbf{\Psi}_n) = 0$.
The next proposition is a classical estimate for stochastic approximation processes. Note that F does not appear, see Benaïm (1999) for a proof and further references.

PROPOSITION 5.6. *Let $\{x_n\}$ given by (III) be a Robbins–Monro process. Suppose that for some $q \geq 2$*

$$\sup_n \mathbf{E}(\|U_n\|^q) < \infty$$

and

$$\sum_n \gamma_n^{1+q/2} < \infty.$$

Then assumption (i) *of Proposition* 5.5 *holds with probability* 1*.*

Remark. Typical applications are
i) U_n uniformly bounded in L^2 and $\gamma_n = \frac{1}{n}$,
ii) U_n uniformly bounded and $\gamma_n = o(\frac{1}{\log n})$.

5.6. Main result.

Consider a random discrete process defined on a compact subset of $\mathbb{R}^K$ and satisfying the differential inclusion :

$$Y_n - Y_{n-1} \in a_n[T(Y_{n-1}) + W_n]$$

where
i) T is an u.s.c. correspondence with compact convex values
ii) $a_n \geq 0$, $\sum_n a_n = +\infty$, $\sum_n a_n^2 < +\infty$
iii) $E(W_n|Y_1, ..., Y_{n_1}) = 0$.

THEOREM 5.3. *The set of accumulation points of $\{Y_n\}$ is almost surely a compact set, invariant and attractor free for the dynamical system defined by the differential inclusion:*

$$\dot{Y} \in T(Y).$$

References

[1] AUBIN J.-P. AND A. CELLINA (1984) *Differential Inclusions*, Springer.
[2] AUER P., CESA-BIANCHI N., FREUND Y. AND R.E. SHAPIRE (2002) The nonstochastic multi-armed bandit problem, *SIAM J. Comput.*, **32**, 48-77.
[3] AUMANN R.J. (1974) Subjectivity and correlation in randomized strategies, *Journal of Mathematical Economics*, **1**, 67-96.
[4] BEGGS A. (2005) On the convergence of reinforcement learning, *Journal of Economic Theory*, **122**, 1-36.
[5] BENAÏM M. (1996) A dynamical system approach to stochastic approximation, *SIAM Journal on Control and Optimization*, **34**, 437-472.
[6] BENAÏM M. (1999) Dynamics of Stochastic Algorithms, *Séminaire de Probabilités*, **XXIII**, Azéma J. and alii eds, Lectures Notes in Mathematics, **1709**, Springer, 1-68.
[7] BENAÏM M. AND M.W. HIRSCH (1996) Asymptotic pseudotrajectories and chain recurrent flows, with applications, *J. Dynam. Differential Equations*, **8**, 141-176.
[8] BENAÏM M. AND M.W. HIRSCH (1999) Mixed equilibria and dynamical systems arising from fictitious play in perturbed games, *Games and Economic Behavior*, **29**, 36-72.
[9] BENAÏM M., J. HOFBAUER AND S. SORIN (2005) Stochastic approximations and differential inclusions, *SIAM J. Opt. and Control*, **44**, 328-348.
[10] BENAÏM M., J. HOFBAUER AND S. SORIN (2006) Stochastic approximations and differential inclusions. Part II: applications, *Mathematics of Operations Research*, **31**, 673-695.
[11] BERGER U. (2005) Fictitious play in $2 \times n$ games, *Journal of Economic Theory*, **120**, 139-154.
[12] BERGER U. (2008) Learning in games with strategic complementarities revisited, *Journal of Economic Theory*, **143**, 292-301.
[13] BÖRGERS T., A. MORALES AND R. SARIN (2004) Expedient and monotone learning rules, *Econometrica*, **72**, 383-406.
[14] BÖRGERS T. AND R. SARIN (1997) Learning through reinforcement and replicator dynamics, *Journal of Economic Theory*, **77**, 1-14.

[15] Brown G. W. (1949) Some notes on computation of games solutions, RAND Report P-78, The RAND Corporation, Santa Monica, California.

[16] Brown G. W. (1951) Iterative solution of games by fictitious play, in Koopmans T.C. (ed.) *Activity Analysis of Production and Allocation* , Wiley, 374-376.

[17] Brown G.W. and J. von Neumann (1950) Solutions of games by differential equations, *Contributions to the Theory of Games I*, Annals of Mathematical Studies, **24**, 73-79.

[18] Cesa-Bianchi N. and G. Lugosi (2006) *Prediction, Learning and Games*, Cambridge University Press.

[19] Cominetti R., E. Melo and S. Sorin (2010) A payoff-based learning procedure and its application to traffic games, *Games and Economic Behavior*, **70**, 71-83.

[20] Conley C.C. (1978) *Isolated Invariant Sets and the Morse Index*, CBMS Reg. Conf. Ser. in Math. 38, AMS, Providence, RI, 1978.

[21] Foster D. and R. Vohra (1997) Calibrated leaning and correlated equilibria, *Games and Economic Behavior*, **21**, 40-55.

[22] Foster D. and R. Vohra (1999) Regret in the on-line decision problem, *Games and Economic Behavior*, **29**, 7-35.

[23] Foster D. and P. Young (1998) On the nonconvergence of fictitons play in coordination games, *Games and Economic Behavior*, **25**, 79-96.

[24] Fudenberg D. and D. K. Levine (1995) Consistency and cautious fictitious play, *Journal of Economic Dynamics and Control*, **19**, 1065-1089.

[25] Fudenberg D. and D. K. Levine (1998) *The Theory of Learning in Games*, MIT Press.

[26] Fudenberg D. and D. K. Levine (1999) Conditional universal consistency, *Games and Economic Behavior*, **29**, 104-130.

[27] Gaunersdorfer A. and J. Hofbauer (1995) Fictitious play, Shapley polygons and the replicator equation, *Games and Economic Behavior*, **11**, 279-303.

[28] Gilboa I. and A. Matsui (1991) Social stability and equilibrium, *Econometrica*, **59**, 859-867.

[29] Hannan J. (1957) Approximation to Bayes risk in repeated plays, in Drescher M., A.W. Tucker and P. Wolfe (eds.),*Contributions to the Theory of Games, III*, Princeton University Press, 97-139.

[30] Harris C. (1998) On the rate of convergence of continuous time fictitious play, *Games and Economic Behavior*, **22**, 238-259.

[31] Hart S. (2005) Adaptive heuristics, *Econometrica*, **73**, 1401-1430.

[32] Hart S. and A. Mas-Colell (2000) A simple adaptive procedure leading to correlated equilibria, *Econometrica*, **68**, 1127-1150.

[33] Hart S. and A. Mas-Colell (2003) Regret-based continuous time dynamics, *Games and Economic Behavior*, **45**, 375-394.

[34] Hofbauer J. (1995) Stability for the best response dynamics, mimeo.

[35] Hofbauer J. (1998) From Nash and Brown to Maynard Smith: equilibria, dynamics and ESS, *Selection*, **1**, 81-88.

[36] Hofbauer J. and W. H. Sandholm (2002) On the global convergence of stochastic fictitious play, *Econometrica*, **70**, 2265-2294.

[37] Hofbauer J. and W. H. Sandholm (2009) Stable games and their dynamics, *Journal of Economic Theory*, **144**, 1665-1693.

[38] Hofbauer J. and K. Sigmund (1998) *Evolutionary Games and Population Dynamics*, Cambridge U.P.

[39] Hofbauer J. and K. Sigmund (2003) Evolutionary games dynamics, *Bulletin A.M.S.*, **40**, 479-519.

[40] Hofbauer J. and S. Sorin (2006) Best response dynamics for continuous zero-sum games, *Discrete and Continuous Dynamical Systems-series B*, **6**, 215-224.

[41] Hofbauer J., S. Sorin and Y. Viossat (2009) Time average replicator and best reply dynamics, *Mathematics of Operations Research*, **34**, 263-269.

[42] Hopkins E. (1999) A note on best response dynamics, *Games and Economic Behavior*, **29**, 138-150.

[43] Hopkins E. (2002) Two competing models of how people learn in games, *Econometrica*, **70**, 2141-2166.

[44] Hopkins E. and M. Posch (2005) Attainability of boundary points under reinforcement learning, *Games and Economic Behavior*, **53**, 110-125.

[45] Krishna V. and T. Sjöstrom (1997) Learning in games: Fictitious play dynamics, in Hart S. and A. Mas-Colell (eds.), *Cooperation: Game-Theoretic Approaches*, NATO ASI Serie A, Springer, 257-273.

[46] Krishna V. and T. Sjöstrom (1988) On the convergence of fictitious play, *Mathematics of Operations Research*, **23**, 479- 511.

[47] Laslier J.-F., R. Topol and B. Walliser (2001) A behavioral learning process in games, *Games and Economic Behavior*, **37**, 340-366.

[48] Leslie D. S. and E.J. Collins, (2005) Individual Q-learning in normal form games, *SIAM Journal of Control and Optimization*, **44**, 495-514.

[49] Littlestone N. and M.K. Warmuth (1994) The weighted majority algorithm, *Information and Computation*, **108**, 212-261.

[50] Maynard Smith J. (1982) *Evolution and the Theory of Games*, Cambridge U.P.

[51] Milgrom P. and J. Roberts (1990) Rationalizability, learning and equilibrium in games with strategic complementarities, *Econometrica*, **58**, 1255-1277.

[52] Milgrom P. and J. Roberts (1991) Adaptive and sophisticated learning in normal form games, *Games and Economic Behavior*, **3**, 82-100.

[53] Monderer D., Samet D. and A. Sela (1997) Belief affirming in learning processes, *Journal of Economic Theory*, **73**, 438-452.

[54] Monderer D. and A. Sela (1996) A 2x2 game without the fictitious play property, *Games and Economic Behavior*, **14**, 144-148.

[55] Monderer D. and L.S. Shapley (1996) Potential games, *Games and Economic Behavior*, **14**, 124-143.

[56] Monderer D. and L.S. Shapley (1996) Fictitious play property for games with identical interests, *Journal of Economic Theory*, **68**, 258-265.

[57] Pemantle R. (2007) A survey of random processes with reinforcement, *Probability Surveys*, **4**, 1-79.

[58] Posch M. (1997) Cycling in a stochastic learning algorithm for normal form games, *J. Evol. Econ.*, **7**, 193-207.

[59] Rivière P. (1997) *Quelques Modèles de Jeux d'Evolution*, Thèse, Université P. et M. Curie-Paris 6.

[60] Robinson J. (1951) An iterative method of solving a game, *Annals of Mathematics*, **54**, 296-301.

[61] Shapley L. S. (1964) Some topics in two-person games, in Dresher M., L.S. Shapley and A.W. Tucker (eds.), *Advances in Game Theory*, Annals of Mathematics **52**, Princeton U.P., 1-28.

[62] Sorin S. (2007) Exponential weight algorithm in continuous time, *Mathematical Programming*, Ser. B , **116**, 513-528.

[63] Taylor P. and L. Jonker (1978) Evolutionary stable strategies and game dynamics, *Mathematical Biosciences*, **40**, 145-156.

[64] Viossat Y. (2007) The replicator dynamics does not lead to correlated equilibria, *Games and Economic Behavior*, **59**, 397-407.

[65] Viossat Y. (2008) Evolutionary dynamics may eliminate all strategies used in correlated equilibrium, *Mathematical Social Science*, **56**, 27-43.

[66] Young P. (2004) *Strategic Learning and its Limits*, Oxford U.P. .

Combinatoire et Optimisation, IMJ, CNRS UMR 7586, Faculté de Mathématiques, Université P. et M. Curie - Paris 6, Tour 15-16, 1ière étage, 4 Place Jussieu, 75005 Paris and Laboratoire d'Econométrie, Ecole Polytechnique, France

E-mail address: sorin@math.jussieu.fr

http://www.math.jussieu.fr/ sorin/