14 December, 1996

A Simple Adaptive Procedure Leading to Correlated Equilibrium[1]

by

Sergiu Hart[2] **and Andreu Mas-Colell**[3]

Abstract

We propose a simple adaptive procedure for playing a game. In this procedure, players depart from their current play with probabilities that are proportional to measures of regret for not having used other strategies (these measures are updated every period). It is shown that our adaptive procedure guaranties that with probability one, the sample distributions of play converge to the set of correlated equilibria of the game. To compute these regret measures, a player needs to know his payoff function and the history of play. We also offer a variation where every player knows only his own realized payoff history (but not his payoff function).

---

[1] Previous version: March 1996 (handout). Research partially supported by grants of: the U.S.-Israel Binational Science Foundation; the Israel Academy of Sciences and Humanities; the Spanish Ministry of Education; and the Generalitat de Catalunya. We want to acknowledge useful discussions with Dean Foster.

[2] Department of Economics; Department of Mathematics; and Center for Rationality and Interactive Decision Theory, The Hebrew University of Jerusalem, Givat-Ram, 91904 Jerusalem, Israel; *e-mail:* hart@math.huji.ac.il.

[3] Department of Economics, Universitat Pompeu Fabra, 08005 Barcelona, Spain.

1. Introduction

The leading non-cooperative equilibrium notions for N-person games in strategic (normal) form are Nash equilibrium (and its refinements) and correlated equilibrium. In this paper we focus on the concept of correlated equilibrium.

A *correlated equilibrium* –– a notion introduced by Aumann [1974] –– can be described as follows: Assume that, before the game is played, each player receives a private signal (which does not affect the payoffs). The player may then choose his action in the game depending on this signal. A correlated equilibrium of the original game is just a Nash equilibrium of the game with the signals. If the signals are (stochastically) independent across the players, this is no different than a Nash equilibrium (in mixed or pure strategies) of the original game. But the signals could well be correlated, in which case one may get new equilibria.

Equivalently[4], a correlated equilibrium is a probability distribution on N-tuples of actions, which can be interpreted as the distribution of play instructions given to the players (by some device). Each player is given –– privately –– instructions for his own play only; the distribution is known to all of them. Also, for every possible instruction that a player receives, the player agrees that the instruction provides a best response to the random

[4] This is sometimes called the "canonical" representation of a correlated equilibrium, or the correlated equilibrium "distribution"; we need not make this distinction explicit here.

estimated play of the other players — assuming that all the other players always follow their instructions.

There is much to be said for correlated equilibrium. See Aumann [1974, 1987] for an analysis and foundational arguments in terms of rationality. Also, from a practical point of view, it could be argued that correlated equilibrium may be the most relevant non-cooperative solution concept. Indeed, with the possible exception of well controlled environments, it may be hard to eliminate a priori the possibility that correlating signals be amply available to the players, and thus find their way into the equilibrium.

The set of correlated equilibria is, in contrast to the set of Nash equilibria, geometrically simple: it is a convex set (actually, with finitely many strategies, a convex polytope) of distributions. It includes the Nash equilibria. Hart and Schmeidler [1989] (see also Nau and McCardle [1990]) provided an elementary (non-fixed point) proof of the non-emptiness of the set of correlated equilibria. This was done by using the Minimax Theorem. Specifically, they proceeded by associating to the given N-person game an auxiliary two-person zero-sum game. As it turns out, the correlated equilibria of the original game correspond to the maximin strategies of player I in the auxiliary game. More precisely, in the Hart-Schmeidler auxiliary game, player I chooses a distribution over N-tuples of actions, and player II chooses an ordered pair of strategies for one of the N original players (interpreted as a play and a suggested deviation from it). The payoff to auxiliary player II is the expected gain of the designated original player if

he were to follow the change suggested by auxiliary player[5] II. In other words, it is the "regret" of that original player of not deviating.

In this paper we move to dynamic considerations. We pose the following question: Are there simple adaptive procedures always leading to correlated equilibrium? Note that we impose a very demanding requirement –– global convergence –– but for a relatively simple, from the mathematical point of view, solution concept. In contrast, this demand would not be reasonable for Nash equilibria, a mathematically more complex notion (as shown by the essential need to resort to fixed point theorems in order to prove existence).

A procedure with the desired properties was suggested by Foster and Vohra [1996]. Fudenberg and Levine [1995] offered a second one. Here we propose a procedure which we view as particularly simple and intuitive. The starting point for our research was the observation that fictitious play applied to the Hart-Schmeidler auxiliary game must converge, by the result of Robinson [1951], and thus yield optimal strategies in the auxiliary game, in particular for player I –– hence correlated equilibria of the original game. A direct application of this idea does not, however, produce anything simple. Yet, our adaptive procedure is based on "no-regret" ideas motivated by this analysis and it is the direct descendant –– several modifications later –– of this research line. Like fictitious play, it is relatively simple and does not entail any sophisticated updating, prediction or fully rational behavior.

---

[5] For a further use of this approach, see Myerson [1995].

In Section 2 we present the model and describe the adaptive procedure proposed. It takes place in discrete time and it specifies that players adjust strategies probabilistically. This adjustment is guided by "regret measures" based on observation of past periods. Players know the past history of play of all players, as well as their own payoff matrix (but not necessarily the payoff matrices of the other players). Our Main Theorem, also stated in Section 2, is: The adaptive procedure generates trajectories of play that almost surely converge to the set of correlated equilibria.

In Section 3 we present a "stylized variation" of the procedure of Section 2. It is a variation that lends itself to a very direct proof, using what turns out to be a key instrument: Blackwell's [1956a] Approachability Theorem.

The proof of the Main Theorem is in Section 4.

In Section 5 we consider the case where players do *not* know the game — neither their own payoff matrix nor the other players' payoff matrices. Each player observes (and remembers) only his own realized past actions and payoffs. The play (and the payoffs) of the other players are not observed. We present a modification[6] of the adaptive procedure of Section 2. Again, the dynamics leads to the set of correlated equilibria. Section 6 includes the proof of the result of Section 5.

---

[6] The modification is based on a suggestion of Dean Foster; see Foster and Vohra [1993] and Auer *et al* [1995].

Finally, an Appendix deals with a related topic: we show that adaptive procedures of the kind presented in this paper apply also to a simpler no-regret problem originally studied by Hannan [1957].

2. The Model and Main Result

Let $\Gamma = (N, (S^i)_{i\in N}, (u^i)_{i\in N})$ be an N-person finite game in strategic (normal) form: N is the set of players, S^i is the set of strategies of player i, and $u^i : \prod_{i\in N} S^i \to \mathbb{R}$ is player i-th payoff function. All sets N and S^i are assumed to be finite. Denote by $S := \prod_{i\in N} S^i$ the set of N-tuples of strategies; the generic member of S is $s = (s^i)_{i\in N}$, and s^{-i} denotes the strategy combination of all players except i, i.e., $s^{-i} = (s^{i'})_{i'\neq i}$. We focus attention on the following solution concept:

Definition: A probability distribution ψ on S is a *correlated equilibrium* of Γ if, for every $i \in N$, every $j \in S^i$ and every $k \in S^i$ we have:

$$\sum_{s\in S:s^i=j} \psi(s)\ [u^i(k,s^{-i}) - u^i(s)] \leq 0.$$

If in the above definition we replace in the right-hand side 0 by an $\varepsilon > 0$, then we obtain the concept of a *correlated* ε - *equilibrium*.

Note that every Nash equilibrium is a correlated equilibrium. Indeed, Nash equilibria correspond to the special case where ψ is a product measure, that is, the play of the different players is independent. Also, the set of correlated equilibria is non-empty, closed and convex, and even in simple games (e.g., "chicken") it may include distributions that are not in the convex hull of the Nash equilibrium distributions.

Suppose now that the game Γ is played repeatedly through time: $t = 1, 2, \ldots$. At time t+1, given a history $h_t = (s_\tau)_{\tau=1}^{t} \in \prod_{\tau=1}^{t} S$, we postulate that each player $i \in N$ chooses $s^i_{t+1} \in S^i$ according to a probability distribution[7] $p^i_{t+1} \in \Delta(S^i)$ which is defined as follows:

For each t and every $j \in S^i$ and $k \in S^i$ with $j \neq k$, let

$$A^i_t(j,k) := \begin{cases} u^i(k,s^{-i}_t) - u^i(s_t), & \text{if } s^i_t = j, \\ 0, & \text{otherwise;} \end{cases}$$

$$D^i_t(j,k) := \frac{1}{t} \sum_{\tau=1}^{t} A^i_\tau(j,k); \quad \text{and}$$

$$R^i_t(j,k) := [D^i_t(j,k)]^+ \equiv \max\{D^i_t(j,k), 0\}.$$

The expressions $R^i_t(j,k)$ have clear interpretations as measures of the (average) regret at t of not having played, every time that j has been played

[7] We write $\Delta(Q)$ for the set of probability distributions over a finite set Q.

in the past, the different strategy k. Indeed, let U be the sum of the payoffs that player i received in all the periods (up to and including t) where he played j; let V be the sum, over the same periods, of the payoffs he would have received if he would have chosen k instead (while the other players would not have changed their choices). If V is smaller than U, then there is no regret — $R^i_t(j,k)$ is zero. If V is larger than U, then the regret $R^i_t(j,k)$ is the difference (per period) between V and U, i.e. (V – U)/t.

Let $\mu > 0$ be a large enough number; specifically, it suffices to take μ so that $1/\mu < 2\, M^i\, (m^i - 1)$ for all $i \in N$, where M^i is an upper bound for $|u^i(\cdot)|$ and[8] $m^i := |S^i|$ (even better: $1/\mu < (m^i - 1)\, [u^i(k,s^{-i}) - u^i(j,s^{-i})]$ for all $j, k \in S^i$, all $s^{-i} \in S^{-i}$ and all[9] $i \in N$). Suppose that at time t player i played strategy $j \in S^i$, i.e., $s^i_t = j$. Then we define the random play $p^i_{t+1} \in \Delta(S^i)$ at t+1 by:

$$\begin{cases} p^i_{t+1}(k) := \dfrac{1}{\mu}\, R^i_t(j,k), & \text{for } k \neq j, \\ p^i_{t+1}(j) := 1 - \displaystyle\sum_{k \in S^i : k \neq j} p^i_{t+1}(k), & \end{cases} \tag{2.1}$$

Note that the choice of μ guarantees that $p^i_{t+1}(j) > 0$; that is, there is always a positive probability of playing the same strategy as in the previous period. The play $p^i_1 \in \Delta(S^i)$ at the initial period is chosen arbitrarily.[10]

---

[8] We write $|Q|$ for the number of elements of a finite set Q.

[9] One may also use different values of μ for the different players.

[10] Actually, the procedure could start with any finite number of periods where the play is arbitrary.

For every t, let $z_t \in \Delta(S)$ be the empirical distribution of the N-tuple of strategies played up to time t. That is, for every $s \in S$,

$$z_t(s) := \frac{1}{t} \left| \{\tau \leq t : s_\tau = s\} \right| \tag{2.2}$$

is the relative frequency that the N-tuple s has been played in the first t periods. We can now state our main result.

> **Main Theorem:** If every player plays according to the adaptive procedure described by (2.1), then the empirical distributions z_t of play converge almost surely as $t \to \infty$ to the set of correlated equilibrium distributions of the game Γ.

More explicitly, the conclusion of the Main Theorem asserts that the following statement holds with probability one: For any $\varepsilon > 0$ there is $T = T(\varepsilon)$ such that for all $t > T$ we can find a correlated equilibrium distribution ψ_t at a distance less than ε from z_t. (Note that this T depends on the history; it is an "a.s. finite stopping time"). That is, the Main Theorem says that, with probability one, for any $\varepsilon > 0$, the (random) trajectory $(z_t)_{t=1,2,...}$ enters and stays in the ε-neighborhood in $\Delta(S)$ of the set of correlated equilibria. Yet in other words: Given any $\varepsilon > 0$, there exists a *constant* $T_0 = T_0(\varepsilon)$ such that, with probability at least $1-\varepsilon$, the

empirical distributions z_t for *all* $t > T_0$ are in the ε-neighborhood of the set of correlated equilibria. Finally, let us note that because the set of correlated equilibrium is non-empty and compact, the statement "the trajectory z_t converges to the set of correlated equilibria" is equivalent to the statement "the trajectory z_t is such that for any $\varepsilon > 0$ there is $t_0 = t_0(\varepsilon)$ with the property that z_t is a correlated ε-equilibrium for all $t > t_0$".

To get some intuition on our result, note first that each player observes the same history h_t of moves of all players; this is the common information that leads to the (possible) correlation. In the fictitious play process, a player looks at the empirical distribution of the choices of his opponents (and plays optimally against it). In our procedure he "allows" for correlation, and thus looks separately at the periods where he played one strategy or another of his.

A few comments are now in order:

(1) Our adaptive procedure (2.1) requires player i to know his own payoff matrix (but not those of the other players) and, at time t+1, the history h_t (actually, the empirical distribution z_t of $(s_1, s_2, ..., s_t)$ suffices). In terms of the computation of the play probabilities, it is enough for player i to remember at t+1 the $m^i(m^i-1)$ numbers $D_t^i(j,k)$ for all $j \neq k$ in S^i (and then update them after t+1 using s_{t+1}).

(2) At every period in time the adaptive procedure that we propose randomizes only over the strategies which exhibit positive regret relative to

the most recently played strategy. Some strategies may, therefore, receive zero probability. Suppose that we were to allow for trembles. Specifically, suppose that at every period we put a $\delta > 0$ probability on the uniform tremble (each strategy is thus played with probability at least δ/m^i). The result still holds with the following two modifications: First, the empirical distribution z_t converges to the set of correlated ε-equilibria (of course, ε depends on δ, and it goes to zero as δ goes to zero); and second, the convergence of the distance is only in probability (rather than almost surely).

(3) It is a significant and important fact that our adaptive procedure guarantees that staying put is always played with positive probability. Suppose that we were to modify the procedure by requiring that the switching probabilities are scaled in such a way that a switch will *necessarily* occur whenever there is at least one strategy exhibiting some positive regret from the most recent play (thus: one "stays put" only when all regrets are zero). Then the result of the Main Theorem may not hold. Consider for example the simplest two-person coordination game:

	a	b
A	1 , 1	0 , 0
B	0 , 0	1 , 1

.

If we start with uncoordinated strategies –– for instance, (A,b) in the first period –– then the play will alternate between the two uncoordinated pairs (A,b) and (B,a). However, no distribution concentrated at (A,b) and (B,a) is a correlated equilibrium (instructions according to such a distribution reveal to each player the instruction given to the other player, and induce therefore responses which are precisely the opposite of the instructions received).

(4) The convergence properties induced by our adaptive procedure are not knife-edge in nature. There are other procedures enjoying a similar property: the procedures of Foster and Vohra [1996], of Fudenberg and Levine [1995], and of Section 3 below. On the other hand, as we saw in the previous observation, not every procedure will have it. The delimitation of general classes of procedures converging to correlated equilibria seems, therefore, an interesting research problem.

(5) We know little about additional convergence properties for z_t. Note that convergence to the *set* of correlated equilibria does not imply that the sequence z_t converges. It may or it may not. It is easy to see that if the set of correlated equilibria has a non-empty interior, then any interior point which is the limit of a trajectory z_t must be a pure strategy Nash equilibrium. Thus there are many games with some correlated equilibria which are unreachable as the limit of some z_t (but it is still possible that they be reachable as limits of a *subsequence* of z_t). More generally, it is easily seen that the empirical distributions z_t either converge to pure Nash equilibria or must be infinitely often outside the set of correlated equilibria.

(6) It is worth pointing out that if a player's last choice was j, then the relative probabilities of switching to k or to k′ do not depend only on the average utilities that would have been obtained if j had been changed to k or to k′ in the past, but also on the average utility that was obtained in those periods by playing j itself (it is the magnitude of the increases in moving from j to k or to k′ that matters). This is in contrast with the procedure of Fudenberg and Levine [1995] (or, in the context of the Appendix, of the "scoring" rules of Auer *et al* [1995]). Formally, our procedure is "best-reply invariant": transformations of the payoff matrix u^i that preserve the best-reply correspondence[11] do not affect it. Note that the procedure of Foster and Vohra [1996] also satisfies this.

(7) Our adaptive procedure depends only on one parameter,[12] μ. This may be viewed as an "inertia" parameter: a higher μ yields smaller probabilities of switch. The convergence to the set of correlated equilibria is always guaranteed, but the speed of convergence changes (by a constant factor) with μ.

---

[11] I.e., $v^i(s) = \alpha\, u^i(s) + \beta(s^{-i})$ for $\alpha > 0$ (note that μ needs to be replaced by $\alpha\,\mu$).

[12] Using a parameter μ (rather than a fixed normalization of the payoffs) was suggested to us by Reinhard Selten.

3. No Regret and Blackwell Approachability

This section may be viewed as a motivational preliminary. What we shall do is to replace the adaptive procedure of Section 2 by another procedure that, while related to it, is more stylized. Then we shall analyze it by means of Blackwell's [1956a] Approachability Theorem, and prove that it yields convergence to the set of correlated equilibria. In fact, the Main Theorem stated in Section 2, and its proof in Section 4, were inspired by consideration and careful study of the result of this section. Still, the procedure of this section is interesting in its own right; see for instance the Remark following the statement of Theorem A.

Fix a player i and recall the procedure of Section 2: At time t+1 the transition probabilities, from the strategy played by player i in period t to the strategies to be played at t+1, are determined by the stochastic matrix defined by the system (2.1). Consider now an invariant probability vector $q_t^i = (q_t^i(j))_{j \in S^i} \in \Delta(S^i)$ for this matrix (such a vector always exists). That is, q_t^i satisfies

$$q_t^i(j) = \sum_{k \neq j} q_t^i(k) \frac{1}{\mu} R_t^i(k,j) + q_t^i(j) \left[1 - \sum_{k \neq j} \frac{1}{\mu} R_t^i(j,k)\right],$$

for every $j \in S^i$. Collecting terms, multiplying by μ, and formally letting $R_t^i(j,j) := 0$, the above expression can be rewritten as:

$$\sum_{k \in S^i} q_t^i(k)\, R_t^i(k,j) = q_t^i(j) \sum_{k \in S^i} R_t^i(j,k), \qquad (3.1)$$

for every $j \in S^i$.

In this section we shall assume that play at time t+1 by player i is determined by a solution q_t^i to the system of equations (3.1); i.e., $p_{t+1}^i(j) := q_t^i(j)$. In a sense, we assume that player i at time t+1 goes instantly to the invariant distribution of the stochastic transition matrix determined by (2.1). We now state the key result. Similar results have been proven by Foster and Vohra [1991, 1996] and by Fudenberg and Levine [1995].

> **Theorem A:** Suppose that at every period t+1 player i chooses strategies according to a probability vector q_t^i that satisfies (3.1). Then the regrets $R_t^i(j,k)$ converge to zero almost surely for every $j \neq k$ in S^i.

Remark: It is important to note that in Theorem A no assumption is made on how players different from i choose their strategies (except for the fact that for every t, given the history up to t, play is independent among players). In the terminology of Fudenberg and Levine [1995], the invariant probability adaptive procedure of this section is "universally conditionally consistent". This is a definite benefit of using the invariant probability. In comparison, the adaptive procedure of Section 2 does not have the property of guaranteeing that the regrets of a player go to zero regardless of how

other players play. The Main Theorem asserts that if *all* players follow the adaptive procedure, then the regrets of *all* players go to zero.

What is the connection between regrets and correlated equilibria? It turns out that a necessary and sufficient condition for the empirical distributions to converge to the set of correlated equilibria is precisely that all regrets converge to zero. More generally:

Proposition: Let $\varepsilon \geq 0$ and let $(s_t)_{t=1,2,...}$ be a sequence of plays (i.e., $s_t \in S$ for all t) such that $\limsup_{t\to\infty} R_t^i(j,k) \leq \varepsilon$ for every $i \in N$ and every $j, k \in S^i$ with $j \neq k$. Then the sequence of empirical distributions z_t (defined by (2.2)) converges to the set of correlated ε-equilibria.

Proof: Let $\psi \in \Delta(S)$ be an accumulation point of the sequence z_t of empirical distributions of play along this history. Consider an arbitrary player i and arbitrary strategies $j \neq k$ in S^i. Then

$$D_t^i(j,k) = \frac{1}{t} \sum_{\tau \leq t : s_\tau^i = j} [u^i(k, s_\tau^{-i}) - u^i(j, s_\tau^{-i})]$$

$$= \sum_{s \in S : s^i = j} z_t(s) \, [u^i(k, s^{-i}) - u^i(j, s^{-i})].$$

This becomes in the limit $\leq \varepsilon$ (it is bounded above by $R_t^i(j,k)$). Taking the limit on the subsequence where the empirical distributions converge to ψ yields:

$$\sum_{s\in S:s^i=j} \psi(s)\ [u^i(k,s^{-i}) - u^i(j,s^{-i})] \leq \varepsilon.$$

So $\psi \in \Delta(S)$ constitutes a correlated ε-equilibrium. ||

Theorem A and the Proposition (for $\varepsilon = 0$) immediately imply

> **Corollary:** Suppose that at each period t+1 every player i chooses strategies according to a probability vector q_t^i that satisfies (3.1). Then the empirical distribution of play $z_t \in \Delta(S)$ converges a.s. to the set of correlated equilibria of the game.

Before addressing the formal proof of Theorem A we make a brief presentation of Blackwell's Approachability Theorem (or, rather, of the part of it that is relevant to the current work).

The basic setup contemplates a decision maker i with a (finite) action set S^i. For a finite indexing set L, the decision maker receives an $|L|$-dimensional vector payoff $v(s^i,s^{-i}) \in \mathbb{R}^L$ that depends on his action $s^i \in S^i$ and on some external action s^{-i} belonging to a set S^{-i}. The decision problem is repeated through time and the issue is which average vector payoffs

$$\frac{1}{t}\sum_{\tau=1}^{t} v(s_\tau) \equiv \frac{1}{t}\sum_{\tau=1}^{t} v(s^i_\tau, s^{-i}_\tau)$$

could the decision maker i guarantee himself in the long-run by appropriate choice of his actions $s^i_t \in S^i$ at every t.

Figure 1 about here

Let $\mathcal{C}$ be a non-empty, closed and convex subset of $\mathbb{R}^L$. For every $D \in \mathbb{R}^L$, $D \notin \mathcal{C}$, denote by $F(D) \in \mathcal{C}$ the "foot" of D in $\mathcal{C}$, that is, the (unique) vector in $\mathcal{C}$ closest to D in the Euclidean distance. Suppose it to be the case that for every $D \notin \mathcal{C}$ the decision maker i could resort to a mixed strategy (of the one-shot decision problem) $q(D) \in \Delta(S^i)$, having the property that[13]

$$\left[\sum_{s^i \in S^i} [q(D)](s^i)\ v(s^i, s^{-i})\right] \bullet \left[D - F(D)\right] \leq 0, \tag{3.2}$$

for every $s^{-i} \in S^{-i}$. Inequality (3.2) — called the *Blackwell condition* — says that no matter what value we take for s^{-i} in S^{-i}, the expected one-shot payoff resulting from q(D) will lie in the half-space $\mathcal{H}$ illustrated in Figure 1. Blackwell's Approachability Theorem then tells us that if at every t+1 the decision maker i randomizes among his strategies according to[14] $q(D_t)$,

[13] For two vectors x and y, their scalar product is denoted $x \bullet y$. We write $[q(D)](s^i)$ for the s^i-coordinate of q(D), i.e., the probability given by the mixed strategy q(D) to s^i.

[14] If $D_t \in \mathcal{C}$, any choice at time t+1 will do.

for $D_t := (1/t)\ \Sigma_{\tau \le t}\ v(s_\tau)$, it follows that the average payoff vector D_t will approach the set C with probability one. Figure 2 suggests why this is so: If condition (3.2) is satisfied and t is large then (we write $\mathbf{E}[\cdot \,|\, h_t]$ for the conditional expectation given the history h_t up to time t)

$$\mathbf{E}[D_{t+1} \,|\, h_t] = \frac{t}{t+1}\ D_t + \frac{1}{t+1}\ \mathbf{E}[v(s^i_{t+1}, s^{-i}_{t+1})]$$

will be *inside* the circle of center $F(D_t)$ and radius $\|D_t - F(D_t)\|$. Hence ("dist" denotes Euclidean distance)

$$\text{dist}(\mathbf{E}[D_{t+1} \,|\, h_t],\ C) \le \|\mathbf{E}[D_{t+1} \,|\, h_t] - F(D_t)\| < \|D_t - F(D_t)\| = \text{dist}(D_t,\ C)$$

(the first inequality follows from the fact that $F(D_t)$ is in C). A precise computation actually shows that the distance not only decreases, but it actually goes to zero.[15]

Figure 2 about here

Proof of Theorem A: As mentioned, the proof of this Theorem consists of an application of Blackwell's Approachability Theorem. Let

$$L := \{(j,k) \in S^i \times S^i : j \ne k\},$$

[15] Note that one looks here at *expected* average payoffs; the law of large numbers (for dependent random variables -- ee.g., see Loève [1978, 32.1.E]) implies that the *actual* average payoffs also converge to the set C.

and define the vector payoff $v(s^i,s^{-i}) \in \mathbb{R}^L$ (for player i) by letting its (j,k) $\in$ L coordinate be:

$$[v(s^i,s^{-i})](j,k) := \begin{cases} u^i(k,s^{-i}) - u(j,s^{-i}), & \text{if } s^i = j, \\ 0, & \text{otherwise.} \end{cases}$$

Let C be the non-positive orthant $\mathbb{R}^L_-$ of $\mathbb{R}^L$ (this is a non-empty, closed, convex subset of $\mathbb{R}^L$). For any vector $D \notin C$, let $R := [D]^+$ be its positive part (that is, $R(j,k) = \max\{D(j,k), 0\}$ for all $j \neq k$) and suppose that $q(D) \in \Delta(S^i)$ is an invariant vector for the non-negative matrix R; again, put formally $R(j,j) = 0$) (such a q(D) always exists). That is, q(D) satisfies

$$\sum_{k \in S^i} [q(D)](k)\ R(k,j) = [q(D)](j) \sum_{k \in S^i} R(j,k), \tag{3.3}$$

for every $k \in S^i$.

Figure 3 about here

We proceed to show that the probability vectors q(D) satisfy Blackwell's condition (3.2) for the set $C = \mathbb{R}^L_-$. Note first that, for any $D \notin C$, we have $D - F(D) = [D]^+ = R$; see Figure 3. Then, for any $s^{-i} \in S^{-i}$, the expression in condition (3.2) takes the form

$$\sum_{(j,k)\in L} \sum_{s^i\in S^i} [q(D)](s^i)\ [v(s^i,s^{-i})](j,k)\ R(j,k) \le 0.$$

That is,

$$\sum_{(j,k)\in L} [q(D)](j)\ [u^i(k,s^{-i}) - u^i(j,s^{-i})]\ R(j,k) \le 0. \tag{3.4}$$

After collecting terms, the left hand side of (3.4) can be written as

$$\sum_{j\in S^i} \alpha(j)\ u^i(j,s^{-i}), \tag{3.5a}$$

where

$$\alpha(j) := \sum_{k\in S^i} [q(D)](k)\ R(k,j) - [q(D)](j) \sum_{k\in S^i} R(j,k). \tag{3.5b}$$

By (3.3) we have therefore that $\alpha(j) = 0$ for all $j \in S^i$ and so, inequality (3.4) holds true (as an equality[16]) for all $s^{-i} \in S^{-i}$. Thus Blackwell's condition (3.2) is satisfied. Because, for all t+1, the mixed strategy p^i_{t+1} played by player i at t+1, is precisely $q(D^i_t)$, it follows that $(1/t)\ \Sigma^t_{\tau=1}\ v(s_\tau)$ approaches $C = \mathbb{R}^L_-$ a.s.; that is, $R^i_t(j,k) = \max\{(1/t)\ \Sigma^t_{\tau=1}\ [v(s_\tau)](j,k),\ 0\}$ goes to zero a.s., for every $j \neq k$. ||

---

[16] Note that this is precisely formula (2) in the Proof of Theorem 1 in Hart and Schmeidler [1989].

Remark: Blackwell's Approachability Theorem provides also bounds on the speed of convergence. In our case, we get the following: The expectation $\mathbf{E}[R_t^i(j,k)]$ of the regrets is of the order of $t^{-1/2}$, and the probability $\mathbf{P}[R_t^i(j,k) > \varepsilon$ for some i in N, some $j \neq k$ in S^i and some $t > T]$ that z_t is *not* a correlated ε-equilibrium for *some* $t > T$, is of the order 1/T (see Mertens, Sorin and Zamir [1994, Theorem 4.3]). Clearly, a better speed of convergence[17] cannot be guaranteed, since for instance, if the other players play stationary mixed strategies, then the errors are of the order $t^{-1/2}$ by the Central Limit Theorem.

4. Proof of the Main Theorem

We return now to our procedure, as defined in Section 2. The proof of the Main Theorem is inspired by the result of Section 3 (Theorem A). It is however rather complex[18] on account of our transition probabilities not being the invariant measures that, as we saw in Section 3, fitted so well with Blackwell's Approachability Theorem.

Fix player i in N. To simplify notation we drop reference to the index i whenever this cannot cause confusion (thus we write D_t and R_t instead of D_t^i and R_t^i, and so on). Let $m := |S^i|$ be the number of strategies of player i, and let M be an upper bound on i-th possible payoffs: $M \geq |u^i(s)|$ for all s in S. Denote $L := \{(j,k) \in S^i \times S^i : j \neq k\}$, then $\mathbb{R}^L$ is the

---

[17] Up to a constant factor.
[18] We only said that the adaptive *procedure* is simple.

m(m–1) - dimensional Euclidean space with coordinates indexed by L. For each t = 1, 2, ... and each (j,k) in L, recall that[19]

$$A_t(j,k) = \mathbf{1}_{\{s^i_t=j\}} \, [u^i(k,s^{-i}_t) - u^i(s_t)];$$

$$D_t(j,k) = \frac{1}{t} \sum_{\tau=1}^{t} A_\tau(j,k);$$

$$R_t(j,k) = D^+_t(j,k) \equiv [D_t(j,k)]^+.$$

We will write A_t for the vector $(A_t(j,k))_{j\neq k} \in \mathbb{R}^L$; the same goes for D_t, D^+_t, R_t, and so on. Let $\Pi_t(\cdot,\cdot)$ denote the transition probabilities from t to t+1 (these are computed after period t, based on h_t):

$$\Pi_t(j,k) := \begin{cases} \frac{1}{\mu} R_t(j,k), & \text{if } k \neq j, \\ 1 - \sum_{k'\neq j} \frac{1}{\mu} R_t(j,k'), & \text{if } k = j. \end{cases}$$

Thus, at time t+1 the strategy used by player i is to choose each $k \in S^i$ with probability $p^i_{t+1}(k) = \Pi_t(s^i_t,k)$. Note that the choice of μ guarantees that $\Pi_t(j,j) > 0$ for all $j \in S^i$ and all t. Finally, let

$$\rho_t := [\text{dist}(D_t, \mathbb{R}^L_-)]^2$$

[19] We write $\mathbf{1}_G$ for the indicator of the event G.

be the squared distance (in $\mathbb{R}^L$) of the vector D_t to the non-positive orthant $\mathbb{R}^L_-$. Since the closest point to D_t in $\mathbb{R}^L_-$ is[20] D_t^-, we have $\rho_t = \|D_t - D_t^-\|^2 = \|D_t^+\|^2 = \sum_{j \neq k} [D_t^+(j,k)]^2$.

It will be convenient to use the standard "O" notation: For two real-valued functions $f(\cdot)$ and $g(\cdot)$ defined on a domain X, "$f(x) = O(g(x))$" means that there exists a constant $K < \infty$ such that $|f(x)| \leq K\, g(x)$ for all x in[21] X. We write **P** for Probability, and **E** for Expectation. From now on, t, v, and w will denote positive integers; $h_t = (s_\tau)_{\tau \leq t}$ will be histories of length t; j, k and s^i will be elements of S^i; s and s^{-i} will be elements of S and S^{-i}, respectively. Unless stated otherwise, all statements should be understood to hold "for all t, v, h_t, j, k, etc."; when histories h_t are concerned, only those that occur with positive probability are considered.

We divide the proof of the Main Theorem into 11 steps, M1 - M11. The final Step M11, together with the Proposition of Section 3, completes the proof. The reader will find a short "intuitive guide" to the proof following the statements of the steps.

• *Step M1:* $\mathbf{E}[(t+v)^2\, \rho_{t+v} | h_t] = t^2\, \rho_t + 2t \sum_{w=1}^{v} R_t \bullet \mathbf{E}[A_{t+w} | h_t] + O(v^2)$.

---

[20] We write $[x]^-$ for $\min\{x, 0\}$, and D_t^- for the vector $([D_t(j,k)]^-)_{(j,k) \in L}$.

[21] The domain X will usually be the set of positive integers, or the set of vectors whose coordinates are positive integers. Thus when we write, say, $f(t,v) = O(v)$, it means $|f(t,v)| \leq K\, v$ for all v and all t. The constants K will always depend only on the game (through N, m, M, and so on) and on the parameter μ.

Define

$$\alpha_{t,w}(j,s^{-i}) := \sum_{k \in S^i} \Pi_t(k,j)\ \mathbf{P}[s_{t+w} = (k,s^{-i}) \,|\, h_t] - \mathbf{P}[s_{t+w} = (j,s^{-i}) \,|\, h_t].$$

- *Step M2:* $R_t \bullet \mathbf{E}[A_{t+w} \,|\, h_t] = \mu \sum_{s^{-i} \in S^{-i}} \sum_{j \in S^i} \alpha_{t,w}(j,s^{-i})\ u^i(j,s^{-i})$.

- *Step M3:* $R_{t+v}(j,k) - R_t(j,k) = O(\frac{v}{t})$.

For each $t > 0$ and each history h_t, define an auxiliary stochastic process $(\hat{s}_{t+w})_{w=0,1,2,...}$ with values in S as follows: The initial value is $\hat{s}_t = s_t$, and the transition probabilities are[22]

$$\mathbf{P}[\hat{s}_{t+w} = s \,|\, \hat{s}_t, ..., \hat{s}_{t+w-1}] := \prod_{i' \in N} \Pi_t^{i'}(\hat{s}^{i'}_{t+w-1}, s^{i'}).$$

The $\hat{s}$-process is thus stationary: it uses the transition probabilities of period t at each period t+w, for all $w \geq 0$.

- *Step M4:* $\mathbf{P}[s_{t+w} = s \,|\, h_t] - \mathbf{P}[\hat{s}_{t+w} = s \,|\, h_t] = O(\frac{w^2}{t})$.

---

[22] We write $\Pi_t^{i'}$ for the transition probability matrix of player i′ (thus Π_t is Π_t^i).

Define

$$\hat{\alpha}_{t,w}(j,s^{-i}) := \sum_{k \in S^i} \Pi_t(k,j)\ \mathbf{P}[\hat{s}_{t+w} = (k,s^{-i}) \,|\, h_t] - \mathbf{P}[\hat{s}_{t+w} = (j,s^{-i}) \,|\, h_t].$$

- *Step M5:* $\alpha_{t,w}(j,s^{-i}) - \hat{\alpha}_{t,w}(j,s^{-i}) = O(\frac{w^2}{t})$.

- *Step M6:* $\hat{\alpha}_{t,w}(j,s^{-i}) = \mathbf{P}[\hat{s}^{-i}_{t+w} = s^{-i} \,|\, h_t]\ [\Pi_t{}^{w+1} - \Pi_t{}^{w}](s^i_t, j)$,

where $\Pi_t{}^{w} \equiv (\Pi_t)^w$ is the w-th power of the matrix Π_t, and $[\Pi_t{}^{w+1} - \Pi_t{}^{w}](s^i_t, j)$ denotes the (s^i_t, j) element of the matrix $\Pi_t{}^{w+1} - \Pi_t{}^{w}$.

- *Step M7:* $\hat{\alpha}_{t,w}(j,s^{-i}) = O(w^{-1/2})$.

- *Step M8:* $\mathbf{E}[(t+v)^2\ \rho_{t+v} \,|\, h_t] = t^2\ \rho_t + O(v^3 + t\ v^{1/2})$.

For each n = 1, 2, ..., let $t_n := \lfloor n^{5/3} \rfloor$ be the largest integer not exceeding $n^{5/3}$.

- *Step M9:* $\mathbf{E}[t^2_{n+1}\ \rho_{t_{n+1}} \,|\, h_{t_n}] = t^2_n\ \rho_{t_n} + O(n^2)$.

- *Step M10:* $\mathbf{E}[\rho_{t_n}] = O(n^{-1/3})$ and $\lim_{n \to \infty} \rho_{t_n} = 0$ a.s.

• *Step M11:* $\lim_{t\to\infty} R_t(j,k) = 0$ a.s.

We now provide a short intuitive guide to the proof. Step M1 is the basic recursion equation, similar to that used for the standard proof of Blackwell's Approachability Theorem. However, we need to consider v periods together, rather than just one. The reader may want to keep in mind that both t and v are very large, and that v is relatively small compared to t. This implies that, in the right-hand expression of M1, only the middle term has to be bounded (from above). Blackwell's condition (3.2) precisely says that the middle term in the corresponding formula is ≤ 0. Here this is not so (we are *not* in the stylized version of Section 3), so we need to estimate this term. Step M2 shows that we get an expression similar to (3.5), but the coefficients α depend also on the moves of the other players. Given h_t, the choices s^i_{t+w} and s^{-i}_{t+w} are *not* independent when w > 1 (since the transition probabilities change with time). Therefore we replace the process $(s_{t+w})_{0\le w\le v}$ by another process $(\hat{s}_{t+w})_{0\le w\le v}$, with a *stationary* transition matrix (that of period t). For w small relative to t, the change in probabilities is small (see Steps M3 and M4), yielding the approximation of Step M5. In Step M6 we factor out the moves of the other players (which, in the $\hat{s}$-process, are independent of the moves of player i) from the coefficients $\hat{\alpha}$ and obtain a formula comparable to (3.5). However, instead of the invariant distribution, we have the transition probabilities after w periods and after w+1 periods. The difference between them is shown in Step M7 to be small, since w is large and the transition matrix has all its diagonal elements strictly positive[23]. This yields the

[23] For further discussion on this point, see the Proof of Step M7.

final recursive formula of Step M8. The proof is now completed in Steps M9 - M11, by considering a carefully chosen subsequence $(t_n)_{n=1,2,...}$.

The rest of this section contains the proofs of the Steps M1 - M11.

• *Proof of Step M1:* Because $D_t^- \in \mathbb{R}_-^L$ we have:

$$\rho_{t+v} \le \|D_{t+v} - D_t^-\|^2 = \left\| \frac{t}{t+v} D_t + \frac{1}{t+v} \sum_{w=1}^{v} A_{t+w} - D_t^- \right\|^2$$

$$= \frac{t^2}{(t+v)^2} \|D_t - D_t^-\|^2 + \frac{2t}{(t+v)^2} \sum_{w=1}^{v} (A_{t+w} - D_t^-) \bullet (D_t - D_t^-)$$

$$+ \frac{v^2}{(t+v)^2} \left\| \frac{1}{v} \sum_{w=1}^{v} A_{t+w} - D_t^- \right\|^2$$

$$\le \frac{t^2}{(t+v)^2} \rho_t + \frac{2t}{(t+v)^2} \sum_{w=1}^{v} A_{t+w} \bullet R_t + \frac{v^2}{(t+v)^2} m (m-1) 4 M^2.$$

Indeed: $|u^i(s)| \le M$, so $|A_{t+w}(j,k)|$, $|D_t(j,k)| \le 2M$, yielding the upper bound on the third term. As for the second term, note that $R_t = D_t^+ = D_t - D_t^-$ and $D_t^- \bullet D_t^+ = 0$. Taking conditional expectation given the history h_t (so ρ_t and R_t are known) completes the proof. ||

• *Proof of Step M2:* We have

$$\mathbf{E}[A_{t+w}(j,k)\,|\,h_t] = \sum_{s^{-i}} \phi(j,s^{-i})\ [u^i(k,s^{-i}) - u^i(j,s^{-i})],$$

where $\phi(j,s^{-i}) := \mathbf{P}[s_{t+w} = (j,s^{-i})\,|\,h_t]$. So

$$R_t \bullet \mathbf{E}[A_{t+w}\,|\,h_t] = \sum_j \sum_{k\neq j} R_t(j,k) \sum_{s^{-i}} \phi(j,s^{-i})\ [u^i(k,s^{-i}) - u^i(j,s^{-i})]$$

$$= \sum_{s^{-i}} \sum_j u^i(j,s^{-i}) \Big[\sum_{k\neq j} R_t(k,j)\ \phi(k,s^{-i}) - \sum_{k\neq j} R_t(j,k)\ \phi(j,s^{-i}) \Big]$$

(we have collected together all terms containing $u^i(j,s^{-i})$). Now $R_t(k,j) = \mu\ \Pi_t(k,j)$ for $k \neq j$, and $\Sigma_{k\neq j}\ R_t(j,k) = \mu\ (1 - \Pi_t(j,j))$ by definition, so

$$R_t \bullet \mathbf{E}[A_{t+w}\,|\,h_t] = \mu \sum_{s^{-i}} \sum_j u^i(j,s^{-i}) \Big[\sum_k \Pi_t(k,j)\ \phi(k,s^{-i}) - \phi(j,s^{-i}) \Big]$$

(note that the last sum is now over *all* k in S^i). ||

- *Proof of Step M3:* This follows immediately from

$$(t+v)\ [D_{t+v}(j,k) - D_t(j,k)] = \sum_{w=1}^{v} A_{t+w}(j,k) - v\ D_t(j,k),$$

together with $|A_{t+w}(j,k)|$, $|D_t(j,k)| \leq 2M$. ||

• *Proof of Step M4:* We need the following Lemma, which gives bounds for the changes in the w-step transition probabilities as a function of changes in the 1-step transitions:

Lemma: Let $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ be two stochastic processes with values in a finite set B. Assume $X_0 = Y_0$ and

$$|\mathbf{P}[X_n = b_n | X_0 = b_0, \ldots, X_{n-1} = b_{n-1}] - \mathbf{P}[Y_n = b_n | Y_0 = b_0, \ldots, Y_{n-1} = b_{n-1}]| \leq \beta_n$$

for all $n \geq 1$ and all $b_0, \ldots, b_{n-1}, b_n \in B$. Then

$$|\mathbf{P}[X_{n+w} = b_{n+w} | X_0 = b_0, \ldots, X_{n-1} = b_{n-1}] - \mathbf{P}[Y_{n+w} = b_{n+w} | Y_0 = b_0, \ldots, Y_{n-1} = b_{n-1}]| \leq |B| \sum_{r=0}^{w} \beta_{n+r}$$

for all $n \geq 1$, $w \geq 0$ and all $b_0, \ldots, b_{n-1}, b_{n+w} \in B$.

Proof of Lemma: We write $\mathbf{P}_X$ and $\mathbf{P}_Y$ for the probabilities of the two processes $(X_n)_n$ and $(Y_n)_n$, respectively (thus $\mathbf{P}_X[b_{n+w} | b_0, \ldots, b_{n-1}]$ stands for $\mathbf{P}[X_{n+w} = b_{n+w} | X_0 = b_0, \ldots, X_{n-1} = b_{n-1}]$, and so on). We use induction on w.

$$\mathbf{P}_X[b_{n+w}|b_0, \ldots, b_{n-1}] = \sum_{b_n} \mathbf{P}_X[b_{n+w}|b_0, \ldots, b_n]\ \mathbf{P}_X[b_n|b_0, \ldots, b_{n-1}]$$

$$\leq \sum_{b_n} \mathbf{P}_Y[b_{n+w}|b_0, \ldots, b_n]\ \mathbf{P}_X[b_n|b_0, \ldots, b_{n-1}] + |B| \sum_{r=1}^{w} \beta_{n+r}$$

$$\leq \sum_{b_n} \mathbf{P}_Y[b_{n+w}|b_0, \ldots, b_n]\ (\mathbf{P}_Y[b_n|b_0, \ldots, b_{n-1}] + \beta_n) + |B| \sum_{r=1}^{w} \beta_{n+r}$$

$$\leq \mathbf{P}_Y[b_{n+w}|b_0, \ldots, b_{n-1}] + |B|\ \beta_n + |B| \sum_{r=1}^{m} \beta_{n+r}$$

(the first inequality is by the induction hypothesis). Exchanging the roles of X and Y completes the proof. ||

We proceed now with the proof of Step M4. From t to t+w there are $|N|$ w transitions (at each period, think of the players moving one after the other, in some arbitrary order). Step M3 implies that each transition probability for the $\hat{s}$-process differs from the corresponding one of the s-process by at most O(w/t), which, when composed by the Lemma, yields $|N|$ w $|S|$ O(w/t) = $O(w^2/t)$. ||

- *Proof of Step M5:* Immediate by Step M4. ||

- *Proof of Step M6:* Given h_t, the random variables $(\hat{s}^{i'}_{t+w})_w$ are independent over the different players i′ in N; indeed, the transition

probabilities are all determined at time t, and the players randomize independently. Hence:

$$\mathbf{P}[\hat{s}_{t+w} = (j,s^{-i}) \,|\, h_t] = \mathbf{P}[\hat{s}^{-i}_{t+w} = s^{-i} \,|\, h_t]\ \mathbf{P}[\hat{s}^{i}_{t+w} = j \,|\, h_t],$$

implying that

$$\hat{\alpha}_{t,w}(j,s^{-i}) = \mathbf{P}[\hat{s}^{-i}_{t+w} = s^{-i} \,|\, h_t] \left[\sum_{k \in S^i} \Pi_t(k,j)\ \mathbf{P}[\hat{s}^{i}_{t+w} = k \,|\, h_t] - \mathbf{P}[\hat{s}^{i}_{t+w} = j \,|\, h_t] \right].$$

Now $\mathbf{P}[\hat{s}^{i}_{t+w} = j \,|\, h_t]$ is the probability of reaching j in w steps starting from s^i_t, using the transition probability matrix Π_t. Therefore $\mathbf{P}[\hat{s}^{i}_{t+w} = j \,|\, h_t]$ is the (s^i_t,j)-element of the w-th power $\Pi_t^w \equiv (\Pi_t)^w$ of Π_t, i.e., $[\Pi_t^w](s^i_t,j)$. Hence

$$\hat{\alpha}_{t,w}(j,s^{-i}) = \mathbf{P}[\hat{s}^{-i}_{t+w} = s^{-i} \,|\, h_t] \left[\sum_{k \in S^i} \Pi_t(k,j)\ [\Pi_t^w](s^i_t,k) - [\Pi_t^w](s^i_t,j) \right]$$

$$= \mathbf{P}[\hat{s}^{-i}_{t+w} = s^{-i} \,|\, h_t] \left[[\Pi_t^{w+1}](s^i_t,j) - [\Pi_t^w](s^i_t,j) \right],$$

completing the proof. ||

• *Proof of Step M7:* It follows from M6 using the following Lemma[24] (recall that $\Pi_t(j,j) > 0$ for all $j \in S^i$).

Lemma: Let Π be an $m \times m$ stochastic matrix with all of its diagonal entries positive. Then $[\Pi^{w+1} - \Pi^w](j,k) = O(w^{-1/2})$ for all $j, k = 1, ..., m$.

Proof of the Lemma: Let $\beta > 0$ be a lower bound on all the diagonal entries of Π, i.e., $\beta := \min_j \Pi(j,j)$. We can then write $\Pi = \beta I + (1-\beta) \Lambda$, where Λ is also a stochastic matrix. Now

$$\Pi^w = \sum_{r=0}^{w} \binom{w}{r} \beta^{w-r} (1-\beta)^r \Lambda^r,$$

and similarly for Π^{w+1}. Subtracting yields

$$\Pi^{w+1} - \Pi^w = \sum_{r=0}^{w+1} \gamma_r \binom{w}{r} \beta^{w-r} (1-\beta)^r \Lambda^r,$$

where $\gamma_r := \beta(w+1)/(w+1-r) - 1$. Now $\gamma_r > 0$ if $r > q := (w+1)(1-\beta)$, and $\gamma_r \leq 0$ if $r \leq q$; together with $0 \leq \Lambda^r(j,k) \leq 1$, we get

---

[24] If Π were a strictly positive matrix, this would be a standard result, because then Π^w would converge to the invariant matrix. However we only know that the diagonal elements are positive. This implies that, if w is large, then with high probability there will be a positive fraction of periods when the process does not move. But this number is random, so the probabilities of going from j to k in w steps or in w+1 steps should be almost the same (since it is like having r "stay put" transitions versus r+1).

$$\sum_{r\le q} \gamma_r \binom{w}{r} \beta^{w-r} (1-\beta)^r \le [\Pi^{w+1} - \Pi^w](j,k) \le \sum_{r>q} \gamma_r \binom{w}{r} \beta^{w-r} (1-\beta)^r.$$

Consider the left-most sum. It equals

$$\sum_{r\le q} \binom{w+1}{r} \beta^{w+1-r} (1-\beta)^r - \sum_{r\le q} \binom{w}{r} \beta^{w-r} (1-\beta)^r = G_{w+1}(q) - G_w(q),$$

where $G_n(\cdot)$ denotes the cumulative distribution function of a sum of n independent Bernoulli random variables, each one having the value 0 with probability β and the value 1 with probability $1-\beta$. Using the normal approximation yields (Φ denotes the standard normal cumulative distribution function):

$$G_{w+1}(q) - G_w(q) = \Phi(x) - \Phi(y) + O(\frac{1}{\sqrt{(w+1)}}) + O(\frac{1}{\sqrt{w}}),$$

where

$$x := \frac{q-(w+1)(1-\beta)}{\sqrt{(w+1)\beta(1-\beta)}} \quad \text{and} \quad y := \frac{q-w(1-\beta)}{\sqrt{w\beta(1-\beta)}};$$

the two error terms $O((w+1)^{-1/2})$ and $O(w^{-1/2})$ are given by the Berry-Esseen Theorem (see Feller [1965, Theorem XV1,5.1]). By definition of q we have $x = 0$ and $y = O(w^{-1/2})$. The derivative of Φ is bounded, so $\Phi(x) - \Phi(y) = O(x - y) = O(w^{-1/2})$. Altogether, the left-most sum is $O(w^{-1/2})$. A similar computation applies to the right-most sum. ||

• *Proof of Step M8:* Steps M5 and M7 imply $\alpha_{t,w}(j,s^{-i}) = O(w^2/t + w^{-1/2})$. The formula of Step M2 then yields

$$R_t \bullet \mathbf{E}[A_{t+w} | h_t] = O(\frac{w^2}{t} + w^{-1/2}).$$

Adding over w = 1, 2, ..., v, and substituting into Step M1 gives the result (recall that $\sum_{w=1}^{v} w^{\gamma} = O(v^{\gamma+1})$ for $\gamma \neq -1$). ||

• *Proof of Step M9:* We use the inequality of Step M8 for $t = t_n$ and $v = t_{n+1} - t_n$. Because $v = \lfloor (n+1)^{5/3} \rfloor - \lfloor n^{5/3} \rfloor = O(n^{2/3})$, we have $v^3 = O(n^2)$ and $t\, v^{1/2} = O(n^{5/3+1/3}) = O(n^2)$, and the result follows. ||

• *Proof of Step M10:* Taking expectation in the expression of Step M9 yields

$$\mathbf{E}[t_{n+1}^2\ \rho_{t_{n+1}}] = \mathbf{E}[t_n^2\ \rho_{t_n}] + O(n^2).$$

So, by iterating, $\mathbf{E}[t_n^2\ \rho_{t_n}] = O(n^3)$, or $\mathbf{E}[\rho_{t_n}] = O(n^3 / n^{10/3}) = O(n^{-1/3})$.

To obtain the almost sure convergence, divide the formula of Step M9 by t_{n+1}^2, to get (note that $t_n^2 / t_{n+1}^2 \leq 1$)

$$\mathbf{E}[\rho_{t_{n+1}} | h_{t_n}] \leq \rho_{t_n} + K\ n^{-4/3}$$

for an appropriate constant $K < \infty$. Define

$$X_n := \rho_{t_n} + K \sum_{\nu=n}^{\infty} \nu^{-4/3}.$$

Then $(X_n)_{n=1,2,...}$ is a non-negative super-martingale, adapted to the sequence of fields generated by the histories h_{t_n}, i.e., $\mathbf{E}[X_{n+1} | h_{t_n}] \leq X_n$. It therefore converges a.s. as $n \to \infty$ (see Loève [1978, 32.3.A(i)]). The limit X_∞ satisfies by Fatou's lemma $\mathbf{E}[X_\infty] \leq \lim_n \mathbf{E}[X_n] = \lim_n \mathbf{E}[\rho_{t_n}] = 0$ (the series $\sum_\nu \nu^{-4/3}$ converges, so its tail goes to 0). Thus $X_\infty = 0$ a.s., or $X_n \to 0$, so $\rho_{t_n} \to 0$. ||

• *Proof of Step M11:* Since $\rho_{t_n} = \sum_{j \neq k} [R_{t_n}(j,k)]^2$, the previous Step M10 implies that $R_{t_n}(j,k) \to 0$ a.s. as $n \to \infty$, for all $j \neq k$. When $t_n \leq t \leq t_{n+1}$, we have $R_t(j,k) - R_{t_n}(j,k) = O(n^{-1})$ by the inequality of Step M3, so $R_t(j,k) \to 0$ a.s. as $t \to \infty$. ||

5. The Case of the Unknown Game

In this section we present a variation of our basic adaptive procedure of Section 2, with the following two properties: First, in choosing play probabilities at time t each player uses information *only* on his past play and payoffs (and not on possible payoffs if past play had been different than what

it was); in other words, players do not need to know the game. Second, the convergence to the correlated equilibrium set still obtains.

The setup is as in Section 2. However, the information of player i at time t+1 consists just of the t numbers $u^i(s_1)$, $u^i(s_2)$, ..., $u^i(s_t)$. There will be two kinds of modifications on the definition of the adaptive procedure of Section 2. The first and most significant modification concerns the definition of the regret measures. Indeed, for each $j \neq k$ in S^i and each t we use instead of $D_t^i(j,k) = (1/t)\,[\sum_{\tau\le t:s_\tau^i=j} u^i(k,s_\tau^{-i}) - \sum_{\tau\le t:s_\tau^i=j} u^i(s_\tau)]$, the following:

$$C_t^i(j,k) := \frac{1}{t}\left[\sum_{\tau\le t:s_\tau^i=k} \frac{p_\tau^i(j)}{p_\tau^i(k)}\, u^i(s_\tau) - \sum_{\tau\le t:s_\tau^i=j} u^i(s_\tau)\right].$$

The probabilities $p_\tau^i(\cdot)$ will be defined below; they are determined by the information available to player i at time τ, i.e., by $u^i(s_1)$, ..., $u^i(s_{\tau-1})$ (thus C_t^i depends on $u^i(s_1)$, ..., $u^i(s_t)$ only). We define the "modified regrets" by

$$Q_t^i(j,k) := [C_t^i(j,k)]^+ = \max\{C_t^i(j,k),\, 0\}.$$

In words, the modified regret of not having used k instead of j measures now the difference (strictly speaking, its positive part) of the payoff average over the periods when k has been used and the periods when j has been used. In addition, the payoffs of these periods are normalized in a manner

that, intuitively speaking, makes the length of the respective periods comparable.[25] We emphasize again that, in order to compute the modified regrets Q_t^i, player i needs to know only the actual payoffs $u^i(s_\tau)$ that he received in the past. In contrast, for the regrets R_t^i of Section 2, he needs to know in addition what the other players chose (i.e., s_τ^{-i}), as well as what his payoff would have been had he played differently than what he actually did (i.e., $u^i(k,s_\tau^{-i})$) --- which requires knowing the payoff matrix u^i.

The second kind of modifications concerns the definition of the play probabilities. We make two changes. First, we need to guarantee that the sum of the proposed probabilities does not exceed one; multiplying the (modified) regrets by a factor $1/\mu$, which we still do, does not suffice because now regrets are not bounded.[26] Second, we need that every strategy be played with some minimal frequency. Thus we shall require that the transitions may "tremble" over every strategy.

Let $\mu > 0$ be a sufficiently large number; as in Section 2, it suffices to take μ so that $1/\mu < 2\, M^i\, (m^i-1)$ for all i, where $m^i = |S^i|$ is the cardinality of the set S^i, and M^i is an upper bound on $|u^i(s)|$ for all $s \in S$. Let $0 < \delta < 1$, and define the play probabilities of player i at time t+1 by

---

[25] For instance, if k is chosen with a probability that is always twice the probability that j is chosen, then, in the long run, there will be roughly twice as many periods when k was played than when j was played. So, in order to compare the sum of the payoffs over those periods when j was played with the parallel sum when k was played, one needs to multiply the k-sum by 1/2. When the probabilities change over time, the formula used makes the correct normalization (see Step B10(i) and the proof of Step B16 in Section 6).

[26] Notice that there is no a priori lower bound on the probabilities appearing in the denominators.

$$\begin{cases} p^i_{t+1}(k) := (1-\delta)\ \min\{\frac{1}{\mu}\, Q^i_t(j,k),\ \frac{1}{m^i-1}\} + \delta\ \frac{1}{m^i}, & \text{for } k \neq j; \\ p^i_{t+1}(j) := 1 - \sum_{k \in S^i : k \neq j} p^i_{t+1}(k), \end{cases} \tag{5.1}$$

where $j := s^i_t$ is the choice in (the previous) period t. Note that $p^i_{t+1}(j)$ may be rewritten as $(1-\delta)\ [1 - \Sigma_{k \neq j}\ \min\{Q^i_t(j,k)/\mu,\ 1/(m^i-1)\}] + \delta/m^i$. As for the first period, the play probabilities there are given by an arbitrary strictly positive vector $p^i_1 \in \Delta(S^i)$; for simplicity, assume that $p^i_1(j) \geq \delta/m^i$ for all j in S^i.

Formula (5.1) says that p^i_{t+1} is a weighted average of two probability vectors. The first, with weight $1-\delta$, is given by the modified regrets in a manner similar to (2.2) (the minimum with $1/(m^i-1)$ guarantees that the resulting sum does not exceed 1). The second term, with weight δ, is just the uniform distribution over S^i: each $k \in S^i$ has equal probability $1/m^i$. This "uniform tremble" guarantees that all probabilities at time t+1 are at least $\delta/m^i > 0$.

The empirical distributions of play up to t are denoted, as in Section 2, by $z_t \in \Delta(S)$ (see (2.3)). Our result is

> **Theorem B:** For every $\varepsilon > 0$ there is a $\delta_0 \equiv \delta_0(\varepsilon) > 0$ such that for all $\delta < \delta_0$, if every player plays according to the adaptive procedure described by (5.1), then the

empirical distributions of play z_t converge in probability to the set of correlated ε-equilibria.

Thus, the probability that z_t is a correlated 2ε-equilibrium converges to one as t goes to infinity. As in the Main Theorem, both μ and δ may differ for different players.

To obtain in the limit correlated equilibria — rather than correlated *approximate* equilibria — we need δ to decrease to 0 as t increases. A simple way to do this is to replace δ in (5.1) by $\delta\, t^{-\gamma}$, where γ is a number strictly between 0 and[27] 1/3. Thus, we now define

$$\begin{cases} p^i_{t+1}(k) := (1 - \delta\, t^{-\gamma}) \min\{\frac{1}{\mu}\, Q^i_t(j,k),\ \frac{1}{m^i - 1}\} + \delta\, t^{-\gamma}\, \frac{1}{m^i}, \quad \text{for } k \neq j; \\ p^i_{t+1}(j) := 1 - \sum_{k \in S^i : k \neq j} p^i_{t+1}(k), \end{cases} \tag{5.2}$$

where, again, $j := s^i_t$ is the choice in period t. At t = 1, let again p^i_1 be an arbitrary vector in $\Delta(S^i)$ with $p^i_1(j) \geq \delta/m^i$ for all j in S^i. We have

Theorem B': If every player plays according to the adaptive procedure described by (5.2), then the empirical distributions of play z_t converge almost surely as $t \to \infty$ to the set of correlated equilibria of the game Γ.

---

[27] The reason for the upper bound of 1/3 will be discussed below.

We have chosen the particular type of sequence $\delta_t := \delta\, t^{-\gamma}$ for simplicity and convenience only. What matters is that, on the one hand, δ_t converge to 0. On the other hand, it should do so "not too fast" (since then the modified regrets C_t^i may become too large; recall that the probabilities p_t^i, which are bounded from below by δ_{t-1}/m^i, appear in the denominators). This explains the need for an upper bound on γ (it turns out that $\gamma < 1/3$ suffices; the reader should consult the proof in Section 6). Moreover, we note that one may well take different sequences δ_t^i for different players i (see the Remark at the end of Section 6).

The results of this section may be summarized as follows: Consider games where each player knows only how many strategies he has (but nothing about the other players and about his own payoff function). Then there exist simple adaptive procedures, based for each player solely on his own past realized payoffs, that lead to correlated equilibria; namely, the frequencies that the N-tuples of strategies are played over time form a correlated (approximate) equilibrium.[28]

6. Proof of the Result for the Unknown Game Case

We will prove the two Theorems B and B′ together. Let thus

---

[28] Of course, this can be perceived only by an outside observer.

$$\delta_t := \frac{\delta}{t^\gamma},$$

where $\delta > 0$ and $0 \le \gamma < 1/3$. For Theorem B take $\gamma = 0$, and for Theorem B′, $\gamma > 0$. As in the proof of the Main Theorem in Section 4, we fix player i in N, and drop reference to the index i whenever this cannot cause confusion. In addition to the notations introduced there, we define for each t = 1, 2, ... and each (j,k) in L

$$Z_t(j,k) := \frac{p_t^i(j)}{p_t^i(k)} \mathbf{1}_{\{s_t^i=k\}} - \mathbf{1}_{\{s_t^i=j\}};$$

$$B_t(j,k) := Z_t(j,k)\, u^i(s_t)$$

(thus $C_t(j,k) = (1/t) \sum_{\tau \le t} B_\tau(j,k)$); and

$$\Pi_t(j,k) := \begin{cases} (1-\delta_t) \min\{\frac{1}{\mu} C_t^+(j,k), \frac{1}{m-1}\} + \delta_t \frac{1}{m}, & \text{if } k \ne j, \\ (1-\delta_t)\Big[1 - \sum_{k' \ne j} \min\{\frac{1}{\mu} C_t^+(j,k'), \frac{1}{m-1}\}\Big] + \delta_t \frac{1}{m}, & \text{if } k = j. \end{cases}$$

Note that $\Pi_t(j,\cdot) \in \Delta(S^i)$ for all $j \in S^i$, thus Π_t is a probability transition matrix on S^i. Both procedures (5.1) and (5.2) satisfy $p_{t+1}^i(k) = \Pi_t(s_t^i,k)$ for all k (where, again, $\gamma = 0$ corresponds to (5.1) and $\gamma > 0$ to (5.2)). Let

$$\rho_t := [\mathrm{dist}(C_t, \mathbb{R}_-^L)]^2$$

be the squared distance (in $\mathbb{R}^L$) of the vector C_t from the non-positive orthant $\mathbb{R}^L_-$. Thus $\rho_t = \Sigma_{j\neq k}\,[C_t^+(j,k)]^2$.

The proof of Theorems B and B′ will be divided into 16 steps, B1 - B16. The final Step B16, together with the Proposition of Section 3, completes the proof. We note that Steps B1 - B10 hold for *any* non-increasing sequence $\delta_t > 0$, whereas Steps B11 - B16 make use of the rate of the special form $\delta_t = \delta\, t^{-\gamma}$. A guide to the proof follows the statement of the steps.

- *Step B1:* $\mathbf{E}[(t+v)^2\,\rho_{t+v}|h_t] = t^2\,\rho_t + 2t\sum_{w=1}^{v} C_t^+ \bullet \mathbf{E}[B_{t+w}|h_t] + O(\frac{v^2}{\delta^2_{t+v}})$.

Define[29]

$$\beta_{t+w}(j) := \sum_{k\neq j} \frac{1}{\mu}\, C_t^+(k,j)\; p^i_{t+w}(k) - \sum_{k\neq j} \frac{1}{\mu}\, C_t^+(j,k)\; p^i_{t+w}(j).$$

- *Step B2:* $C_t^+ \bullet \mathbf{E}[B_{t+w}|h_t] = \mu\,\mathbf{E}\Big[\sum_{j\in S^i} u^i(j,s^{-i}_{t+w})\;\beta_{t+w}(j)\,\Big|\,h_t\Big]$,

- *Step B3:* $\Pi_{t+v}(j,k) - \Pi_t(j,k) = O(\frac{v}{t\,\delta_{t+v}} + [\delta_t - \delta_{t+v}])$.

---

[29] Note that $\beta_{t+w}(j)$ is measurable with respect to h_{t+w} (actually, it depends only on h_{t+w-1} and s^{-i}_{t+w}, but not on s^i_{t+w}).

Define:

$$\vartheta := \frac{1}{m-1} - \frac{2M}{\mu} > 0, \quad \text{and} \quad Y_t := \sum_{(j,k)\in L} [\frac{M}{\mu} \bar{Z}_t(j,k) - \vartheta]^+.$$

- *Step B4:* $\Pi_t(j,k) - \frac{1}{\mu} C_t^+(j,k) = O(Y_t + \delta_t)$.

- *Step B5:* $C_t^+ \bullet \mathbf{E}[B_{t+w} | h_t] = \mu \, E\Big[\sum_{j\in S^i} u^i(j,s^{-i}_{t+w}) \, [\Pi_t^2 - \Pi_t](s^i_{t+w-1},j) \Big| h_t \Big]$
$$+ O(Y_t + \delta_t + \frac{w}{t\,\delta_{t+w}}).$$

Recall the auxiliary process $(\hat{s}_{t+w})_{w=0,1,2,...}$ as defined in Step M4 in Section 4.

- *Step B6:* $\mathbf{P}[s_{t+w} = s | h_t] - \mathbf{P}[\hat{s}_{t+w} = s | h_t] = O(\frac{w^2}{t\,\delta_{t+w}} + w\,[\delta_t - \delta_{t+w}])$.

- *Step B7:* $C_t^+ \bullet \mathbf{E}[B_{t+w} | h_t] = \mu \, \mathbf{E}\Big[\sum_{j\in S^i} u^i(j,\hat{s}^{-i}_{t+w}) \, [\Pi_t^2 - \Pi_t](\hat{s}^i_{t+w-1},j) \Big| h_t \Big]$
$$+ O(Y_t + \delta_t + \frac{w^2}{t\,\delta_{t+w}} + w\,[\delta_t - \delta_{t+w}]).$$

- *Step B8:*[30] $\mathbf{E}\Big[\sum_{j \in S^i} u^i(j, \hat{s}^{-i}_{t+w}) \; [\Pi_t^{\,2} - \Pi_t](\hat{s}^i_{t+w-1}, j) \,\Big|\, h_t \Big]$

$$= \sum_{s^{-i}} \mathbf{P}[\hat{s}^{-i}_{t+w} = s^{-i} \,|\, h_t] \; [\Pi_t^{\,w+1} - \Pi_t^{\,w}](s^i, j) = O((1-\delta_t)^w).$$

- *Step B9:* $\mathbf{E}[(t+v)^2 \, \rho_{t+v} \,|\, h_t] = t^2 \, \rho_t$

$$+ O(t \; v \; Y_t + t \; v \; \delta_t + \frac{v^3}{\delta_{t+v}} + t \; v^2 \; [\delta_t - \delta_{t+v}] + \frac{t}{\delta_t} + \frac{v^2}{\delta^2_{t+v}}).$$

- *Step B10:* (i) $\mathbf{E}[Z_t(j,k) \,|\, h_{t-1}] = 0$;

(ii) $\mathrm{Var}[Z_t(j,k)] = O(\delta_t^{-1})$; and

(iii) $\mathbf{E}[Y_t] = O(t^{-1/2} \; \delta_t^{-1/2})$.

- *Step B11:* $\bar{Z}_t(j,k) \dashrightarrow 0$ a.s. as $t \dashrightarrow \infty$.

For each $n = 1, 2, \ldots$, let $t_n ::= \lfloor n^{3/2} \rfloor$ be the largest integer not exceeding $n^{3/2}$.

- *Step B12:* $\mathbf{E}[t^2_{n+1} \, \rho_{t_{n+1}} \,|\, h_{t_n}] = t^2_n \, \rho_{t_n} + O(\delta \; n^{2-3\gamma/2} + Y_{t_n} \; n^2 + n^{3/2+3\gamma/2})$.

- *Step B13:* $\mathbf{E}[\rho_{t_n}] = O(\delta \; n^{-3\gamma/2} + n^{3\gamma/2-1/2})$.

- *Step B14:* If $\delta_t = \delta$ for all t, then $\mathbf{E}[\rho_t] = O(\delta)$.

---

[30] One could use here another estimate, $O(w^{-1/2})$, as in Steps M6 and M7 in the proof of the Main Theorem; it would have made no difference.

• *Step B15:* If $\delta_t \to 0$ (i.e., $\gamma > 0$), then $\mathbf{E}[\rho_{t_n}] = O(n^{-\lambda})$, where $\lambda := \min\{3\gamma/2, (1-3\gamma)/2\} > 0$, and $\rho_{t_n} \to 0$ a.s. as $n \to \infty$.

• *Step B16:* (i) If $\delta_t = \delta$ for all t, then $\limsup_{t\to\infty} \mathbf{E}[D_t^+(j,k)] = O(\sqrt{\delta})$.

(ii) If $\delta_t \to 0$ (i.e., if $\gamma > 0$), then $\lim_{t\to\infty} D_t^+(j,k) = 0$.

We now provide a short intuitive overview of the steps of the proof. The proof is based on the proof of the Main Theorem in Section 4, with a number of additional steps needed to take care of the modifications. Most of the proof is devoted to show that the modified regrets $Q_t^i \equiv [C_t^i]^+$ are small (see Steps B14 and B15). From this one readily gets in Step B16 that the actual regrets $R_t^i \equiv [D_t^i]^+$ are also small, since the difference $C_t^i - D_t^i$ is a martingale converging almost surely to 0 (see Steps B10 and B11). The main steps in the proof are as follows: We start with the basic recursion equation in Step B1 (similar to M1). Next, we estimate the "middle term" on the right-hand of B1 by approximating the s-process with the $\hat{s}$-process, which is independent across players (Steps B2 - B7; parallel to M2 - M6). This leads to a formula comparable to (3.5), except that the invariant distribution in (3.5b) is replaced here by the transitions after w and w+1 periods, which are close to one another (Step B8; compare with M7). Finally, we obtain the recursion formula in Step B9. When comparing this formula with the parallel one in M8, we see that in B9 there are additional terms; one of them (that comes about because the modified regrets are not bounded) contains a random variable Y_t. Step B10(iii), Step B11 and the proof of Step B15 show that this term also

goes to zero. The next steps, B12 - B15, complete the proof for the modified regrets, in a manner that is similar (though more complicated, partly because of the Y_t terms) to M9 - M11. As we have indicated above, Step B16 yields the result for the actual regrets.

We now proceed to the proofs of B1 - B16.

• *Proof of Step B1:* Similar to that of Step M1, with B and C, respectively, replacing A and D there. Note that $Z_{t+w}(j,k)$ is bounded by m/δ_{t+w} which, for $w \leq v$, is at most m/δ_{t+v} (the sequence δ_t is non-increasing), therefore $\|B_{t+w}\|$ and $\|C_t\|$ are $O(\delta_{t+v}^{-2})$ — which yields the bound on the third term. ||

• *Proof of Step B2:* Note that $B_{t+w}(j,k)$ vanishes except when $s^i_{t+w} = j$ or k. We condition on h_{t+w-1} and s^{-i}_{t+w} (i.e., h_{t+w} *except* player i–th choice at time t+w):

$$\mathbf{E}[B_{t+w}(j,k) \,|\, h_{t+w-1},\ s^{-i}_{t+w} = s^{-i}]$$

$$= p^i_{t+w}(k)\ \frac{p^i_{t+w}(j)}{p^i_{t+w}(k)}\ u^i(k,s^{-i}) - p^i_{t+w}(j)\ u^i(j,s^{-i})$$

$$= p^i_{t+w}(j)\ [u^i(k,s^{-i}) - u^i(j,s^{-i})].$$

Hence

$$C_t^+ \bullet \mathbf{E}[B_{t+w} | h_{t+w-1},\ s_{t+w}^{-i} = s^{-i}] = \sum_j \sum_{k \neq j} C_t^+(j,k)\ p_{t+w}^i(j)\ [u^i(k,s^{-i}) - u^i(j,s^{-i})]$$

$$= \sum_j u^i(j,s^{-i}) \left[\sum_{k \neq j} C_t^+(k,j)\ p_{t+w}^i(k) - \sum_{k \neq j} C_t^+(j,k)\ p_{t+w}^i(j) \right]$$

(we have collected together all terms containing $u^i(j,s^{-i})$). Conditioning now on h_t yields the result. ||

- *Proof of Step B3:* We have

$$(t+v)\ |C_{t+v}(j,k) - C_t(j,k)| \leq \sum_{w=1}^{v} |B_{t+w}(j,k) - C_t(j,k)|.$$

Together with $B_{t+w}(j,k)$, $C_t(j,k) = O(\delta_{t+v}^{-1})$, we get $C_{t+v}(j,k) - C_t(j,k) = O(vt^{-1}\delta_{t+v}^{-1})$. For $j \neq k$, the difference between $\Pi_{t+v}(j,k)$ and $\Pi_t(j,k)$ is therefore at most $(1-\delta_t)\ O(vt^{-1}\delta_{t+v}^{-1}) + (\delta_t - \delta_{t+v})\ (m-1)^{-1}$. For $j = k$, it is at most $m-1$ times this amount. ||

- *Proof of Step B4:* We need to consider only $j \neq k$ (the error for $j = k$ is at most $m-1$ times larger). If $(1/\mu)\ C_t^+(j,k) \leq 1/(m-1)$, then

$$\Pi_t(j,k) - \frac{1}{\mu} C_t^+(j,k) = \delta_t \left[\frac{1}{m} - \frac{1}{\mu} C_t^+(j,k)\right],$$

and this lies between 0 and δ_t/m. If $(1/\mu)\ C_t^+(j,k) \geq 1/(m-1)$, then $\Pi_t(j,k) = (1-\delta_t)/(m-1) + \delta_t/m \leq 1/(m-1)$, so

$$\frac{1}{\mu} C_t^+(j,k) \leq \frac{1}{\mu} |C_t(j,k)| \leq \frac{1}{\mu} \frac{1}{t} \sum_{\tau=1}^{t} |Z_\tau(j,k)|\ |u^i(s_\tau)|$$

$$\leq \frac{2M}{\mu} + \frac{M}{\mu} \bar{Z}_t(j,k) = \frac{1}{m-1} - \vartheta + \frac{M}{\mu} \bar{Z}_t(j,k)$$

$$\leq \Pi_t(j,k) + O(\delta_t) + \left[\frac{M}{\mu} \bar{Z}_t(j,k) - \vartheta\right]^+,$$

where we have used the inequality $|Z_\tau(j,k)| \leq 2 + Z_\tau(j,k)$ (since the only possible negative value of $Z_\tau(j,k)$ is -1). Together with

$$\Pi_t(j,k) \leq \frac{1}{m-1} \leq \frac{1}{\mu} C_t^+(j,k),$$

the proof is complete. ||

- *Proof of Step B5:* Denote s^i_{t+w-1} by r, then

$$\beta_{t+w}(j) = \sum_{k \neq j} \frac{1}{\mu} C_t^+(k,j)\ \Pi_{t+w-1}(r,k) - \sum_{k \neq j} \frac{1}{\mu} C_t^+(j,k)\ \Pi_{t+w-1}(r,j).$$

Also,

$$[\Pi_t^2 - \Pi_t](r,j) = [\Pi_t^2](r,j) - \Pi_t(r,j)$$

$$= \sum_{k \in S^i} \Pi_t(k,j)\ \Pi_t(r,k) - \Pi_t(r,j) \sum_{k \in S^i} \Pi_t(j,k)$$

$$= \sum_{k \neq j} \Pi_t(k,j)\ \Pi_t(r,k) - \sum_{k \neq j} \Pi_t(j,k)\ \Pi_t(r,j)$$

(we have subtracted the j'th term from both sums). Comparing the last expression with $\beta_{t+w}(j)$, we see that $[\Pi_t^2 - \Pi_t](r,j)$ is obtained by replacing each C_t^+/μ and each Π_{t+w-1} in $\beta_{t+w}(j)$ by Π_t. Thus

$$\begin{aligned}\beta_{t+w}(j) - [\Pi_t^2 - \Pi_t](r,j) = \sum_{k \neq j} &[\frac{1}{\mu}\ C_t^+(k,j) - \Pi_t(k,j)]\ \Pi_{t+w-1}(r,k) \\ &+ \sum_{k \neq j} \Pi_t(k,j)\ [\Pi_{t+w-1}(r,k) - \Pi_t(r,k)] \\ &- \sum_{k \neq j} [\frac{1}{\mu}\ C_t^+(j,k) - \Pi_t(j,k)]\ \Pi_{t+w-1}(r,j) \\ &- \sum_{k \neq j} \Pi_t(j,k)\ [\Pi_{t+w-1}(r,j) - \Pi_t(r,j)]\end{aligned}$$

$$= O(Y_t + \delta_t + \frac{w}{t}\delta_{t+w}),$$

by the estimates of Steps B3 and B4. It only remains to substitute this into the formula of Step B2. ||

• *Proof of Step B6:* Identical to the proof of Step M4, using the estimate of Step B3 instead of that of Step M3. ||

• *Proof of Step B7:* Replacing $(s_{t+w})_w$ by $(\hat{s}_{t+w})_w$ in the formula of Step B5 gives an additional error that is estimated in Step B6 (note that the two processes $(s_{t+w})_w$ and $(\hat{s}_{t+w})_w$ start from the same history h_t). ||

• *Proof of Step B8:* As in Step M6, $\hat{s}^{-i}_{t+w}$ and $\hat{s}^{i}_{t+w-1}$ are independent given h_t. Hence:

$$\mathbf{E}\Big[\sum_{j\in S^i} u^i(j,\hat{s}^{-i}_{t+w})\ [\Pi_t^{\,2} - \Pi_t](\hat{s}^{i}_{t+w-1},j) \,\Big|\, h_t \Big]$$

$$= \sum_{s^{-i}\in S^{-i}} \mathbf{P}[\hat{s}^{-i}_{t+w} = s^{-i} \,|\, h_t] \sum_{r\in S^i} \mathbf{P}[\hat{s}^{i}_{t+w-1} = r \,|\, h_t]\ [\Pi_t^{\,2} - \Pi_t](r,j)$$

$$= \sum_{s^{-i}} \mathbf{P}[\hat{s}^{-i}_{t+w} = s^{-i} \,|\, h_t] \sum_{r} \Pi_t^{\,w-1}(s^i_t,r)\ [\Pi_t^{\,2} - \Pi_t](r,j)$$

$$= \sum_{s^{-i}} \mathbf{P}[\hat{s}^{-i}_{t+w} = s^{-i} \,|\, h_t] \; [\Pi_t^{w+1} - \Pi_t^{w}](s^i, j).$$

Now Π_t is an $m \times m$ stochastic matrix with all entries at least $\delta_t/m > 0$, implying, as it is well-known, that $|[\Pi_t^{w+1}](k,j) - [\Pi_t^{w}](k,j)| \leq (1 - m\,\delta_t/m)^w$. ||

• *Proof of Step B9:* Putting together the estimates of Steps B8 and B9 (note that $\sum_{w=1}^{v} (1-\delta_t)^w \leq 1/\delta_t$) yields

$$2t \sum_{w=1}^{v} C_t^+ \bullet \mathbf{E}[B_{t+w} \,|\, h_t] = O(t\; v\; Y_t + t\; v\; \delta_t + \frac{t\; v^3}{\delta_{t+v}} + t\; v^2\; [\delta_t - \delta_{t+v}] + \frac{t}{\delta_t}).$$

Recalling the formula of Step B1 completes the proof. ||

• *Proof of Step B10:* By definition $Z_t(j,k)$,

$$\mathbf{E}[Z_t(j,k) \,|\, h_{t-1}] = \frac{p_t^i(j)}{p_t^i(k)}\; p_t^i(k) - p_t^i(j) = 0.$$

Therefore $(Z_t(j,k))_t$ are martingale differences, and $\mathbf{E}[Z_t(j,k)] = 0$. Also

$$\mathrm{Var}[Z_t(j,k)] = \mathbf{E}[Z_t^2(j,k)] = \mathbf{E}\Big[\mathbf{E}[Z_t^2(j,k)\,|\,h_{t-1}]\Big]$$

$$= \mathbf{E}\left[\frac{(p_t^i(j))^2}{(p_t^i(k))^2}\, p_t^i(k) + (-1)^2\, p_t^i(j)\right] \le \mathbf{E}\left[\frac{1}{p_t^i(k)}\right] \le \frac{m}{\delta_t}.$$

From this we get

$$\mathrm{Var}[\bar{Z}_t(j,k)] = \frac{1}{t^2} \sum_{\tau=1}^{t} \mathrm{Var}[Z_\tau(j,k)] \le \frac{m}{t\,\delta_t}$$

(the variance of the sum of martingale differences is the sum of their variances). Finally, since $\vartheta > 0$ we have $[\zeta - \vartheta]^+ \le |\zeta|$, so

$$\mathbf{E}[\tfrac{M}{\mu}\,\bar{Z}_t(j,k) - \vartheta]^+ \le \mathbf{E}|\tfrac{M}{\mu}\,\bar{Z}_t(j,k)| \le \left[\mathrm{Var}[\tfrac{M}{\mu}\,\bar{Z}_t(j,k)]\right]^{1/2} \le \left[\frac{m}{t\,\delta_t}\right]^{1/2}.$$

Summing over all $j \neq k$ gives the estimate for $\mathbf{E}[Y_t]$. ||

- *Proof of Step B11:* By Step B10,

$$\sum_{t=1}^{\infty} \frac{1}{t^2}\,\mathrm{Var}[Z_t(j,k)] \le \sum_{t=1}^{\infty} \frac{m}{t^2\,\delta_t}\,.$$

This series converges, since $\gamma < 1/3$. Therefore $\bar{Z}_t(j,k) \to 0$ a.s. (see Loève [1978, 32.1.E]). ||

• *Proof of Step B12:* Apply the inequality of Step B10 with $t = t_n = n^{3/2}$ and $v = t_{n+1} - t_n$. Then: $v = O(n^{1/2})$; $\delta_t = \delta\ (n^{3/2})^{-\gamma} = \delta\ n^{-3\gamma/2}$; $\delta_{t+v} = \delta\ (n+1)^{-3\gamma/2} = O(n^{-3\gamma/2})$; $\delta_t - \delta_{t+v} = O(n^{-3\gamma/2-1})$, so

$$\mathbf{E}[t_{n+1}^2\ \rho_{t_{n+1}} | h_{t_n}] = t_n^2\ \rho_{t_n}$$

$$+ O(Y_{t_n}\ n^2 + \delta\ n^{2-3\gamma/2} + n^{3/2+3\gamma/2} + n^{3/2-3\gamma/2} + n^{3/2+3\gamma/2} + n^{1+3\gamma}).$$

Now $\gamma < 1/3$ implies $3/2 - 3\gamma/2$, $1 + 3\gamma < 3/2 + 3\gamma/2 < 2$. ||

• *Proof of Step B13:* By Step B10, $\mathbf{E}[Y_{t_n}] = O(t_n^{-1/2}\ \delta_{t_n}^{-1/2}) = O(n^{-3/4+3\gamma/4}) = o(n^{-1/2})$ (since $\gamma < 1/3$); thus $n^2\ \mathbf{E}[Y_{t_n}] = o(n^{3/2})$. Taking expectation in the formula of Step B12, we thus get

$$\mathbf{E}[t_{n+1}^2\ \rho_{t_{n+1}}] = \mathbf{E}[t_n^2\ \rho_{t_n}] + O(\delta\ n^{2-3\gamma/2} + n^{3/2+3\gamma/2}).$$

So, by iterating, $\mathbf{E}[t_n^2\ \rho_{t_n}] = O(\delta\ n^{3-3\gamma/2} + n^{5/2+3\gamma/2})$, or $\mathbf{E}[\rho_{t_n}] = O(\delta\ n^{-3\gamma/2} + n^{-1/2+3\gamma/2})$. ||

• *Proof of Step B14:* By Step B13, $\mathbf{E}[\rho_{t_n}] = O(\delta)$. When $t_n \le t \le t_{n+1}$, we have $\mathbf{E}[C_t^+(j,k)] - \mathbf{E}[C_{t_n}^+(j,k)] = O(\delta)$ for all $j \ne k$ (combine the estimates of Steps B3, B4 and B10), so $\mathbf{E}[\rho_t] - \mathbf{E}[\rho_{t_n}] = O(\delta^2)$. ||

• *Proof of Step B15:* By Step B13, $\mathbf{E}[\rho_{t_n}] = O(n^{-\lambda}) \to 0$ as $n \to \infty$. Step B12 gives

$$\mathbf{E}[t_{n+1}^2\ \rho_{t_{n+1}} | h_{t_n}] = t_n^2\ \rho_{t_n} + O(n^{2-\lambda} + Y_{t_n}\ n^2).$$

Divide by t_{n+1}^2, to get (note that $t_n^2 / t_{n+1}^2 \leq 1$)

$$\mathbf{E}[\rho_{t_{n+1}} | h_{t_n}] \leq \rho_{t_n} + \eta_n,$$

where $\eta_n := K\ (n^{-1-\lambda} + n^{-1}\ Y_{t_n})$ for an appropriate constant $K < \infty$.

By Step B10, the following statements hold with probability 1: $\bar{Z}_t(j,k)$ converges to 0 as t goes to ∞; therefore, $[M/\mu\ \bar{Z}_t(j,k) - \vartheta]^+ = 0$ for all except finitely many t's (recall that $\vartheta > 0$); so the series $\sum_t \eta_t$ converges. Moreover, the estimate for $\mathbf{E}[Y_t]$ in Step B10 implies that $\sum_t \mathbf{E}[\eta_t]$ also converges.

Let

$$X_n := \rho_{t_n} - \sum_{\nu=1}^{n-1} \eta_\nu.$$

Then $\mathbf{E}[X_{n+1} | h_{t_n}] \leq X_n$, so $(X_n)_{n=1,2,...}$ is a super-martingale (adapted to the sequence of fields generated by the histories $(h_{t_n})_n$). Moreover, $\sup_n \mathbf{E}[X_n^-] \leq$

$\sup_n \mathbf{E}[\sum_\nu^{n-1} \eta_\nu] \leq \sum_{\nu=1}^{\infty} \mathbf{E}[\eta_\nu] < \infty$, so $(X_n)_n$ converges a.s. as as $n \to \infty$ (see Loève [1978, 32.3.A(i)]). Call the limit X_∞, and define

$$\rho_\infty := X_\infty + \sum_{\nu=1}^{\infty} \eta_\nu.$$

Then

$$\rho_{t_n} + \sum_{\nu=n}^{\infty} \eta_\nu = X_n + \sum_{\nu=1}^{\infty} \eta_\nu \to X_\infty + \sum_{\nu=1}^{\infty} \eta_\nu = \rho_\infty.$$

But $\sum_{\nu=n}^{\infty} \eta_\nu \to 0$ as $n \to \infty$, so $\rho_{t_n} \to \rho_\infty$. Fatou's lemma implies $\mathbf{E}[\rho_\infty] \leq \lim_n \mathbf{E}[\rho_{t_n}] = 0$, so $\rho_\infty = 0$ a.s. ||

- *Proof of Step B16:* We have $\mathbf{1}_{\{s_t^i=k\}}\, u^i(s_t) = \mathbf{1}_{\{s_t^i=k\}}\, u^i(k,s_t^{-i})$, so

$$\begin{aligned} &B_t(j,k) - A_t(j,k) \\ &\quad = \left[\frac{p_t^i(j)}{p_t^i(k)} \mathbf{1}_{\{s_t^i=k\}} - \mathbf{1}_{\{s_t^i=j\}} \right] u^i(s_t) - \mathbf{1}_{\{s_t^i=j\}} \left[u^i(k,s_t^{-i}) - u^i(s_t)\right] \\ &\quad = \left[\frac{p_t^i(j)}{p_t^i(k)} \mathbf{1}_{\{s_t^i=k\}} - \mathbf{1}_{\{s_t^i=j\}} \right] u^i(k,s_t^{-i}) = Z_t(j,k)\, u^i(k,s_t^{-i}). \end{aligned}$$

But s_t^i and s_t^{-i} are independent given h_{t-1}, therefore $\mathbf{E}[B_t(j,k) - A_t(j,k) \,|\, h_{t-1}] = \mathbf{E}[Z_t(j,k) \,|\, h_{t-1}]\ \mathbf{E}[u^i(k,s_t^{-i}) \,|\, h_{t-1}] = 0$, since the first term is 0 by Step B10. So $(B_t(j,k) - A_t(j,k))_t$ are martingale differences, and

$$\mathrm{Var}[B_t(j,k) - A_t(j,k)] = \mathbf{E}[Z_t^2(j,k)\ (u^i(k,s_t^{-i}))^2] \le M^2\ \mathbf{E}[Z_t^2(j,k)] = O(\delta_t^{-1}).$$

Thus the series $\sum_t \mathrm{Var}[B_t(j,k) - A_t(j,k)] / t^2$ converges, implying that $C_t(j,k) - D_t(j,k) = (1/t) \sum_{\tau \le t} [B_\tau(j,k) - A_\tau(j,k)] \to 0$ a.s. as $t \to \infty$ (again, see Loève [1978, 32.1.E]).

In the case $\gamma > 0$ we have $C_t^+(j,k) \le \sqrt{\rho_t} \to 0$ by Step B15, so $D_t^+(j,k) \to 0$ a.s. In the case $\delta_t = \delta$ for all t we get $\mathbf{E}^2[C_t^+(j,k)] \le \mathbf{E}[[C_t^+(j,k)]^2] \le \mathbf{E}[\rho_t] = O(\delta)$ by Step B14, completing the proof since $\mathbf{E}[D_t(j,k)] = \mathbf{E}[C_t(j,k)]$. ||

Remark: For simplicity, we have assumed that all players use the same sequence $(\delta_t)_t$. In case there are different sequences $(\delta_t(i'))_t$ for the different players i′, where $\delta_t(i') = \delta(i')\ t^{-\gamma(i')}$ for some $\delta(i') > 0$ and $0 \le \gamma(i') < 1/3$, then it is straightforward to check that the estimate of B6 becomes $\sum_{i' \in N} O(w^2\ t^{-1}\ (\delta_{t+w}(i'))^{-1} + w\ [\delta_t(i') - \delta_{t+w}(i')])$. This implies that the right-hand side in Step B12 is $t_n^2\ \rho_{t_n} + O(\delta(i)\ n^{2-3\gamma(i)/2} + Y_{t_n}\ n^2 + n^{3/2+3\bar{\gamma}/2})$, where $\bar{\gamma} := \max_{i' \in N} \gamma(i')$, and the final result in B16 becomes:

(i) If $\delta_t(i) = \delta(i)$ for all t, then $\mathrm{limsup}_t\ \mathbf{E}[[D_t^i]^+(j,k)] = O(\sqrt{\delta(i)})$.

(ii) If $\delta_t(i) \to 0$ (i.e., if $\gamma(i) > 0$), then $\lim_t\ [D_t^i]^+(j,k) = 0$.

Appendix: A Simple Adaptive Procedure Leading to Asymptotic Optimal Payoff

A line of research started by Hannan [1957] considers the following question: Can a player do asymptotically at least as well as if he knew beforehand the proportion of time that each (N–1)-tuple of strategies will be used by the other players? More precisely, let z_t^{-i} denote the empirical distribution of the choices made by all players except i; i.e., for each s^{-i} in S^{-i},

$$z_t^{-i}(s^{-i}) := \frac{1}{t} \left| \{\tau \leq t : s_\tau^{-i} = s^{-i}\} \right|.$$

Hannan [1957] proved that player i can guarantee in the long-run that his average payoff $(1/t) \sum_{\tau \leq t} u^i(s_\tau)$ will be at least as large as his best reply payoff against z_t^{-i}, i.e.[31] $\max_{j \in S^i} u^i(j, z_t^{-i})$. The same result was shown by Blackwell [1956b] to follow from his Approachability Theorem (see also Luce and Raiffa [1957, pp. 482-483]).

Our adaptive procedures also lead to this "asymptotic optimal payoff". Indeed, for each j in S^i, they actually guarantee it when restricted to those periods when player i chose that particular j; this being true for *all* j in S^i, the result follows.[32] However, the application of Blackwell's

---

[31] In other words, the average payoff is no worse than if i were to play any *constant* strategy $j \in S^i$ for all t; of course, the best payoff against z_t^{-i} using *non-stationary* strategies is usually much higher. (We write $u^i(j, z_t^{-i})$ for $(1/t) \sum_{\tau \leq t} u^i(j, s_\tau^{-i})$).

[32] We note that the two conditions -- asymptotic optimal payoff, and regrets converging to zero -- are equivalent for player i when he has just two

Approachability Theorem in Section 3 suggests the following much simpler procedure.[33]

At time t, for each strategy j in S^i, let

$$D_t^i(j) := \frac{1}{t} \sum_{\tau=1}^{t} \left[u^i(j, s_\tau^{-i}) - u^i(s_\tau) \right], \tag{A.1a}$$

$$p_{t+1}^i(j) := \frac{[D_t^i(j)]^+}{\sum_{k \in S^i} [D_t^i(k)]^+}, \tag{A.1b}$$

if the denominator is positive, and $(p_{t+1}^i(j))_{j \in S^i}$ in $\Delta(S^i)$ arbitrary otherwise. The strategy of player i is then, at time t+1, to choose j in S^i with probability $p_{t+1}^i(j)$. Thus, the probabilities are proportional to the "regrets" $[D_t^i(j)]^+$. At time 1, $p_1^i \in \Delta(S^i)$ is arbitrary.

Theorem C. If player i plays according to the adaptive procedure given by (A.1), then

$$\liminf_{t \to \infty} \left[\frac{1}{t} \sum_{\tau=1}^{t} u^i(s_\tau) - \max_{s^i \in S^i} u^i(s^i, z_t^{-i}) \right] \geq 0 \quad \text{a.s.} \tag{A.2}$$

---

strategies, i.e. when $|S^i| = 2$.

[33] The proof in Blackwell [1956b] does not seem to lead to a simple procedure (see Luce and Raiffa [1957]).

Note that, as in Theorem A of Section 3, no assumption is made on the behavior of the other players; (A.2) is guaranteed no matter how they play.

Proof: The proof is similar to the proof of Theorem A in Section 3 — a direct application of Blackwell's Approachability Theorem. Let $m := |S^i|$, and define the vector payoff $v(s^i,s^{-i}) \in \mathbb{R}^m$ by letting its coordinate j (for $j \in S^i$) be

$$v(s^i,s^{-i})(j) := u^i(j,s^{-i}) - u^i(s^i,s^{-i}).$$

We claim that the non-positive orthant $\mathbb{R}^m_-$ of $\mathbb{R}^m$ is approachable by player i, using the strategy given by (A.1).

Indeed, let $D \notin \mathbb{R}^m_-$. The equation of the corresponding supporting hyperplane is $D^+ \bullet x = 0$. Let $q \equiv q(D) \in \Delta(S^i)$ be proportional to D^+ (i.e., $q(j) := D^+(j)/d$, where $d := \sum_k D^+(k)$). Then the (expected) payoff vector x when i uses q and the choice of the other players is an arbitrary s^{-i} satisfies:

$$D^+ \bullet x = \sum_j D^+(j) \sum_k q(k)\ v(k,s^{-i})(j)$$

$$= \frac{1}{d} \sum_j \sum_k D^+(j)\ D^+(k) \left[u^i(j,s^{-i}) - u^i(k,s^{-i})\right] = 0.$$

Therefore the Blackwell condition is satisfied, and the strategy given by $p_{t+1} = q(D^i_t)$ -- which is precisely (A.1) -- yields the result. ||

Let $v^i \equiv v^i(\Gamma)$ denote the (minimax) value of the game Γ to player i:

$$v^i := \max_{\xi^i \in \Delta(S^i)} \min_{\xi^{-i} \in \Delta(S^{-i})} u^i(\xi^i,\xi^{-i}) = \min_{\xi^{-i} \in \Delta(S^{-i})} \max_{\xi^i \in \Delta(S^i)} u^i(\xi^i,\xi^{-i}).$$

(note that ξ^{-i} is a distribution over (N–1)-tuples of strategies -- *not* necessarily a product of N–1 independent mixtures). Because $\max_j u^i(j,z^{-i}_t) \geq v^i$, Theorem C implies that the long-run average payoff will be at least as large as the value. Thus

> **Corollary.** If player i plays according to the adaptive procedure given by (A.1), then

$$\liminf_{t \to \infty} \frac{1}{t} \sum_{\tau=1}^{t} u^i(s_\tau) \geq v^i \quad \text{a.s.} \tag{A.3}$$

Of course, the result of Theorem C is stronger -- since the other players may play in such a way that the best response of player i yields a higher payoff than the value.

Let us move now to the setup of the "unknown game" of Section 5, where player i knows only his set of strategies and his actual past payoffs. Baños [1968] and Meggido [1980] showed that (A.3) can still be guaranteed: player i

has a strategy that yields him an average payoff that, in the limit, is no less than the (unknown to him) value of the game. Recently, Auer *et al* [1995] have proved that a stronger result of the type (A.2) can also be guaranteed in this setup.[34]

Our adaptive procedure (A.1) can be modified to fit the unknown game setup, in the same way that (5.1) was obtained from (2.1). Fix $0 < \delta < 1$ and $0 < \gamma < 1/2$. For each $t = 1, 2, \ldots$, let $\delta_t := \delta\, t^{-\gamma}$, and define for each j in S^i

$$C_t^i(j) := \frac{1}{t}\left[\sum_{\tau \le t : s_\tau^i = j} \frac{1}{p_\tau^i(j)}\, u^i(s_\tau) - \sum_{\tau=1}^{t} u^i(s_\tau)\right], \tag{A.4a}$$

$$q_{t+1}^i(j) := \frac{[C_t^i(j)]^+}{\sum_{k \in S^i} [C_t^i(k)]^+}, \tag{A.4b}$$

$$p_{t+1}^i(j) := (1-\delta_t)\, q_{t+1}^i(j) + \delta_t \frac{1}{m^i}; \tag{A.4c}$$

again, if the denominator of $q_{t+1}^i(j)$ vanishes, take $q_{t+1}^i \in \Delta(S^i)$ arbitrary. The strategy of player i is to choose j in S^i at time $t+1$ with probability $p_{t+1}^i(j)$; at time 1, $p_1^i \in \Delta(S^i)$ is arbitrary with $p_1^i(j) \ge \delta/m^i$ for all j.

---

[34] Auer *et al* [1995] show that the inequality of (A.2) is attained in probability, with a procedure that depends on the length of the game t. Our procedure (A.4) below is independent of t, and it yields the almost sure result (A.2).

Theorem C'. If player i plays according to the adaptive procedure given by (A.4), then (A.2) (hence also (A.3)) is satisfied.

Again, no assumption is made on the behavior of the other players. The proof of Theorem C′ is similar to –– but simpler than –– the proofs of the Main Theorem and especially of Theorems B and B′. We omit it here.

REFERENCES

Auer, P., N. Cesa-Bianchi, Y. Freund and R.E. Schapire [1995], "Gambling in a Rigged Casino: The Adversarial Multi-Armed Bandit Problem", *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, 322-331.

Aumann, R.J. [1974], Subjectivity and Correlation in Randomized Strategies, *Journal of Mathematical Economics* 1, 67-96.

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

Baños, A. [1968], On Pseudo-Games, *The Annals of Mathematical Statistics* 39, 1932-1945.

Blackwell, D. [1956a], An Analog of the Minmax Theorem for Vector Payoffs, *Pacific Journal of Mathematics* 6, 1-8.

Blackwell, D. [1956b], Controlled Random Walks, *Proceedings International Congress of Mathematicians 1954*, vol. III, 335-338, Amsterdam: North-Holland.

Feller, W. [1965], *An Introduction to Probability Theory and its Applications*, *vol. II*, 2nd edition, Wiley.

Foster, D. and R.V. Vohra [1991], Asymptotic Calibration, (mimeo).

Foster, D. and R.V. Vohra [1993], A Randomization Rule for Selecting Forecasts, *Operations Research* 41, 704-709.

Foster, D. and R.V. Vohra [1996], Calibrated Learning and Correlated Equilibrium, (1st version: 1993); *Games and Economic Behavior*, (forthcoming).

Fudenberg, D. and D.K. Levine [1995], Conditional Universal Consistency, (mimeo).

Fudenberg, D. and D.K. Levine [1996], *Theory of Learning in Games*, (mimeo).

Hannan, J. [1957], Approximation to Bayes Risk in Repeated Play, in *Contributions to the Theory of Games, vol. III* (*Annals of Mathematics Studies* 39), M. Dresher, A.W. Tucker and P. Wolfe (eds.), 97-139, Princeton: Princeton University Press.

Hart, S. and D. Schmeidler [1989], Existence of Correlated Equilibria, *Mathematics of Operation Research* 14, 18-25.

Loève, M. [1978], *Probability Theory, vol. II*, 4th Edition, Springer-Verlag.

Luce, R.D. and H. Raiffa [1957], *Games and Decisions*, Wiley.

Megiddo, N. [1980], "On Repeated Games with Incomplete Information Played by Non-Bayesian Players", *International Journal of Game Theory* 9, 157-167.

Mertens, J.-F., S. Sorin and S. Zamir [1995], Repeated Games, Part A, CORE DP-9420 (mimeo).

Myerson, R.B. [1995], Dual Reduction and Elementary Games, DP-1113, Northwestern University, (mimeo).

Nau, R.F. and K.F. McCardle [1990], Coherent Behavior in Noncooperative Games, *Journal of Economic Theory* 50, 424-444.

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

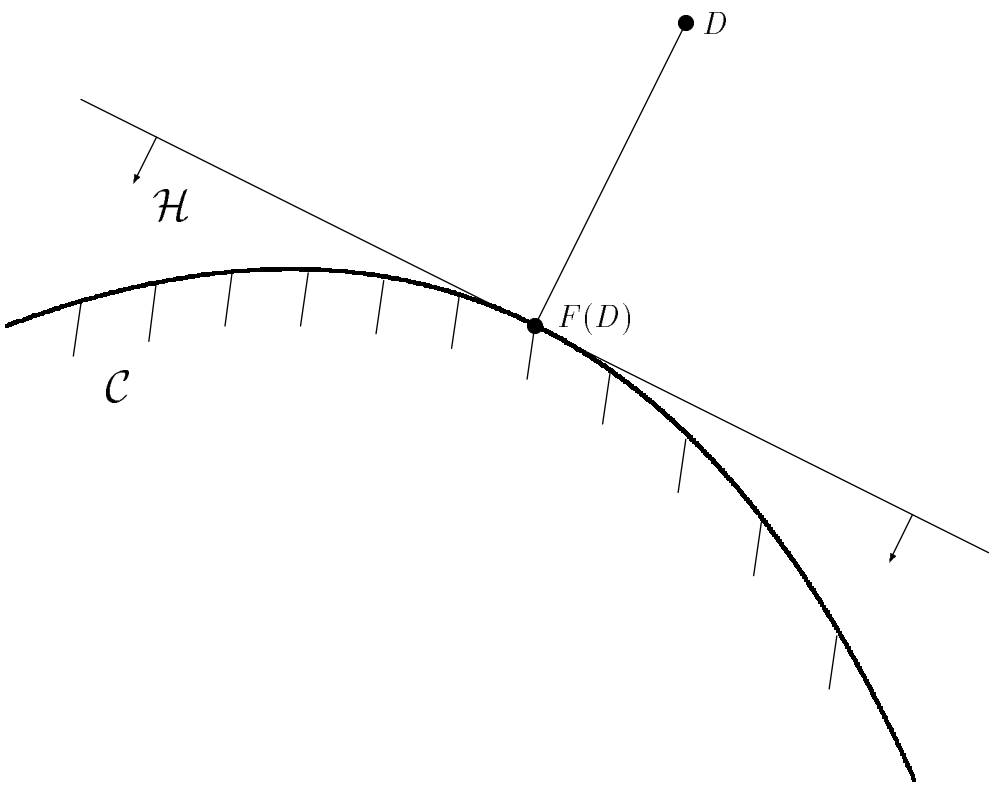

Figure 1. The Blackwell condition

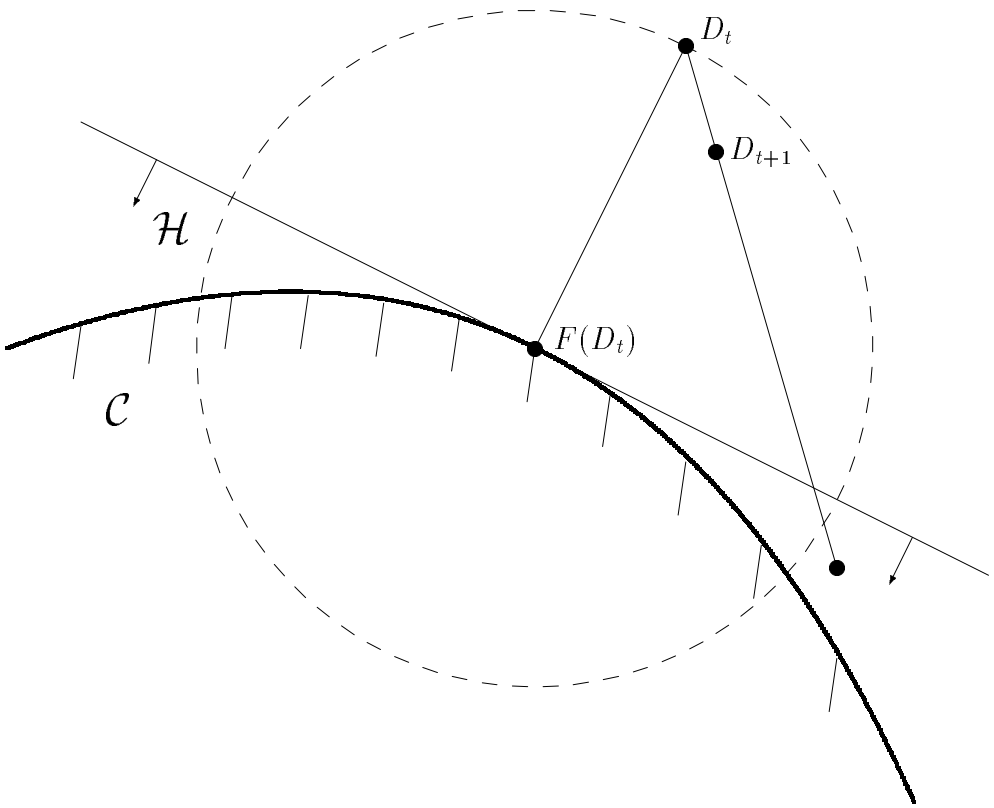

Figure 2. Approachability

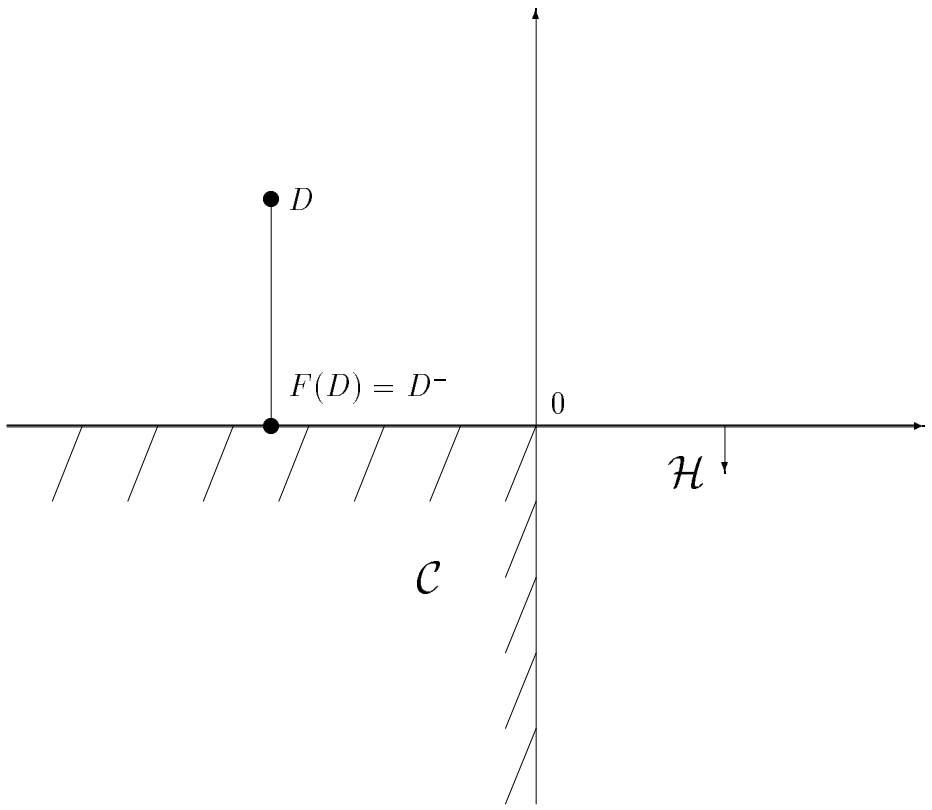

Figure 3. The Blackwell condition for R_{-}^{L}