Limit Behavior of No-regret Dynamics[*]

Andriy Zapechelnyuk[†]

University of Bonn and KSE-KEI

October 1, 2009

Abstract

Consider a repeated game where all players follow *no-regret strategies* by reinforcing the actions that they regret not having played enough in the past. We show that a resulting no-regret dynamic approaches in the long run a best-response dynamic and leads to its invariant sets: rest points (Nash equilibria) or periodic orbits. The convergence results for best-response dynamics known in the literature immediately apply to no-regret dynamics. Thus, every no-regret dynamic leads to Nash equilibrium in zero-sum games, weighted potential and two-player ordinal potential games, supermodular games with diminishing returns, and some other special classes.

Keywords: Regret minimization, no-regret strategy, best-response dynamic, Nash equilibrium, Shapley polygon, curb set

JEL classification numbers: C73, D81, D83

[*] The author thanks Sergiu Hart, Alexander Matros, Karl Schlag, and Eilon Solan for helpful comments and suggestions. The author also thanks the Center for Rationality, the Hebrew University, for its hospitality while a part of this research was being conducted, as well as the Lady Davis and Golda Meir Fellowship Funds, the Hebrew University, for their financial support.

[†] University of Bonn, Economic Theory II, Lennéstrasse 37, 53113 Bonn, Germany. *E-mail:* zapechelnyuk@hcm.uni-bonn.de

1 Introduction

How do individuals that interact in a game learn to play rationally, for instance, to play a Nash equilibrium? The traditional approach suggests that players carefully scrutinize rules of the game and incentives of opponents, and determine what strategy they should play. In practice, however, this approach is very hard to use, due to a few reasons. Firstly, there may be multiple equilibria, thus the problem of equilibrium selection arises. Secondly, the problem of finding an equilibrium may be too complex to carry out the analysis and to be confident that every opponent has carried out this analysis as well. Thirdly, players may be informed imprecisely about the opponents' utility functions, and even a small uncertainty may lead to arbitrarily large error in calculation of equilibrium behavior.

An alternative approach sees rational behavior as a result of some dynamic learning process through repeated interactions, where players make inference about opponents' future behavior on the basis of their past moves. This approach includes Bayesian learning, where a player has prior beliefs about opponents' behavior that are updated with every new observation, as well as adaptive heuristics, where players use simple adaptive rules which, nevertheless, may lead to highly sophisticated rational behavior.

No-regret strategies are simple adaptive learning rules that recently received a lot of attention (e.g., Littlestone and Warmuth, 1994; Fudenberg and Levine, 1995; Foster and Vohra, 1998, 1999; Freund and Schapire, 1999; Hart and Mas-Colell, 2000, 2001, 2003; Lehrer, 2003; Young, 2004; Cesa-Bianchi and Lugosi, 2003, 2006; Lehrer and Solan, 2009). In a repeated game, a player has a *regret* for an action if, loosely speaking, she could have obtained a greater average payoff had she played that action more often in the past. In the course of the game, the player reinforces the actions that she regrets not having played enough in the past, for instance, by choosing an action with probability proportional to the regret for that action (as in Hart and Mas-Colell's (2000) *regret matching* rule). A player's objective is to have *no regrets*, that is, to select a sequence of actions which guarantees to her, *no matter what other players do*, no regrets in the long run. A behavior rule of a player which fulfills this objective is called a *no-regret strategy.*[1]

A no-regret play is a simple adaptive behavior of an unsophisticated, myopic, non-Bayesian decision maker. It is especially appealing in a setting where every player's payoff function is private information. Existence of no-regret strategies is known since Hannan (1957); wide classes of no-regret behavior rules are identified by Hart and Mas-Colell (2001) and Cesa-Bianchi and Lugosi (2003).

How much can players learn using simple no-regret behavior rules, such as regret-matching? The only known fact (Hart and Mas-Colell, 2000) is that if all players use no-regret strategies, the long-run distribution of the joint play converges to the Hannan set, which is tautologically defined as the set of joint mixed actions where players have no regrets.[2] In many games the Hannan set is rather large, in particular, it contains (and generally, properly contains) the set of correlated equilibria and thus, a fortiori, the set of Nash equilibria (Hart and Mas-Colell, 2003). Aside of eliminating (iteratively) strictly dominated strategies, the result of convergence of the average play to the Hannan set has very little cutting power on the prediction for no-regret dynamics, as convergence to this set does not imply convergence to any particular point or orbit on this set.

[1] This paper deals with the simplest notion of regret known as *unconditional* regret (Fudenberg and Levine, 1995; Hart and Mas-Colell, 2000, 2001, 2003). For more sophisticated regret notions, see Hart and Mas-Colell (2000), Lehrer (2003), and Cesa-Bianchi and Lugosi (2006).

[2] The Hannan set of a game is the set of all mixed action profiles that satisfies the no-regret condition first stated by Hannan (1957). It is also known as the set of *coarse correlated equilibria* first appeared in Moulin and Vial (1978), but explicitly defined as a solution concept by Young (2004, Ch.3).

By a *no-regret dynamic* we understand a stochastic process that describes the trajectories of the average joint play of players and that emerges when every player follows a no-regret strategy (different players may play different strategies). In this paper we show that every no-regret dynamic approaches a *best-response dynamic*, and thus leads in the long-run to invariant sets of best-response dynamics: its rest points (Nash equilibria) or periodic orbits. The proof uses the stochastic approximation theory (Benaïm and Weibull, 2003; Benaïm et al., 2005, 2006) to relate discrete-time and continuous-time no-regret dynamics, and then shows that every limit orbit of a continuous-time no-regret dynamic is contained in the set of limit orbits of best-response dynamics.

A best-response dynamic is a continuous-time version of the well-known fictitious play (Brown, 1951; Robinson, 1951), where every player plays only best-response actions to the past average (joint) play of the others. It is shown in the literature that on many classes of games best-response dynamics lead to the set of Nash equilibria: zero-sum sum games (see Robinson, 1951; Harris, 1998), dominance solvable games (Milgrom and Roberts, 1991), generic $2 \times n$ games (Berger, 2005), weighted potential games (Monderer and Shapley, 1996), two-player ordinal potential games (Berger, 2007), supermodular games[3] with diminishing returns (Berger, 2007), and some other special classes (see, e.g., Berger, 2007; Sparrow et al., 2008). As we show that every limit orbit of a no-regret dynamic is contained in the set of limit orbits of best-reply dynamics, it follows immediately that on the above classes of games every no-regret dynamic leads to a Nash equilibrium.

A related adjustment process, stochastic fictitious play (see Fudenberg and Kreps, 1993; Hofbauer and Sandholm, 2002; Hopkins, 2002; Hofbauer and Hopkins, 2005), stipulates that in each period every player chooses only best-response actions after her payoffs are perturbed by random shocks. The

[3] Also known as games with strategic complementarities (e.g., Tirole, 1988).

presence of shocks, in effect, smoothes out the best-response correspondence of every player. As the perturbation level approaches zero, the stochastic fictitious play strategy resembles a no-regret dynamic where the best-response actions (i.e., the actions with the highest regret) are played with probability nearly one. However, we note that a no-regret dynamic is less selective and may lead to play weakly dominated actions, while the stochastic fictitious play does not (see Section 6).

Finally, we observe that a particular invariant set for a best-response dynamic need not be an attractor for a no-regret dynamic. An example is the mixed Nash equilibrium in the 2×2 coordination game, which is, normally, an unstable rest point. A natural question arises: Under what conditions will a subset of action profiles attract a no-regret dynamic? In other words, when does a no-regret stochastic process that starts close enough to a set converge with probability one to an orbit or a rest point on this set? We show that strict Nash equiliblria and, more generally, *strict curb sets* are attracting sets for a no-regret dynamic.[4]

2 Preliminaries

Let $\Gamma = (N, (A^i, u^i)_{i \in N})$ be a finite n-person game in normal form, where $N = \{1, 2, \ldots, n\}$ is a set of players, A^i is a set of actions of player i, and $u^i : A \to \mathbb{R}$ is a payoff function of player i, $A = A^1 \times \ldots \times A^n$.

The game is played repeatedly in discrete time periods $t = 1, 2, \ldots$. In every period t each player i chooses an action $a^i(t) \in A^i$ and receives payoff $u^i(a(t))$, $a(t) = (a^1(t), \ldots, a^n(t))$. Denote by $h(t) = (a(1), a(2), \ldots, a(t))$ the

[4] A product set of action profiles is called *closed under rational behavior (curb)* if it contains all best responses of each player whenever she believes that no (product mixed) actions outside this set are being played by the other players (Basu and Weibull, 1991). We use a stronger definition of curb where we require that it contain best responses to all *joint* (rather than product) mixed actions of the other players (see Section 7).

history of play up to t, and let $\mathcal{H}$ be the set of all finite histories (including the empty history). A strategy of player i is a function[5] $p^i : \mathcal{H} \to \Delta(A^i)$ that stipulates to play in every period $t = 1, 2, \ldots$ a mixed action $p^i(t) \equiv p^i(h(t-1))$ as a function of the history before t.

For every mixed joint action $z \in \Delta(A)$, let z^i and z^{-i} be the marginal distributions for player i and for the other players, respectively. Note that z need not be equal to $z^i \times z^{-i}$. We extend the definition of u^i to mixed actions in $\Delta(A)$ via statistical expectation and write $u^i(z) = \sum_{a \in A} u^i(a)z(a)$ for player i's expected payoff if mixed joint action z is played, and $u^i(k, z^{-i}) = \sum_{a^{-i} \in A^{-i}} u^i(k, a^{-i}) z^{-i}(a^{-i})$, for i's expected payoff if she deviates to a pure action $k \in A^i$ while the joint play of the others remains $z^{-i} \in \Delta(A^{-i})$.

Next, denote by $z(t) \in \Delta(A)$ the empirical distribution of play up to period t. That is, for every $a \in A$, $z_a(t)$ is the frequency of joint action a in the history up to time t,

$$z_a(t) = \frac{1}{t} \left| \{ \tau \leq t : a(\tau) = a \} \right| . \tag{1}$$

In these notations,

$$u^i(z(t)) = \frac{1}{t} \sum\nolimits_{\tau=1}^{t} u^i(a(\tau))$$

is the time-average payoff of player i up to period t.

Imagine that the objective of each player is to obtain the long-run average payoff that is at least as high as the best response against the empirical distribution of the others, i.e.,

$$\liminf_{t \to \infty} \left[u^i(z(t)) - \max_{k \in A^i} u^i(k, z^{-i}(t)) \right] \geq 0 \tag{2}$$

Denote $R^i_k(t) = u^i(k, z^{-i}(t)) - u^i(z(t))$. The term $R^i_k(t)$ is called player i's *regret for action* $k \in A^i$ and interpreted as the average gain that player i could have received had she always played k in the past instead of her

[5] We denote by $\Delta(B)$ the set of probability measures over a set B.

actual past play (given that the opponents' past actions remain unchanged). Denote by $R^i(t)$ the vector of i's regrets in period t. Thus, objective (2) can now be called *no regrets for player* i, i.e., $\limsup_{t\to\infty}[R^i(t)] \leq 0$.

We say that a strategy of player i is a no-regret strategy if it guarantees to satisfy this objective *irrespectively of what other players do.* Formally:

Definition 1 A strategy p^i of player i is a *no-regret strategy* if for every tuple p^{-i} of strategies of the other players inequality (2) holds with probability one.

The notion of *no regret* as applied to a strategy is the same as the notion of *Hannan consistency* (Hart and Mas-Colell, 2001) or *universal consistency* (Fudenberg and Levine, 1995).

It is well known in the literature starting from Hannan (1957) that there exist simple no-regret strategies. The theorem below is due to Hart and Mas-Colell (2001) who describe a wide class of *potential based* no-regret strategies (cf. Cesa-Bianchi and Lugosi, 2003).

A continuously differentiable function $P^i : \mathbb{R}^{A^i} \to \mathbb{R}$ is called a *potential* if satisfies the following conditions:

(R1) $P^i(x) \geq 0$, and $P^i(x) = 0$ for all $x \in \mathbb{R}^{A^i}_-$;

(R2) $\nabla P^i(x) \geq 0$, and $\nabla P^i_k(x) > 0$ if and only if $x_k > 0$, $k \in A^i$.

Function P^i can be viewed as a generalized differentiable distance function between a vector $x \in \mathbb{R}^{A^i}$ and the nonpositive orthant $\mathbb{R}^{A^i}_-$.

Theorem 1 (Hart and Mas-Colell, 2001) *Let P^i be a potential, and suppose that strategy p^i satisfies*

$$p^i_k(t+1) = \frac{\nabla P^i_k(R^i(t))}{\sum_{a^i\in A^i} \nabla P^i_{a^i}(R^i(t))}, \quad k \in A^i, \tag{3}$$

whenever $\max_{a^i\in A^i} R^i_{a^i}(t) > 0$. *Then p^i is a no-regret strategy.*

Notice that by condition (R2), $p^i_k(t+1) > 0$ if and only if $R^i_k(t) > 0$, or equivalently, $u^i(k, z^{-i}(t)) > u^i(z(t))$. Thus, condition (R2) is the *better response property* of p^i that stipulates to assign a positive probability only on better response actions to the opponents' empirical joint distribution of play.

We provide two examples of no-regret strategies that satisfy (3).

Example 1. The simplest example is the Hart and Mas-Colell's *regret-matching strategy* (Hart and Mas-Colell, 2000) that stipulates to play an action in the next period with probability proportional to the regret for that action, i.e., for every $k \in A^i$, whenever $\max_{a^i \in A^i} R^i_{a^i}(t) > 0$, [6]

$$p^i_k(t+1) = \frac{[R^i_k(t)]_+}{\sum_{a^i \in A^i} [R^i_{a^i}(t)]_+}.$$

Example 2. More generally, let P be the $l_{\mathbf{p}}$-norm on $\mathbb{R}^{A^i}_+$, $P(x) = (\sum_{k \in A^i} x_k^{\mathbf{p}})^{1/\mathbf{p}}$ and $1 < \mathbf{p} < \infty$. Then p^i is called the $l_{\mathbf{p}}$*-norm* strategy (Hart and Mas-Colell, 2001; Cesa-Bianchi and Lugosi, 2003) if it is defined for every $k \in A^i$ and every period $t+1$ by

$$p^i_k(t+1) = \frac{[R^i_k(t)]_+^{\mathbf{p}-1}}{\sum_{a^i \in A^i} [R^i_{a^i}(t)]_+^{\mathbf{p}-1}},$$

whenever $\max_{a^i \in A^i} R^i_{a^i}(t) > 0$. In particular, the l_2-norm strategy is equal to the regret matching strategy. For large $\mathbf{p}$, the $l_{\mathbf{p}}$-norm strategies approximate fictitious play and called *stochastic fictitious play.*[7] Note that the l_∞-norm strategy assigns probability 1 on actions with the highest regret, thus it is equivalent to the exact fictitious play (Brown, 1951) that assigns probability one on the *best* actions (not just the *better* ones). The l_∞ potential is not differentiable, and this strategy is *not* a no-regret strategy (see Young (1993) and Hart and Mas-Colell (2001) for counterexamples).

[6] We write $[c]_+$ for the positive part of a number c, i.e., $[c]_+ = \max\{c, 0\}$.

[7] The original Fudenberg and Levine's (1995) stochastic fictitious play assigns a positive probability on all (not only best response) actions, and thus it cannot be defined using a potential that satisfies the better-response condition (R2).

3 No-regret Dynamics

Suppose that every player i plays a no-regret strategy p^i given by (3) based on a potential P^i that satisfies (R1) – (R2).[8] First, observe that if player i has no regrets in some period t, $R^i(t) \leq 0$, rule (3) does not specify what she should play, her behavior is arbitrary. On the other hand, if player i has a positive regret in period t, rule (3) specifies her next-period play uniquely. We now state an important property of a no-regret strategy that if a player's maximal regret is positive in some period, it remains positive in all further periods. It will follow immediately that rule (3) defines uniquely all future behavior of player i starting from the period where she first had a positive regret.

Lemma 1 *Let player i play a no-regret strategy (3). If $\max_{a^i \in A^i} R^i_{a^i}(t) > 0$ for some period t, then $\max_{a^i \in A^i} R^i_{a^i}(t') > 0$ for all $t' > t$.*

Proof.[9] Let $\max_{a^i \in A^i} R^i_{a^i}(t) > 0$. Denote by k an action played in period $t+1$, $k = a^i(t+1)$. By (3) it follows that $\nabla P^i_k(R^i(t)) > 0$ and by (R2), $R^i_k(t) > 0$. Therefore,

$$\begin{aligned}
\max_{a^i \in A^i} R^i_{a^i}(t+1) &\geq R^i_k(t+1) \\
&= \frac{t}{t+1} R^i_k(t) + \frac{1}{t+1}(u^i(k, a^{-i}(t+1)) - u^i(k, a^{-i}(t+1))) \\
&= \frac{t}{t+1} R^i_k(t) > 0.
\end{aligned}$$

The proof is complete by induction. □

Note that Lemma 1 applies for player i's behavior regardless whether other players also follow no-regret strategies or not.

As rule (3) does not specify the play of players who have no regrets, a no-regret dynamic is well-defined only on the set of histories where every player

[8] Note that players may follow different no-regret strategies.

[9] We adapt the proof of Proposition 4.3 in Hart and Mas-Colell (2001).

has a positive regret. Thus, in what follows we shall assume that there exists an initial period t_0 where each player has a positive regret, $\max_{a^i \in A^i} R^i_{a^i}(t_0) > 0$, $i \in N$.[10]

Denote by Z the subset of joint mixed actions where for every player i at least one of i's regrets is positive,

$$Z = \left\{ z \in \Delta(A) \;\middle|\; \max_{k \in A^i} u^i(k, z^{-i}) > u^i(z) \quad \text{for all } i \in N \right\}.$$

Definition 2 A *no-regret dynamic* is a discrete-time stochastic process $\{z(t)\}_{t=1,2\ldots}$ defined by a triple (Z, z_0, p), where Z is the state space, $z_0 = z(t_0) \in Z$ is an initial state (and t_0 is an initial period), and $p = (p^1, \ldots, p^n)$ is a tuple of no-regret strategies that satisfy (3).

Denote by $\mathcal{R}$ the class of no-regret dynamics.

Since $z(t)$ can be written recursively as[11] $z(t) = \frac{1}{t}\mathbf{1}_{a(t)} + \frac{t-1}{t} z(t-1)$, its dynamic has the form

$$z(t) - z(t-1) = \frac{1}{t}\left(\mathbf{1}_{a(t)} - z(t-1)\right). \tag{4}$$

Note that by Lemma 1, the process never leaves Z.

Define the *Hannan set* H of stage game Γ as follows,

$$H = \left\{ z \in \Delta(A) \;\middle|\; \max_{k \in A^i} u^i(k, z^{-i}) \le u^i(z) \quad \text{for all } i \in N \right\}.$$

To put it differently, the Hannan set is the set of all mixed joint actions of the players where each player has no regrets. The definition of a point in H resembles Nash equilibrium, with the difference that here z is a mixed joint

[10] This assumption is not crucial, as there are many natural ways to define play in the situation of no regrets that leads to the same results as we obtain in this paper, such as playing a constant pure action, or repeating the last-period action, or playing a best-response action to the empirical distribution of the others (as in the fictitious play) (see also Hart and Mas-Colell, 2003, Appendix A).

[11] We write $\mathbf{1}_a \in \Delta(A)$ for the elementary vector that has one in a-th component and zero everywhere else.

action, rather than a product of individual mixed actions. The following properties of H are straightforward:

(i) H is a closed, convex and nonempty polytope;

(ii) H contains the set of correlated equilibria and, a fortiori, the set of Nash equilibria of game Γ;

(iii) If $z \in H$ is a product measure, i.e., $z \in \Delta(A^1) \times \ldots \times \Delta(A^n)$, then it is a Nash equilibrium.

The next property of a no-regret dynamic is straightforward by Theorem 1 (see, e.g., Hart and Mas-Colell, 2001).

Corollary 1 *Every no-regret dynamic in $\mathcal{R}$ converges to H with probability one.*

Note that convergence of the average play $z(t)$ to set H does not imply its convergence to any particular point in H. In fact, $z(t)$ may approach some orbits over H. Also, in case $z(t)$ does converge to a point in H, this point need not be a Nash equilibrium: despite the fact that period-by-period joint play of players, $p(t)$, is a product measure (i.e., $p(t) \in \Delta(A^1) \times \ldots \times \Delta(A^n)$), its trajectory may approach a complex orbit that does not yield, on average, a product measure.

We will now make the statement in Corollary 1 more precise. Observe that set Z (where $z(t)$ lives) and set H (where $z(t)$ converges to) have empty intersection. It follows that $z(t)$ converges to the intersection of H and the closure of Z.

Let $\bar{Z}$ be the closure of Z, and denote by Λ the intersection of H and $\bar{Z}$. That is,

$$\Lambda = H \cap \bar{Z} = \left\{ z \in \Delta(A) \;\middle|\; \max_{k \in A^i} u^i(k, z^{-i}) = u^i(z) \quad \text{for all } i \in N \right\}.$$

Set Λ contains all joint action profiles where every player has non-positive regrets for all actions, and exactly zero regret for at least one action.

Corollary 2 *Every no-regret dynamic in $\mathcal{R}$ converges to Λ with probability one.*

4 Convergence to Best Response Dynamics

In this section we define a continuous time best-response dynamic and its invariant sets, and state our main result that every no-regret dynamic behaves in the limit like a best-response dynamic, and hence converges to an invariant set of the latter.

Let β^i be the best-response correspondence of player i in mixed actions, given for every $y \in \Delta(A)$ by

$$\beta^i(y) = \arg\max_{q \in \Delta(A^i)} u^i(q, y^{-i}).$$

Note that $\beta^i(y)$ is functionally independent of y^i, which is included for notational convenience only. Let $\beta(y) = (\beta^1(y), \ldots, \beta^n(y))$.

A continuous time *best response dynamic* (BRD) is an absolutely continuous mapping $z : \mathbb{R}_+ \to \Delta(A)$ which is a solution of the differential inclusion

$$\dot{z} \in \beta(z) - z \tag{5}$$

with an initial condition $z(0) = z_0 \in \Delta(A)$.

This dynamic is a well known continuous time analogy of the fictitious play (Brown, 1951). In some special classes of games (e.g., zero-sum games, potential games, certain types of supermodular games) this dynamic is known to converge to a Nash equilibrium (Monderer and Shapley, 1996; Harris, 1998; Berger, 2007). However, differential inclusion (5) typically admits multiple solutions from a single initial condition, and their trajectories may be quite complex (e.g., Matsui, 1992; Sparrow et al., 2008). Here we are interested only in the invariant sets of this dynamic.

Differential inclusion (5) induces a set-valued dynamic system $\Phi^B(t) : \Delta(A) \times \mathbb{R}_+ \rightrightarrows \Delta(A)$ defined for every initial condition $z_0 \in \Delta(A)$ as the set of all solutions of (5) with that initial condition,

$$\Phi^B(z_0, t) = \{z(t) : z \text{ is a solution to (5) with } z(0) = z_0\}.$$

A set $Y \subset \Delta(A)$ is said to be *strongly invariant* for dynamic system Φ^B if $\Phi^B(Y, t) = Y$ for all $t > 0$. A set $Y \subset \Delta(A)$ is said to be *invariant* for dynamic system Φ^B if for every initial condition $z_0 \in Y$ there exists a solution z with $z(0) = z_0$ that never leaves Y, i.e., $z(t) \subset Y$ for all $t > 0$. The essential difference between these two invariance concepts is that for every initial condition z_0 in Y the former requires *every* solution that originates at z_0 to remain in Y forever, while the latter requires that for *some* solution. The latter may seem to be a very weak notion of invariance. However, it is more appropriate to describe *every* possible limit behavior of BRD, without specifying tie-breaking rules for choosing among multiple best-response actions (see also the discussion in Benaïm et al., 2005, p. 336).

We can now state our main result.

Theorem 2 *For every no-regret dynamic in $\mathcal{R}$, with probability one the average joint play $z(t)$ converges to an invariant set for the best-response dynamics.*

We defer the proof to the next section.

Invariant sets for BRD often take simple forms. So, an invariant set is a singleton (i.e., a rest point) if and only if it is a Nash equilibrium of the game. For instance, in the 2×2 coordination game (Fig. 1a) there are three invariant sets for BRD: pure action profiles (L,L) and (R,R), and the mixed Nash equilibrium that assigns equal probabilities on L and R for each player

	L	R
L	1,1	0,0
R	0,0	1,1

(a)

	L	R
L	1,-1	-1,1
R	-1,1	1,-1

(b)

Fig. 1: Coordination game (a) and Matching Pennies game (b)

(under some standard technical assumptions, the first two are attractors of BRD, while the last one is not). In the Matching Pennies game (Fig. 1b) the unique invariant set (the global attractor of BRD) is the mixed Nash equilibrium (e.g., Hopkins, 1999).

It is known that every BRD converges to a Nash equilibrium on some special classes of games, such as *dominance-solvable* games (e.g., Milgrom and Roberts, 1991), *zero-sum* games (Brown, 1951; Robinson, 1951; Harris, 1998), *potential* and *weighted potential* games (Monderer and Shapley, 1996), *two-player ordinal potential* games (Berger, 2007), *(quasi-) supermodular games with diminishing returns* (Berger, 2007), and some other special classes (see, e.g., Berger, 2007; Sparrow et al., 2008). It is immediate by Theorem 2 that on these classes every no-regret dynamic in $\mathcal{R}$ converges to a Nash equilibrium.

	R	P	S
R	0,0	1,-2	-2,1
P	-2,1	0,0	1,-2
S	1,-2	-2,1	0,0

Fig. 2: Shapley's modified Rock-Paper-Scissors game

Shapley (1964) was first to point out that in a modified Rock-Paper-Scissors (RPS) game (Fig. 2) the best response dynamic does not converge to a Nash equilibrium. Instead, it approaches some periodic orbit whose projection on the mixed action set of some player is commonly referred as the

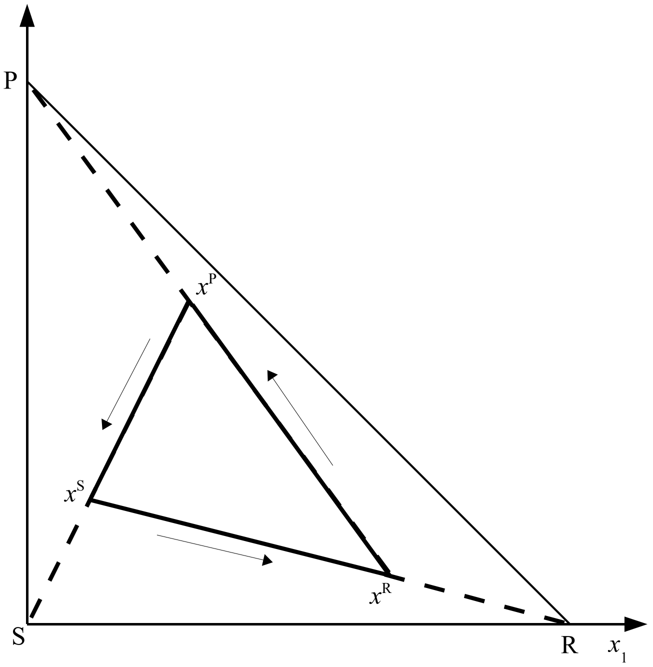

Fig. 3: Shapley triangle for the RPS game

Shapley triangle (Fig. 3). Further literature (Gaunersdorfer and Hofbauer, 1995; Sela, 2000; Benaïm et al., 2009) refers to limit orbits of BRD along which the best response correspondence is almost everywhere a singleton as *Shapley polygons*. A Shapley polygon is a cyclic orbit obtained by consecutive play of $M \geq 1$ pure action profiles in A, $a_1, a_2, \ldots, a_M$. If $M = 1$, it is a pure Nash equilibrium. Otherwise, every two consecutive action profiles a_m and a_{m+1} (index $M+1$ is identified with 1) coincide in all individual actions except for only one of the players. For that player the transition from a_m to a_{m+1} is an improvement, i.e., $u^i(a_m) < u^i(a_{m+1})$ (however, not every sequence of improvements forms a Shapley polygon (see Sela, 2000)).

5 Proof of the Main Result

The proof of our main result (Theorem 2) proceeds as follows. First, we define the continuous-time analogy of a discrete-time no-regret dynamic and use the theory of stochastic approximations to relate the limit behavior in the two settings. Second, we show that a continuous-time no-regret dynamic behaves in the limit like a best-response dynamic, and hence converges to its invariant sets.

Given a sequence $\{z(t)\}_{t=t_0,t_0+1,\ldots}$ generated by a no-regret dynamic, we say that a point $z^* \in \bar{Y}$ is a *limit point* of $\{z(t)\}$ if $\lim_{m\to\infty} z(t_m) = z^*$ for some subsequence $t_m \to \infty$. The *limit set* $L\{z(t)\}$ of $z(t)$ is the set of all its limit points. Observe that by Corollary 2, $L\{z(t)\}$ is almost surely contained in Λ.

We now define a continuous-time dynamic that corresponds to the discrete-time no-regret play and state the Limit Set Theorem (Benaïm et al., 2005, 2006) that relates limit behavior in discrete and continuous-time dynamics.

Let us extend a discrete-time no-regret dynamic to $\bar{Z}$ (the closure of Z) as in a natural way. Let $z(t_0) \in \bar{Z}$, and for every $t > t_0$ define

$$z(t) - z(t-1) \in \left\{ \frac{1}{t}(\mathbf{1}_{a(t)} - z(t-1)) \,:\, a(t) \in Q(z(t-1)) \right\}, \qquad (6)$$

where $Q(z) \subset \Delta(A^1) \times \ldots \times \Delta(A^n)$ is the set of transition rules given for every $z \in \bar{Z}$ as follows. If $z \in Z$, then $Q(z)$ is a singleton that contains only $p(z)$. For every $z \in \bar{Z} \backslash Z$, $Q(z)$ contains every distribution $q = (q_1, \ldots, q_n)$ that satisfies two conditions:

(a) q is consistent with some limit behavior of the players, i.e., there exists a sequence (x_m) on Z such that $x_m \to z$ and $p(x_m) \to q$ as $m \to \infty$;

(b) set Λ is strongly positive invariant under Q, i.e., with probability one $z(t) \in \Lambda$ entails $z(t+1) \in \Lambda$.

Since by Corollary 2 every trajectory of the dynamic, $z(t)$, almost surely converges to Λ, the actual play $p(z(t))$ almost surely converges to $Q(\Lambda)$.

Next, the extended dynamic (6) can be rewritten as

$$z(t) - z(t-1) \in \frac{1}{t} \left(F(z(t-1)) + G_t \right), \tag{7}$$

where $F : \bar{Z} \to \bar{Z}$ is the deterministic vector field given by

$$F(z(t-1)) = Q(z(t-1)) - z(t-1),$$

and G_t is the stochastic component,

$$G_t = \mathbf{1}_{a(t)} - Q(z(t-1)),$$

Consider the associated continuous-time dynamic on $\bar{Z}$ given by the following differential inclusion,[12]

$$\dot{z}(t) \in F(z(t)) \equiv Q(z(t)) - z(t). \tag{8}$$

The stochastic approximation theory (Benaïm et al., 2005, 2006) allows us to describe a limit set $L\{z_t\}$ in terms of the dynamics of the vector field F.

Consider a dynamic system $\Phi : \bar{Z} \times \mathbb{R}_+ \rightrightarrows \bar{Z}$ generated by F, where for each $y \in \bar{Z}$, $\Phi(y, \cdot)$ is the set of all solutions of differential inclusion $\dot{z} \in F(z)$ with initial condition $z(0) = y$.

A set $Y \subset \bar{Z}$ is called an *attracting set* for Φ if

(i) Y is nonempty, compact and strongly positive invariant; and

(ii) Y has a neighborhood $\mathcal{U} \subset \Delta(A)$ such that[13] $dist(\Phi(y,t), Y) \to 0$ as $t \to \infty$ uniformly in $y \in \mathcal{U} \backslash Y$.

Loosely speaking, a set is attracting if it captures the orbits of all nearby points. An attracting set $Y \subset \Delta(A)$ is called an *attractor* if Y is strongly

[12] Note that $\dot{z}(t) = (1/t)(Q(t) - z(t))$ can be translated to (8) by an appropriate time rescale, $\tilde{t} = e^t$.

[13] We denote by $dist(\cdot, \cdot)$ the Euclidean distance on $\Delta(A)$.

invariant. A set $Y \subset \Delta(A)$ is called *attractor-free* if no strict subset of Y is an attractor. Note that an attractor may or may not be attractor-free.

Benaïm et al. (2005) proved the following proposition that establishes the connection between a limit behavior of the discrete-time stochastic process $\{z(t)\}_{t=1,2...}$ and the associated continuous-time deterministic dynamic.

Proposition 1 (Limit Set Theorem) *With probability one, a limit set $L\{z(t)\}$ is a connected invariant attractor-free set for dynamic system Φ.*[14]

We now relate continuous time no-regret dynamics to BRD. Our main result, Theorem 2, will follow immediately.

Proposition 2 *Every invariant set for a no-regret dynamic system Φ is contained in an invariant set for a best-response dynamic system Φ^B.*

Proof. The proof consists of three steps. Step 1 shows that a no-regret dynamic never leaves $\bar{Z}$ and converges to Λ.

Lemma 2 *Let Φ be a dynamic system for a no-regret dynamic in $\mathcal{R}$. Then for every initial condition $z_0 \in \bar{Z}$ and every t*

(i) *$\Phi(z_0, t) \in \bar{Z}$, and*

(ii) *$Y \subset \bar{Z}$ is an invariant set for Φ only if $Y \subset \Lambda$.*

The proof of this lemma can be found in Hart and Mas-Colell (2003) and thus omitted here.

Step 2 is similar to Step 1, except that it applies to BRD instead. We show that if a BRD originates in set $\bar{Z}$, then it never leaves $\bar{Z}$ and converges to Λ. As we mentioned in Section 3, a BRD can be described as a potential-based dynamic with the l_∞-potential, however, Lemma 2 cannot be directly applied, as we have to deal with the non-differentiability of the l_∞-potential.

[14] In fact, Benaïm et al. (2005) proved a stronger result, that a limit set $L\{z(t)\}$ is *internally chain transitive*, implying that it is connected, invariant and attractor-free.

Lemma 3 *Let Φ^B be the best-response dynamic system. If $z(t_0) \in \bar{Z}$, then for every $t > t_0$*

(i) *$\Phi^B(z_0, t) \in \bar{Z}$, and*

(ii) *$Y \subset \bar{Z}$ is an invariant set for Φ^B only if $Y \subset \Lambda$.*

Proof. To simplify notations, we will further suppress t whenever possible. For every $i \in N$ and every $k \in A^i$ we have

$$\begin{aligned} \dot{R}^i_k(z) &\equiv u^i(k, \dot{z}^{-i}) - u^i(\dot{z}) = u^i(k, p^{-i} - z^{-i}) - u^i(p - z) \\ &= (u^i(k, p^{-i}) - u^i(p)) - (u^i(k, z^{-i}) - u^i(z)) \\ &= u^i(k, p^{-i}) - u^i(p) - R^i_k(z). \end{aligned} \tag{9}$$

Multiplying by p^i and summing over $k \in A^i$ yields

$$p^i \cdot \dot{R}^i(z) = -p^i \cdot R^i(z). \tag{10}$$

Next, define for every $z \in \bar{Z}$

$$\pi^i(z) = \max_{a^i \in A^i} R^i_{a^i}(z).$$

Recall that p^i assigns a positive probability only on best response actions w.r.t. z^{-i}, i.e., $p^i \in \beta^i(z)$, or equivalently,

$$p^i_k > 0 \text{ only if } k \in \arg\max_{a^i \in A^i} R^i_{a^i}(z),\ k \in A^i.$$

Hence (10) can be rewritten as

$$p^i \cdot \dot{R}^i(z) = -\max_{a^i \in A^i} R^i_{a^i}(z) = -\pi^i(z).$$

Note that the above equality holds for every $p^i \in \beta^i(z)$. Hence, the directional derivative of $\pi^i(z)$ in every direction $p^i \in \beta^i(z)$ satisfies

$$\lim_{\varepsilon \downarrow 0} \frac{\pi^i(z + \varepsilon p^i) - \pi^i(z)}{\varepsilon} = -\pi^i(z).$$

It follows that $\dot{\pi}^i(z) = \frac{d}{dt}\pi^i(z(t))$ exists, and moreover

$$\dot{\pi}^i(z) = -\pi^i(z). \tag{11}$$

Let $\pi(z) = \sum_{i \in N} \pi^i(z)$. Observe that $\pi^i(z) \geq 0$ on $\bar{Z}$ for every i, and $\pi(z) = 0$ if and only if $z \in \Lambda$. Also, by (11), $\dot{\pi}(z) = -\pi(z)$. Consequently, π is a strict Lyapunov function for set Λ of the dynamic on $\bar{Z}$. Part (i) follows from the observation that $\pi^i(z(t)) \geq 0$ for all i and all t whenever $z(0) \in \bar{Z}$, and hence the dynamic never leaves $\bar{Z}$. Part (ii) follows by $\pi(z(t)) \to 0$, and hence $z(t) \to \Lambda$. □

We established that every invariant set of either a no-regret dynamic or a best-response dynamic on $\bar{Z}$ is contained in Λ. We now show that the set of solution trajectories of a no-regret dynamic restricted to Λ is contained in the set of solution trajectories of BRD. It follows immediately that every invariant set for a no-regret dynamic is contained in an invariant set for BRD.

Lemma 4 *Let Φ be a dynamic system for a no-regret dynamic in $\mathcal{R}$ and let Φ^B be the BRD system. Then for every $z_0 \in \Lambda$ and every $t > 0$, $\Phi(z_0, t) \subset \Phi^B(z_0, t)$.*

Proof. Consider no-regret dynamic system Φ. For every $z \in \Lambda$ each player i has no regrets, with regrets for some actions exactly equal to zero. In fact, $R^i_k(z) = 0$ if and only if k is a best response action to z^{-i}. Since $R^i(\cdot)$ is continuous, in every ε-neighborhood of z, denoted by $U_\varepsilon(z) \subset \bar{Z}$, only best response actions to z^{-i} may have a strictly positive regret, and at least one of them does. Thus by (3) and condition (R2), for every $y_\varepsilon \in U_\varepsilon(z)$, $p^i(y_\varepsilon)$ has support only on the best-response actions, i.e., $p^i(y_\varepsilon) \in \beta^i(z)$. Taking the limit $\varepsilon \to 0$, we have $y_\varepsilon \to z$ and $p^i(y_\varepsilon) \to Q^i(z) \subset \beta^i(z)$. Hence,

$$\dot{z} \in Q^i(z) - z \subset \beta^i(z) - z.$$

Consequently, the trajectories of Φ on Λ are contained in those of Φ^B. □.

6 Stochastic Fictitious Play

One may wonder how the stochastic fictitious play relates to the no-regret dynamics. The original fictitious play (Brown, 1951; Robinson, 1951) requires a player to choose *best* response actions (to the average play of the opponents), rather than *better* ones, like no-regret strategies do. Due to the discontinuity of the fictitious play strategy, in some cases a behavior which is inconsistent with best-responding has been observed (see, e.g., the example in Young 2004, Ch.6). To get around the problem, Fudenberg and Kreps (1993) proposed to smooth out the best response choice problem by adding small shocks to the observed average payoffs. Formally, we define the stochastic (or smooth) fictitious play as follows. Let $\varepsilon(t) \in A^i$ be an i.i.d. sequence of vectors of shocks, $t = 1, 2, \ldots$. For every period t, let player i choose an action k that maximizes $u^i(k, z^{-i}) + \varepsilon_k$, i.e., let i choose a best response against the average past play of the others, when the payoffs are disturbed by realized shocks. Then the probability to choose action $k \in A^i$ in the next period is given by

$$p^i_k(t+1) = \Pr\left(\arg\max_{a^i \in A^i}[u^i(a^i, z^{-i}) + \varepsilon_{a^i}] = k\right).$$

The most well known stochastic fictitious play formula is described by the logistic function (see, e.g., McFadden, 1974; Blume, 1993, 1997) obtained by assuming that the shocks are i.i.d. with the extreme value distribution that yields for every $k \in A^i$

$$p^i_k(t+1) = \frac{e^{u^i(k,z^{-i})/\eta}}{\sum_{a^i \in A^i} e^{u^i(a^i,z^{-i})/\eta}}. \tag{12}$$

The parameter $\eta \in (0, \infty)$ is interpreted as the level of noise. This strategy approaches the unperturbed best response as η approaches zero. Observe that this strategy on its own is does not lead to no regrets, as it will assign a positive probability on actions with negative regret when the maximal regret approaches zero (that is, the potential for strategy (12) violates condition (R2)). However, Fudenberg and Levine (1995) show that the stochastic fictitious play with noise level η guarantees ε-regrets, where ε depends on η and approaches zero as $\eta \to 0$.

Let us look at the limit behavior of an SFP dynamic, where the level of noise for the SFP strategies of all players approaches zero. Similarly to our result, such a dynamic will approach a best-response dynamic and converge to its invariant sets (Hofbauer and Sandholm, 2002). However, an SFP is more selective than a no-regret dynamic: some invariant sets that may be approached by a no-regret dynamic are never approached by an SFP. For example, an SFP never converges to weakly dominated actions, and the dominance solvable games have a unique attractor for an SFP, the Nash equilibrium that remains after iterated elimination of all (weakly) dominated strategies. To the contrary, a no-regret dynamic may converge to a weakly dominated Nash equilibrium.

	L	M	R
L	2,2	1,1	0,0
M	1,1	1,1	1,1
R	0,0	1,1	0,0

Fig. 4: A dominance solvable game.

For illustration, consider the following identical-interest game (Fig. 4). This game is dominance solvable. After elimination of dominated action R for each player, action M becomes dominated by L, and (L, L) is the unique solution. However, another Nash equilibrium, (M, M), may be a limit of a no-regret dynamic. Suppose that both players play no-regret strategies and assume that by chance they played (R, R) in period 1. After period 1 each player has a positive regret for action M (as she could have obtained 1 instead of zero by having played M) and a zero regret for actions L and R.

As any no-regret strategy requires to choose actions with positive regret (as long as there are such), the play in period 2 will be (M, M) with probability one. The regrets for actions L and R remain zero (as by playing any of those actions instead of the actual past play would yield exactly the same payoff), and the regret for M remains positive (as each player still could have gotten a greater average payoff by having played M in period 1). This holds for every further period, and thus (M, M) will be played forever.

Finally, we would like to remark that a no regret dynamic may converge to weakly dominated actions, but never to strictly dominated ones. Indeed, if action j is strictly dominated by action k, the regret for action k is always strictly greater than that for j, and hence in the limit, as the dynamic approaches no-regret set H, the regret for j is strictly negative.

7 Attracting Sets for No-regret Dynamics

Theorem 2 asserts that every no-regret dynamic leads to the collection of invariant sets for BRD. But does every such invariant set contain limit points or orbits of a no-regret dynamic? In general, it is not true. For instance, in the 2 × 2 coordination game, every BRD (and therefore every no-regret dynamic) from almost every initial condition leads to a pure Nash equilibrium, while the mixed Nash equilibrium is unstable (under some standard technical conditions, e.g., Hopkins, 1999).

We provide below a sufficient condition for an invariant set for BRD to be an attracting set for a no-regret dynamic.

Clearly, a strict Nash equilibrium is an absorbing point of a no-regret play: if a strict Nash equilibrium $a \in A$ is played long enough, any deviation of player i to $k \in A^i$, $k \neq a^i$, will generate a positive regret for not playing a^i, thus reinforcing the play of the equilibrium action in the future. Let us now consider a generalization of strict Nash equilibrium concept to product subsets, so called sets that are closed under rational behavior (curb).

Let $\mathcal{B}$ be the collection of all non-empty product subsets of A, i.e., $B \in \mathcal{B}$ if and only if B is a Cartesian product of nonempty subsets $B^i \subset A^i$, $i = 1, \ldots, n$. For a product set $B \in \mathcal{B}$ denote by $\Delta_B(A)$ the set of all probability measures on A with support on B only,

$$\Delta_B(A) = \{y \in \Delta(A) : y(a) = 0 \text{ for all } a \in A \backslash B\}.$$

A set $B \in \mathcal{P}$ is said to be *strict closed under rational behavior* (*strict curb*) if $\beta(\Delta_B(A)) \subset \Delta_B(A)$. That is, a set B is strict curb if players' best responses are contained in B whenever they believe that no (joint) actions outside of B should be played. Obviously, if B is a singleton, then it is a strict Nash equilibrium. A strict curb set is said to be *minimal* if it does not contain any proper subset which is a strict curb set.

Note that our concept of a *strict curb* set is the same in spirit as Basu and Weibull's (1991) *curb*, but it sets somewhat stronger requirements. A set $B \in \mathcal{B}$ is curb (in the sense of Basu and Weibull) if for every player i, B^i contains all best responses of i to every *product mixed action* of $N \backslash \{i\}$ with support on B^{-i}, while B is strict curb if B^i contains all best responses to every *joint mixed action* of $N \backslash \{i\}$ with support on B^{-i}. For two-player games these two concepts coincide; for games with three and more players, every strict curb set is curb, but the converse need not be true.

Being closely related, strict curb sets inherit most properties of curb sets (Basu and Weibull, 1991). Firstly, in a game Γ there exists at least one minimal strict curb set. Secondly, every strict curb set contains the support of at least one Nash equilibrium, in particular, a singleton strict curb set contains a strict Nash equilibrium. Thirdly, every minimal strict curb set B is 'tight', i.e., $\beta(\Delta_B(A)) = \Delta_B(A)$. Fourthly, minimal strict curb sets are disjoint. Finally, sets of Nash equilibria contained in minimal strict curb sets satisfy some of the strongest known set-wise refinement criteria: Kohlberg and Mertens' (1986) hyperstability and strategic stability (see Ritzberger and Weibull, 1995).

Note that, as every minimal curb set is equal to or contained in some minimal strict curb set, all dynamics or adjustment processes that converge to the former (Hurkens, 1995; Young, 1998; Matros, 2003) converge to the latter as well.

Proposition 3 *Every minimal strict curb set is an attracting set for every BRD and every no-regret dynamic in $\mathcal{R}$.*

Proof. Let $B \in \mathcal{B}$ be a minimal strict curb set. Then there exists a number $\gamma > 0$ such that for every mixed joint action profile $x \in \Delta_B(A)$, every player i and every action $a^i \notin B^i$, $u^i(x) - u^i(a^i, x^{-i}) > \gamma$. Then there exists an open neighborhood of $\Delta_B(A)$, $U_\gamma(B) \subset \Delta(A)$, where any action outside of B^i is a "worse" response for every player i. Formally, $x \in U_\gamma(B)$ implies that for all $i \in N$ and all $a^i \notin B^i$

$$u^i(a^i, x^{-i}) < u^i(x).$$

Thus, if the initial condition $z(t_0)$ is in $U_\gamma(B)$, then for every player i her best responses are contained in B^i, and hence every BRD will remain in $U_\gamma(b)$ and converge to $\Delta_B(A)$. Also, every action outside of B^i will have a strictly negative regret for i. Thus, every no-regret dynamic will select only actions in B. $\square$.

The following corollary is immediate.

Corollary 3 *A collection of invariant sets for BRD contained in some (minimal) strict curb set is an attracting set for every no-regret dynamic in $\mathcal{R}$. Furthermore, if a strict curb set contains a unique invariant set for BRD, that set is an attractor.*

We conclude by providing an example where a set that is not (strict) curb may attracts a no-regret dynamic. Consider the following game (Fig. 5). For

	H	T	O
H	3,0	0,3	0,4
T	0,3	3,0	0,-2
O	4,0	-2,0	-1,-1

Fig. 5: An Matching Pennies game with an outside option.

each player, action O is strictly dominated by the mixed action that assigns probability 1/2 to H and T. The game that remains after elimination of action O, $B = ((H, T), (H, T))$ is the Matching Pennies game, with the mixed Nash equilibrium being the global attractor for every BR dynamic (e.g., Hopkins, 1999), and hence for every no-regret dynamic. However, product subset B is not a curb set, as action O is a best response to H.

References

Basu, K. and J. W. Weibull (1991). Strategy subsets closed under rational behavior. *Economic Letters 36*, 141–146.

Benaïm, M., J. Hofbauer, and E. Hopkins (2009). Learning in games with unstable equilibria. *Journal of Economic Theory 144*, 1694–1709.

Benaïm, M., J. Hofbauer, and S. Sorin (2005). Stochastic approximations and differential inclusions. *SIAM Journal on Control and Optimization 44*, 328–348.

Benaïm, M., J. Hofbauer, and S. Sorin (2006). Stochastic approximations and differential inclusions. Part II: Applications. *Mathematics of Operations Research 31*, 673–695.

Benaïm, M. and J. Weibull (2003). Deterministic approximation of stochastic evolution in games. *Econometrica 71*, 873–903.

Berger, U. (2005). Fictitious play in $2 \times n$ games. *Journal of Economic Theory 120*, 139–154.

Berger, U. (2007). Two more classes of games with the continuous-time fictitious play property. *Games and Economic Behavior 60*, 247–261.

Blume, L. E. (1993). The statistical mechanics of strategic interaction. *Games and Economic Behavior 5*, 387–424.

Blume, L. E. (1997). Population games. In W. B. Arthur, S. N. Durlauf, and D. A. Lane. (Eds.), *The Economy as an Evolving Complex System II*, pp. 425–460. Addison-Wesley.

Brown, G. (1951). Iterative solutions of games by fictitious play. In T. Koopmans (Ed.), *Activity Analysis of Production and Allocation*, Volume 13 of *Cowles Commission Monograph*, pp. 374–376. New York: Wiley.

Cesa-Bianchi, N. and G. Lugosi (2003). Potential-based algorithms in on-line prediction and game theory. *Machine Learning 51*, 239–261.

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

Foster, D. and R. Vohra (1998). Asymptotic calibration. *Biometrika 85*, 379–390.

Foster, D. and R. Vohra (1999). Regret in the online decision problem. *Games and Economic Behavior 29*, 7–35.

Freund, Y. and R. Schapire (1999). Adaptive game playing using multiplicative weights. *Games and Economic Behavior 29*, 79–103.

Fudenberg, D. and D. M. Kreps (1993). Learning mixed equilibria. *Games and Economic Behavior 5*, 320–367.

Fudenberg, D. and D. Levine (1995). Universal consistency and cautious fictitious play. *Journal of Economic Dynamics and Control 19*, 1065–1089.

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

Hannan, J. (1957). Approximation to Bayes risk in repeated play. In M. Dresher, A. W. Tucker, and P. Wolfe (Eds.), *Contributions to the Theory of Games, Vol. III*, Annals of Mathematics Studies 39, pp. 97–139. Princeton University Press.

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

Hart, S. and A. Mas-Colell (2000). A simple adaptive procedure leading to correlated equilibrium. *Econometrica 68*, 1127–1150.

Hart, S. and A. Mas-Colell (2001). A general class of adaptive procedures. *Journal of Economic Theory 98*, 26–54.

Hart, S. and A. Mas-Colell (2003). Continuous-time regret-based dynamics. *Games and Economic Behavior 45*, 375–394.

Hofbauer, J. and E. Hopkins (2005). Learning in perturbed asymmetric games. *Games and Economic Behavior 52*, 133–152.

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

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

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

Hurkens, S. (1995). Learning by forgetful players. *Games and Economic Behavior 11*, 304–329.

Lehrer, E. (2003). A wide range no-regret theorem. *Games and Economic Behavior 42*, 101–115.

Lehrer, E. and E. Solan (2009). Approachability with bounded memory. *Games and Economic Behavior 66*, 995–1004.

Littlestone, N. and M. Warmuth (1994). The weighted majority algorithm. *Information and Computation 108*, 212–261.

Matros, A. (2003). Clever agents in adaptive learning. *Journal of Economic Theory 111*, 110–124.

Matsui, A. (1992). Best response dynamics and socially stable strategies. *Journal of Economic Theory 57*, 343–362.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *RAND Journal of Economics 25*, 242–262.

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

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

Moulin, H. and J. P. Vial (1978). Strategically zero-sum games: the class of games whose completely mixed equilibria cannot be improved upon. *International Journal of Game Theory 7*, 201–221.

Ritzberger, K. and J. W. Weibull (1995). Evolutionary selection in normal-form games. *Econometrica 63*, 1371–1399.

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

Sela, A. (2000). Fictitious play in 2×3 games. *Games and Economic Behavior 31*, 152–162.

Shapley, L. S. (1964). Some topics in two person games. In M. Dresher, L. S. Shapley, and A. W. Tucker (Eds.), *Advances in Game Theory*, pp. 1–28. Princeton University Press.

Sparrow, C., S. van Strien, and C. Harris (2008). Fictitious play in 3 × 3 games: The transition between periodic and chaotic behaviour. *Games and Economic Behavior 63*, 259–291.

Tirole, J. (1988). *The Theory of Industrial Organization.* MIT Press.

Young, H. P. (1993). The evolution of conventions. *Econometrica 61*, 57–84.

Young, H. P. (1998). *Individual Strategy and Social Structure.* Princeton University Press.

Young, H. P. (2004). *Strategic Learning and Its Limits.* Oxford University Press.