Average Testing and the Efficient Boundary

Itai Arieli* Yakov Babichenko‡

March 31, 2011

Abstract

We propose a simple adaptive procedure for playing strategic games: *average testing*. In this procedure each player sticks to her current strategy if it yields a payoff that exceeds her average payoff by at least some fixed $\varepsilon > 0$; otherwise she chooses a strategy at random. We consider generic two-person games where both players play according to the average testing procedure on blocks of k-periods. We demonstrate that for all k large enough, the pair of time-average payoffs converges (almost surely) to the 3ε-Pareto efficient boundary.

*University of Oxford and Nuffield College. e-mail:itai.arieli@economics.ox.ac.uk

†Center for the Study of Rationality, Institute of Mathematics, The Hebrew University of Jerusalem. e-mail:yak@math.huji.ac.il.

‡We would like to thank Sergiu Hart, Ron Pertz, Eilon Solan, and Peyton Young for useful discussions and comments. Itai would like to acknowledge financial support by the US Air Force Office of Scientific Research AF 9550-09-1-0538 and the ERC 030-7950-411. Yakov would like to acknowledge financial support by ERC 030-7950-411 and ISF 039-7679-411.

1 Introduction

In a two-player strategic game, a pair of payoffs is *Pareto efficient* if there is no other feasible pair of payoffs that are better for both players. Naturally, efficiency is a prominent and desirable property in equilibrium selection, mechanism design, networks, bargaining, and many other areas. A non-efficient outcome for a game might be interpreted as paradoxical simply because there exists an outcome that is better for both players. Unfortunately, in a one-shot interaction it is not always possible to obtain an efficient equilibrium, as the well known prisoner's dilemma demonstrates.

In a repeated game framework, however, all the individually rational outcomes (particularly, but not exclusively, efficient outcomes) might be obtained in an equilibrium by the folk theorem. Achieving effiecency in this setup using a tools of equalibria selection has been investigated by Aumann and Sorin [1]. As their work reveals, finding a mechanism where the efficient outcome will be the only selected equilibrium is not easy, even in the case where there exists an action profile that maximizes the payoffs for all the players (i.e., a unique efficient outcome, which is also a pure Nash equilibrium).

Here we tackle efficiency from a dynamical perspective. Specifically, we pose the following question: *Is there a simple adaptive procedure leading to Pareto efficiency in every two-player strategic game?* We answer this question in the affirmative for a generic class of two player games. We present the average-testing dynamic that leads to an average payoff that approaches an environment of the Pareto efficient boundary.

Average-testing is a *completely uncoupled*[1] *aspiration-level based*

[1] This notion is sometimes called a payoff-based dynamic.

dynamic. That is, the strategy of each player depends only on her own past payoffs. Complete uncoupledness is a desirable dynamical property since it allows that each player have only a limited amount of information about the game (see Foster and Young [4]).

Aspiration-level formation is a guiding principle of decision theory; each player forms an aspiration level that can evolve over time. If the payoff is above the aspiration level, then the player sticks to the same action; otherwise she chooses a new action uniformly. Learning through aspiration levels is a basic intuitive behavioral procedure; indeed this learning process has recently garnered a great deal of attention in economics, biology, psychology, and computer science (see, for example, [10], [9] and [5]).

Specifically, in our case, the aspiration level evolves in accordance with the average payoff each player has received so far. That is, the player is satisfied if her current payoff is ε above her average payoff. This form of satisfycing behavior may be understood as an overestimation or overconfidence player exhibits with respect to her past performances. Namely, player evaluates her average payoff as if it were ε higher than it actually is, and determines his satisfaction level accordingly. Overestimating past performances is frequently observed empirically (see Svenson [11] for an example concerning driving skills assessment). Alternatively we can consider aspiration level that is symmetric with respect to the average, by taking ε to be a random variable which represents a mistake in the average calculation made by a player, see Remark 5 in Section 4.

The dynamic proposed above does not, however, necessarily leads the average payoff close to the Pareto efficient boundary in all games (see Section 3.1 for examples). To achieve convergence to Pareto efficiency we operate the dynamic over the k-stage game, that is, the game where every strategy of a player is just a k-tuple of strategies from the original game. Alternatively, one may think of it as a process in which, after every k periods of time, each player decides how she should play in the next k periods in accordance with her satisfaction level.[2] Essentially, our main Theorem asserts that for large enough k the average payoff will eventually be in an environment of the Pareto efficient boundary.

The connection between learning through aspiration levels and efficiency is already being established in the literature. This work consolidates this connection and present, an unequivocal relationship between the two. To the best of our knowledge, this is the first work to address a learning process that converges to the Pareto efficient boundery.

This connection was established by Karandikar Mookherjee and Ray [8], who focus on 2×2 games and characterize the asymptotic behavior of the aspiration level. In their work the aspiration level in each period is the weighted average of the previous period's level and the current payoff. They characterize the asymptotic behavior in the class of 2×2 games and, in particular, they show that in the prisoner's dilemma cooperation is formed for sufficiently slow updating of aspirations and some small tremble of probability.

In a subsequent paper Borgers and Sarin [2] use aspiration levels to examine a singled-agent learning process. They showed that aspiration level adjustments may improve the decision maker's long-run performance; however, they also demonstrate that such a process may lead to persistent deviations from expected payoff maximization by creating "probability matching" effects.

[2] Blocks division was introduced previously in the literature, see Foster and Young [4] for example.

Cho and Matsui [3] characterize the asymptotic behavior of the average payoff in a satisficing learning process applied to 2×2 symmetric games. In their work the aspiration level is also formed in accordance with the average payoff each player receives. The satisfaction of a player is determined by how "far" the average payoff is from the current payoff. That is, a player is more likely to randomize if she gets a payoff that is much smaller than the average payoff she had been getting up until then. Specifically, the probability of randomization is determined by some sort of smooth sigmoid function. Cho and Matsui use a deterministic differential approximation result to establish their main results. We conjecture that adopting their learning process to our setup will eventually lead to results similar to those establish in this paper.

A recent paper by Pardelski and Young [6] presents a completely uncoupled learning rule that selects an efficient pure Nash equilibrium in an all generic n-person game.[3] This work, also establishes a connection between a satisfycing behavior procedure and efficiency, by incorporating a technique of log linear learning.

Our paper proceeds as follows. In Section 2 we present our dynamic and main Theorem (Theorem 2). In Section 3 we give a sketch of the proof of main Theorem. A discussion follows in Section 4. In Section 5 we provide a formal proof of the Main Theorem.

2 Formal Treatment

Fix a two-player strategic game $G = (A^1, A^2, u^1, u^2)$. $A^i = \{a^i_1, \ldots, a^i_{m_i}\}$ is the finite action set of player i. $U^i : A = A^1 \times A^2 \to \mathbb{R}$ is the payoff function for player i.

[3] More precisely, their procedure selects the equalibria that maximizes the welfare.

The k-stage game that is derived from the game G is defined as follows:

Definition 1. Given a two-player strategic game G, define the k-stage game $G^k = (S^1, S^2, u^1, u^2)$ to be the game where,

- $S^i := (A^i)^k$, the action set of i, is a k-tuple of actions from the original game G.
- $u^i : S = S^1 \times S^2 \to [0,1]$ is the payoff function. Given $s^1 = (a_1, \ldots, a_k)$ and $s^2 = (b_1, \ldots, b_k)$, define
$$u^i(s^1, s^2) = \frac{1}{k} \sum_{m=1}^{k} U^i(a_m, b_m).$$

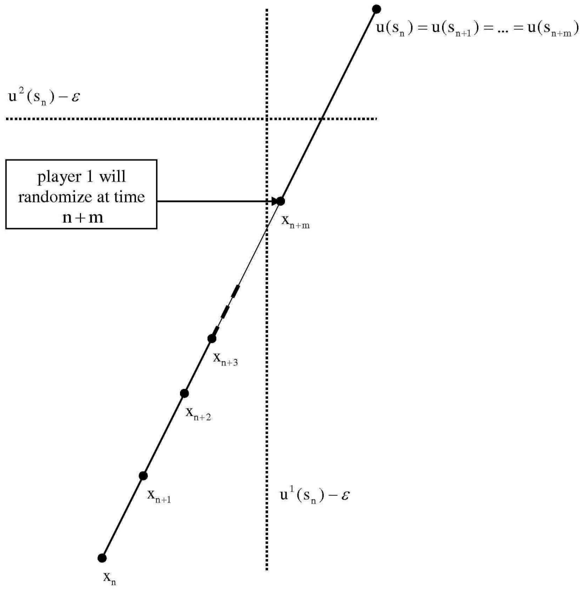
For notational convenience we omit the subscript k from the strategy set S^i; we let k be fixed throughout. Set $u(s^1, s^2) = (u^1(s^1, s^2), u^2(s^1, s^2))$.

Let s^i_n and $u^i_n = u^i(s^1_n, s^2_n)$ be player i's action and payoff at time n, and let $x^i_n = \frac{1}{n} \sum_{m=1}^{n} u^i_m$ be i's average payoff at time n. For a fixed small $\varepsilon > 0$, define the aspiration level for player i at stage n as $\alpha^i_n = x^i_n + \varepsilon$. The satisfaction or the mood of player i is determined in accordance with her current payoff and aspiration level. That is, player i is satisfied at time n if her current payoff u^i_n exceeds her aspiration level α^i_n; otherwise player i is unsatisfied. In case player i is satisfied she sticks with the action s^i_n also at time $n+1$. If, however, she is unsatisfied, then she chooses the action s^i_{n+1} uniformly among the elements of S^i.[4]

Say that a player plays in accordance with the *average testing* with parameters k, ε (write $AT(k, \varepsilon)$), if she plays the k-stage game in accordance with the above procedure.

[4] In fact, the only thing that matters in this case is that player assigns a positive probability to every pure action.

Figure 1:


u(s_n) = u(s_n+1) = ... = u(s_n+m)
u^2(s_n) − ε
player 1 will randomize at time n+m
x_n+m
x_n+3
x_n+2
x_n+1
u^1(s_n) − ε
x_n

Assume for example that both players are satisfied at time n (see Figure 1). Then they will keep on playing their current action, and as a result, the average payoff of both of them (lying on the line that connects $u_n = (u_n^1, u_n^2)$ with the point $u(s_n)$) will gradually increase. In some time $n + m$ it has to be the case that one of them (Figure 1 describes a case where player 1 would be the first to be unsatisfied) will no longer be satisfied with her payoff. At this point she will start to randomize by looking for a better action.

For the game G, we let $F(G)$ be the set of all feasible payoffs in the convex hull of the payoff matrix and let $PO(G)$ be the set of all feasible payoffs that are (weakly) Pareto efficient. That is, there is no other feasible payoff that is strictly better for both players. We let $IR(G)$ be the set of payoffs that are also purely individually rational for both players. That is, let $v^i = \max_{a^i \in A^i} \min_{a^j \in A^j} u^i(a^i, a^j)$ be the purely individually rational level of player i and

$$IR(G) = \{y = (y^1, y^2) : y \in F \text{ and } y^i \geq v^i \text{ for } i = 1, 2\}.$$

Finally, let $PIR(G) = IR(G) \cap PO(G)$ be the set of Pareto efficient payoffs that are purely individually rational for both players.

We note that $F(G), PO(G), IR(G)$, and $PIR(G)$ are equal to $F(G^k), PO(G^k), IR(G^k)$, and $PIR(G^k)$ respectively. Since G is fixed we omit the reference for G and simply write F, PO, IR, and PIR respectively.

Let $V \subseteq \mathbb{R}^2$; for $\varepsilon > 0$ we let V^ε be the set of points that lie at a distance of at most ε in the $\| \|_\infty$ norm from the set V. Similarly, V_ε is the set of points that lie at a distance of at most ε in the $\| \|_2$ norm.

We say that a sequence of points $\{y_n\}_{n=1,2,...} \subseteq \mathbb{R}^2$ converges to V if $d(y_n, V) \rightarrow_{n\rightarrow\infty} 0$ where $d(y, V)$ is the distance of the point y from the set V.

For the fixed set of strategy profiles $A = A^1 \times A^2$ we let $\mathcal{G}$ be the set of games such that every two different strategy profiles yield a different payoff and there are no three different profiles whose corresponding payoffs lie on the same line in the plane. Every game with an action profile set A, can be identified with a vector in $\mathbb{R}^{2|A|}$. Thus, the set $\mathcal{G}$ is a generic set in the sense that $\mathbb{R}^{2|A|} \setminus \mathcal{G}$ has a zero Lebesgue measure.

Our main Theorem asserts the following:

Theorem 2. For every game $G \in \mathcal{G}$ and $\varepsilon > 0$ there exists a $k_0(\varepsilon)$ such that for every $k > k_0$, if each player i plays in accordance with average testing $AT(k, \varepsilon/3)$, then the average payoff vectors converge almost surely to the set of ε-Pareto efficient and purely individually rational payoffs ($PIR_\varepsilon(G)$).

Note that the convergence of the average payoff to PIR_ε yields that $\sqrt{\varepsilon}$-efficient profiles are played with a limit proportion of at least

$1 - \sqrt{\varepsilon}$, since otherwise, by considering the distance of the average payoff from the efficient boundary, we get a contradiction. As a result we have the following corollary:

Corollary 3. For every game $G \in \mathcal{G}$ and $\varepsilon > 0$ there exists a k_0 such that for every $k > k_0$, if both players play in accordance with $AT(k, \varepsilon^2/3)$, then ε- Pareto efficient profiles in the original game G are played with a limit proportion of at least $1 - \varepsilon$.

We can choose ε small enough such that the only ε-Pareto efficient profiles in the original game will be Pareto efficient. In that case Corollary 3 guarantees that Pareto-efficient profiles are played with frequency $1 - \varepsilon$. So we have the following corollary.

Corollary 4. For every game $G \in \mathcal{G}$ there exists ε_0 and k_0 such that for every $\varepsilon < \varepsilon_0$ and every $k > k_0$, if both players play in accordance with $AT(k, \varepsilon^2/3)$, then Pareto-efficient profiles in the game G are played with a limit proportion of at least $1 - \varepsilon$.

3 Informal Sketch of the Proof

In this section we lay out informally the main ideas in the proof of our main Theorem, the proof is divided into two main parts. The first part is devoted to the choice of the right value of k_0 and the role that the k -stage game plays in our dynamic. In the second part we prove the convergence result, based on the first part.

For simplicity, we assume throughout the proof that all payoffs lie in the segment $[0, 1]$. For the ease of the exposition the payoffs on the presented examples will be integers, clearly the conclusion will not change if we multiply all payoffs by a constant.

3.1 The Choice of k_0

Let us first point out two types of cases where for $k = 1$ the average-testing dynamic does not lead anywhere close to the Patero-efficient boundary.

(a) Consider the following game:

$$\Gamma_1 : \begin{array}{c|c|c|} & L & R \\ \hline T & 2,0 & 0,2 \\ \hline B & 1,3 & 3,1 \\ \hline \end{array}$$

Assume that the average is close to the point $(1.5, 1.5)$. One can see that for this average and small enough ε, the process, dictated by the dynamic, will behave as follows:

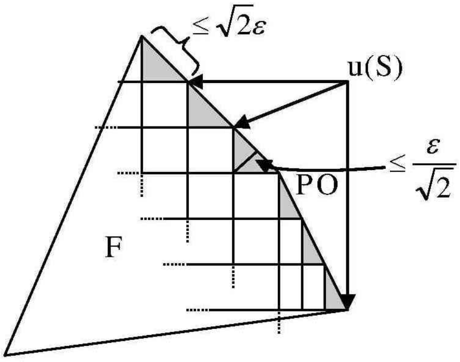
$$(T, L) \to (T, R) \to (B, R) \to (B, L) \to (T, L) \to ...,$$

where → represents the route of the dynamic. For example, if the current state is (T, R), then player 2 is satisfied and player 1 is unsatisfied. Therefore player 1 will randomize until the action B is chosen, and as a result (B, R) will be the new state. We can see that in this stochastic cycle the average for the players is $(1.5, 1.5)$, and in this case the average will converge to $(1.5, 1.5)$, which is bounded away from the Pareto-efficient boundary.

The reason for that has to do with the fact that no profile in the game dominates $(1.5, 1.5)$. The following lemma demonstrates that choosing large enough k ensures that every feasible payoff, with a large enough distance from the efficient boundary, will be dominated by a Pareto-efficient payoff of the k-stage game.

Notation 5. For convenience, denote $(a, b) >\leq (c, d)$ wherever $a > c$ and $b \leq c$. Similarly, let $>>, \leq>$ represent the appropriate relations over $\mathbb{R}^2$.

Figure 2:


$\leq \sqrt{2}\varepsilon$
u(S)
$\leq \frac{\varepsilon}{\sqrt{2}}$
PO
F

Lemma 6. For $k \geq \frac{1}{\varepsilon}$ and $x \in F$, if $x \notin PO_{\varepsilon/\sqrt{2}}$, then there exists a profile $s \in S$ such that $u(s) >> x$.

Proof. From our assumption, the distance between every pair of payoffs in G is at most $\sqrt{2}$. Therefore, for $k \geq \frac{1}{\varepsilon}$ the distance between every two adjacent payoffs on PO, in the k-stage game G^k, is at most $\sqrt{2}\varepsilon$. Set $PO(S) = \{u(s) : s = (s^1, s^2) \in S^1 \times S^2\} \cap PO$; after deleting all the feasible payoffs that are dominated by a payoff from $PO(S)$ we remain with the set E (see the shaded area in Figure 2 above), where by a simple geometric consideration we have $x \in PO_{\varepsilon/\sqrt{2}}$ for every $x \in E$. □

(b) The second problem that may arise is demonstrated using the following example:

Γ_2 :

	L	M	R
T	$\frac{3}{4}, \frac{3}{4}$	$0, 1$	$0, 1$
M	$1, 0$	$1, 0$	$0, 1$
B	$1, 0$	$0, 1$	$1, 0$

This game is a variant of a game introduced by Hart and Mas-Colell [7]. Assume that the average is close to the point $(\frac{3}{4}, \frac{3}{4})$, and the players play some action $s \neq (T, L)$. For every action $s \neq (T, L)$ exactly one of the players randomizes, and it may be seen that for every average $0 << x << (1-\varepsilon, 1-\varepsilon)$ the players will never reach the point $(\frac{3}{4}, \frac{3}{4})$, and in this stochastic process the average will converge to $(\frac{1}{2}, \frac{1}{2})$, which is not close to the Pareto boundary.

We next try to characterize this type of phenomenon.

For every $z \in \mathbb{R}^2$, let $\mathbb{P}_z$ be the Markov chain on S obtained where each player i uses a fixed aspiration level z_i. That is, player i is satisfied at time n, if and only if $u^i(s_n) > z_i$.

Definition 7. A nonempty subset $L \subseteq S$ is called *invariant* with respect to z if for every state $s \in L$, $\mathbb{P}_z(L|s) = 1$. A subset $L \subseteq S$ is called a *z-loop* if it is minimal z-invariant and $1 < |L| < |S|$.

In words, a z-loop L is a minimal invariant set that is not a singleton and not the whole state space S. Note that in the above example the set $S \setminus \{(T, L)\}$ is a z-loop for $0 << z << 1$ for $i = 1, 2$. Potentially, if the average payoff plus ε lies in this range, the Pareto-efficient boundary will not be reached.

We show in Proposition 8 that by choosing k_0 to be large enough one can avoid z-loops for every $z \in IR$.

For every game $G \in \mathcal{G}$ we can define $\alpha = \alpha(G) > 0$ to be the minimal angle between three different payoff profiles in $u(A) = \{u(a) : a \in A\}$, and $\delta = \delta(G) > 0$ to be the minimal difference between two different payoffs in $u_i(A)$.

Proposition 8. For every game $G \in \mathcal{G}$ set $k_0 = \frac{8}{\alpha\delta}$; if $k \geq k_0$, then there are no z-loops in G^k for every $z \in IR$.

The proof of Proposition 8, that relies on the unique structure that a loop poses, is relegated to Section 5.

By combining Lemma 6 and Proposition 8, we have the following corollary.

Corollary 9. For every game $G \in \mathcal{G}$ and $\varepsilon > 0$, take $k_0 = \max(\frac{1}{\varepsilon}, \frac{8}{\alpha\delta})$; then for every $k > k_0$, the game G^k has the following two properties:

1. For every average $z \notin PO_\varepsilon$ there exists a profile of the game G^k, $s = (s^1, s^2)$ such that $u(s) \in PO$ and $u(s) >> z$.
2. For every $z \in IR$ there is no z-loop.

To sum up: by choosing k_0 to be large enough we avoid the two types of problems demonstrated above. This guarantees us that whenever the average payoff $x_n \in IR^\varepsilon \backslash PO_\varepsilon$, there will be an action $s \in S$ such that $u(s) - (\varepsilon, \varepsilon)$ dominates z (first property in Corollary 9), and there will be a positive probability of reaching such an action in at most $|S|$ steps (second property in Corollary 9).

3.2 The Convergence

Let $k > k_0$ determined by Corollary 9. We prove that $AT(k, \varepsilon)$ leads to $PIR_{3\varepsilon}$, which is clearly equivalent to the argument that $AT(k, \varepsilon/3)$ leads to PIR_ε.

The proof is done in a few lemmas that investigate the behavior of the average payoff vector, x_n. We present the lemmas below and provide the main ideas of their proofs. The formal proofs are relegated to Section 5.

Let $\mathbb{P}$ be the probability distribution over all histories governed by the average-testing dynamic. First we prove that the average of every player is infinitely often above $v^i - \varepsilon$.

Lemma 10. $\mathbb{P}(x_n^i > v^i - \varepsilon \ i.o.) = 1$.

The lemma follows from the fact that every player makes infinitely many randomizations, i.e., there are infinitely many periods in which she is not satisfied, and when a player randomizes there is a positive probability that she will randomize the action that guarantees her v^i. If it happens, then she will continue to play this action at least until her average will rise above $v^i - \varepsilon$.

Given Lemma 10, we prove that for every $\delta > 0$ the average payoff $x_n \in IR^{\varepsilon+\delta}$ from some time on, with probability 1.

Lemma 11. $\forall \delta > 0,\ \mathbb{P}(\exists n_0 \text{ s.t. } \forall n \geq n_0,\ x_n^i \geq v^i - \varepsilon - \delta) = 1.$

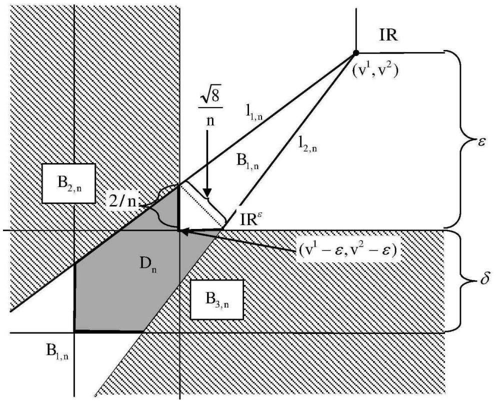
The idea of the proof is the following. When the average of a player is below $v^i - \varepsilon$, in every randomization the player can "catch" the maxmin action that guarantees him v^i, and the average will then rise above $v^i - \varepsilon$. So for time n large enough, the probability of moveing δ below $v^i - \varepsilon$ is exponentially small.

Now we prove that $x_n \in IR^\varepsilon$ infinitely often. From Lemma 10 we know that for each player $x_n^i > v^i - \varepsilon$ infinitely often. We prove that it occurs simultaneously for both players infinitely often.

Lemma 12. $\mathbb{P}(x_n \in IR^\varepsilon\ i.o) = 1.$

The idea is the following: If players "catch" an action s such that $u(s) \in IR$ and the average is close to the line $x^2 - v^2 = x^1 - v^1$, then there exists a probability, bounded away from 0, that the average will lie inside the area IR^ε. More precisely, define an area D_n (see Figure 3) that is wide enough, so that on the one hand the average x_n cannot cross it without lying inside it, and on the other hand the points in D_n are close enough to the line $x^2 - v^2 = x^1 - v^1$. If, in contrary, from some time on the average never lies inside IR^ε, then by Lemma 10 the average infinity often crosses D_n. Therefore the average infinitely often lies inside D_n (because D_n separates two areas

Figure 3:



where the average visits infinitely often). If at time n the average lies in D_n then we show that there exists a positive probability bounded away from 0 that by time $a \cdot n$ the average enters IR^ε for some fixed integer a. This completes the proof of Lemma 12.

Let $PO(S) \subset S$ be the set of actions that are not dominated by any other. Formally we define

$$PO(S) = \{s \in S | \text{ there is no } s' \in S \text{ such that } u(s') >> u(s)\}.$$

We define $E \subset F$ as follows.

$$E := \{x \in F : \text{There is no } s \in S \text{ s.t. } u(s) - (\varepsilon, \varepsilon) >> x\}$$

which is equal to

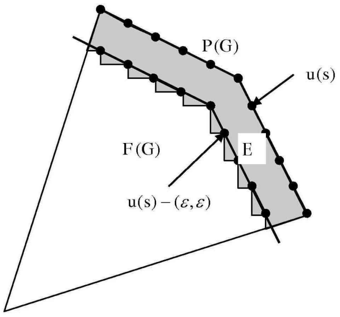
$$E := \{x \in F : \text{ there is no } s \in PO(S) \text{ s.t. } u(s) - (\varepsilon, \varepsilon) >> x\}$$

(see Figure 4).

By arguments similar to those of Lemma 6 we obtain $E \subset P_{\frac{3}{\sqrt{2}}\varepsilon}$.

We show that there exists a fixed positive probability of reaching E every time the average is $x_n \in IR^\varepsilon$. Taking this together with Lemma 11 we have

Figure 4:



Lemma 13. $\mathbb{P}(x_n \in E \textit{ i.o.}) = 1$.

The idea behind the proof is the following. By the selection of k we know that there are no loops whenever the average $x_n \in IR^{\varepsilon}$ and so every time $x_n \in IR^{\varepsilon} \setminus E$ the average is ε-dominated by an action that can be reached with positive probability (in at most $|S|$ steps). Therefore, every time $x_n \in IR^{\varepsilon} \setminus E$, we have a sequence of improvement that happen with positive probability which cause $x_m \in E$ for some $m > n$.

Finally, we prove that for every $\delta > 0$, from some time on the average lies at a distance of at most δ from the set E (in $|| \; ||_{\infty}$norm).

Lemma 14. $\forall \; \delta > 0 \; \mathbb{P}(\exists n_0 \text{ s.t.}, \forall n > n_0 \; x_n \in E^{\delta}) = 1$.

By arguments similar to those presented in the proof of Lemma 14 the probability that the average crosses a distance of δ, is exponentially small.

By the same arguments used in the proof of Lemma 6 we know that for $\delta = \varepsilon$, $E \subseteq PO_{3\varepsilon}$; therefore Lemmas 14 and 11 together prove that from some time on $x_n \in PIR_{3\varepsilon}$.

4 Remarks

1. Using the same ε for both players in the average-testing dynamic is unnecessary. The following corollary asserts that the players could use different ε and the theorem would still hold.

   Corollary 15. For every game $G \in \mathcal{G}$ and $\varepsilon > 0$ there exists a k_0 such that for every $k > k_0$, if the players play in accordance with $AT(k, \frac{\varepsilon_1}{3})$ and $AT(k, \frac{\varepsilon_2}{3})$ respectively, then the average payoff converges to the set $PIR_\varepsilon(G)$ almost surely, where $\varepsilon = \max\{\varepsilon_1, \varepsilon_2\}$.

   By similar considerations to those in the proof of the theorem we can prove that if each player plays according to $AT^i(\varepsilon_i)$ where $\varepsilon_1, \varepsilon_2 < \varepsilon$, then the average payoff will converge to the set $PIR_{3\varepsilon}(G)$.

2. **Multi-player games.** For games with more than two players, the average-testing dynamic fails. The following three-player game demonstrates the shortcomings of the average-testing dynamic in multi-player games:

4,4,0	3,3,3	0,0,0
0,0,0	0,0,0	0,0,0
0,0,0	0,0,0	0,0,0

4,4,0	0,0,0	0,0,0
4,0,4	4,0,4	4,0,4
0,0,0	0,0,0	0,0,0

4,4,0	0,0,0	0,4,4
0,0,0	0,0,0	0,4,4
0,0,0	0,0,0	0,4,4

   Note that if the players reach the payoff $(4, 4, 0)$, then they will leave it only when the average payoff to one of the players 1 or 2 rises above $4-\varepsilon$, because up to $x^1, x^2 \leq 4-\varepsilon$, players 1 and 2 get a payoff of 4 and so they won't change their action and player 3 cannot influence the payoffs for 1 and 2. When $x^i \geq 4-\varepsilon$, where i equals 1 or 2 (assume w.l.o.g. $i = 1$), then after a few randomizations the players will reach a payoff of (0,4,4) (because any other payoff is unstable). After it, the average payoff of players 2 will rise above $4-\varepsilon$. Again, after a few randomizations they will reach the payoff $(4,0,4)$ and play it until the average of player 3 is above $4-\varepsilon$. And so on. It is easy to verify that in the play described above the average is infinitely often far from Pareto efficiency (because of the existence of payoff $(3,3,3)$).

Moreover, one can see, using a similar argument to the above, that increasing the k or slightly perturbing the payoffs in the above example won't be effective.

3. **Universal k.** In Theorem 2 we choose k_0, given the game. We want universal k such that $AT(k, \varepsilon/3)$ will lead the average payoff to PIR^{ε}, for every game G.

   Let $\mathcal{H}(\varepsilon)$ be the set of games for which every two different profiles are at a distance of at least ε and an angle between any three payoffs is at least ε (in radians). The set $\mathcal{H}(\varepsilon)$ is "almost" generic in the sense that if $\varepsilon \to 0$, then the measure of the games that are not in $\mathcal{H}(\varepsilon)$ converges to 0.

   Corollary 16. For every $\varepsilon > 0$, let $k_0 = \frac{8}{\varepsilon^2}$; then for every $k > k_0$ and every game $G \in \mathcal{H}(\varepsilon)$, if both players play in accordance with $AT(k, \varepsilon)$, the average payoff converges to the set $PIR_{3\varepsilon}$ almost surely.

   The idea is that in the proof of convergence we have only used two properties of the k-stage game—the two properties of Corollary 9. To guarantee these two conditions, we can take $k \geq \max(\frac{1}{\varepsilon}, \frac{8}{\alpha\delta})$ where α is the minimal angle and δ is the minimal distance of two different payoffs. In the class $\mathcal{H}(\varepsilon)$ $\alpha \geq \varepsilon$ and $\delta \geq \varepsilon$, and so $k_0 = \frac{8}{\varepsilon^2}$ will be sufficient.

4. **A non-identical choice of k.** In fact, the conclusion of our

main Theorem remains valid under the mild assumption that each player i plays in accordance with $AT(k_i, \varepsilon)$ respectively, where $d = \gcd(k_1, k_2) > k_0(\varepsilon)$.[5] By considering blocks of size k_1k_2 that are divided to sub-blocks of size d we can see that there is a positive probability to "catch" a dominate outcome by repetition of the same sub-blocks of size d. By employing considerations similar to the ones that being used in the proof of our main Theorem, one can show that the dynamic leads to $\varepsilon - PIR$.

5. **Random ε.** Another possible interpretation to the aspiration level formation previously introduced, is that players make a computational mistakes when calculating their average payoff. Under this approach to have ε as a random small noise, rather then deterministic, is more appropriate. We note that if the random mistakes players made during the play, governed by the noise, are i.i.d. throughout time with support $[-\varepsilon_0, \varepsilon_0]$ that overlap the positive orthant, our main theorem still holds: There exists a $k_0 = k_0(\varepsilon_0)$ such that for every $k > k_0$, if the dynamic is operated on the k stage game then the average payoff converges to $3\varepsilon_0 - PIR$ (a.s).

5 The Formal Proof

5.1 Proof of Proposition 8

We start by a characterizing the structure of a loop.

For $E^1 \subset S^1$ and $E^2 \subset S^2$ we denote

$$E^1 \vee E^2 = \{(s^1, s^2) | s^1 \in E^1 \text{ or } s^2 \in E^2\} \subset S.$$

[5] $\gcd(k_1, k_2)$ is the greatest common divisor of k_1 and k_2.

Lemma 17. For every $z \in \mathbb{R}^2$ and a z-loop L there exists $E^1(L) \subset S^1$ and $E^2(L) \subset S^2$ such that $L = E^1(L) \vee E^2(L)$.

Proof. Let L be a z-loop. Set

$$E^1(L) = \{s^1 : \exists s^2 \text{ s.t., } u^1(s^1, s^2) \leq x^1 \text{ and } (s^1, s^2) \in L\},$$

and symmetrically for player 2

$$E^2(L) = \{s^2 : \exists s^1 \text{ s.t., } u^2(s^1, s^2) \leq x^2 \text{ and } (s^1, s^2) \in L\}.$$

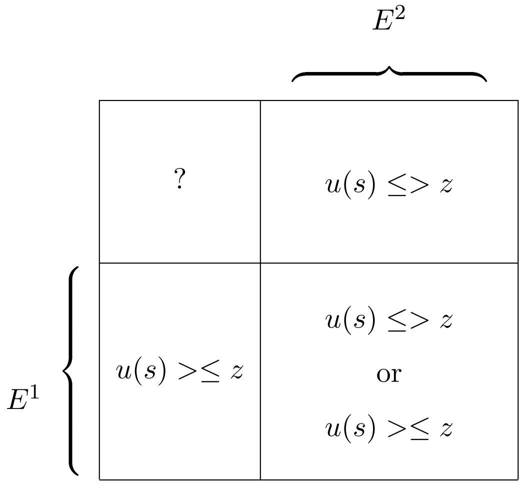
Obviously $E^1(L) \vee E^2(L) \subseteq L$. To see the other inclusion, note that by the definition of a z-loop for every $(s^1, s^2) \in L$ one of the following mast hold: $u^1(s^1, s^2) \leq z^1$ or $u^2(s^1, s^2) \leq z^2$ but not both. □

We can conclude that a z-loop L has the following structure:

- For every $s^1 \in E^1, s^2 \in S^2 \setminus E^2$ $u(s^1, s^2) >\leq z$.
- For every $s^1 \in S^1 \setminus E^1, s^2 \in E^2$ $u(s^1, s^2) \leq> z$.
- For every $s^1 \in E^1, s^2 \in E^2$ $u(s^1, s^2) \leq> z$ or $u(s^1, s^2) >\leq z$.

Where the first and second inequality symbols represent an appropriate inequality in the first and second coordinates respectively.

This structure can be deduced by the fact that in each state $s \in L$ *exactly one* of the players is satisfied and the other is unsatisfied. We can summarize the above structure using the following table:



?	$u(s) \leq> z$
$u(s) >\leq z$	$u(s) \leq> z$ or $u(s) >\leq z$

A loop in the game G^k is a complex object. It will be easier for us to focus on constant actions in the loop, i.e., actions where players play that same action k number of times. To do so we first need to prove that constant actions exist in a loop.

Lemma 18. For every $z \in IR$, and for every z-loop $L = E^1 \vee E^2 \subset S$, there exist for both players two actions in the original game $a^1_{i_1}, a^1_{i_2} \in A^1$, $i_1 \neq i_2$ and $a^2_{j_1}, a^2_{j_2} \in A^2$, $j_1 \neq j_2$ such that $(a^1_{i_1})^k, (a^1_{i_2})^k \in E^1$ and $(a^2_{j_1})^k, (a^2_{j_2})^k \in E^2$.

Proof. Note that $E^1 \neq \varnothing$, because otherwise take $s^2 \in E^2$; then for every $s^1 \in S^1$ $u(s^1, s^2) \leq> z$, which contradicts the assumption that z^2 is at least the minmax of player 2. Symmetrically we have that $E^2 \neq \varnothing$.

First we prove that there exists an action for one of the players $a^1_i \in S^1$ or $a^2_j \in S^2$ such that $(a^1_i)^k \in E^1$ or $(a^2_j)^k \in E^2$. Assume to the contrary that $(a^1_i)^k \in S^1 \smallsetminus E^1$, $(a^2_j)^k \in S^2 \smallsetminus E^2$ for every $1 \leq i \leq m^1$ and $1 \leq j \leq m^2$. Take $s^1 \in E^1$ and $s^2 \in E^2$.

For $1 \leq i \leq m_1$ let x_i be the number of times that the action a^1_i is played in the sequence s^1. From the above table, $u(s^1, (a^2_j)^k) >\leq z$;

therefore, we deduce that for every $a_j^2 \in A^2$

$$\sum_{i=1}^{m^1} \frac{x_i}{k} u^1(a_i^1, a_j^2) > z.$$

For $1 \le j \le m^2$ denote by y_j number of times that the action a_j^2 is played in the sequence s^2. From the above table, $u((a_i^1)^k, s^2) \le> z$, and so for every $a_i^1 \in S^1$

$$\sum_{j=1}^{m^1} \frac{y_j}{k} u^1(a_i^1, a_j^2) \le z$$

Therefore,

$$z_1 < \sum_{j=1}^{m^2} \frac{y_j}{k} \sum_{i=1}^{m^1} \frac{x_i}{k} u^1(a_i^1, a_j^2) = \sum_{i=1}^{m_1} \frac{x_i}{k} \sum_{j=1}^{m^2} \frac{y_j}{k} u^1(a_i^1, a_j^2) \le z_1$$

which is a contradiction.

Now, assume without loss of generality, that the player with a constant action in the loop is player 1; i.e., there exists $a_{i_1}^1 \in A^1$ such that, $(a_{i_1}^1)^k \in E^1$. There exists $a_{j_1}^2 \in A^2$ such that $u_1(a_{i_1}^1, a_{j_1}^2) \le z_1$, because $z_1 \ge v^1$. So $u_1((a_{i_1}^1)^k, (a_{j_1}^2)^k) \le z_1$, and so it follows that $(a_{j_1}^2)^k \in E^2$ and $u_2(a_{i_1}^1, a_{j_1}^2) > z_2$. By the same considerations there exists $a_{i_2}^1 \in A^1$ such that $u_2(a_{i_2}^1, a_{j_1}^2) \le z_2$ (clearly $a_{i_2}^1 \ne a_{i_1}^1$) and so $a_{i_2}^1 \in E^1$ and $u_1(a_{i_2}^1, a_{j_1}^2) > z_2$. Apply for the third time the same consideration to the action $a_{i_2}^1$, to get that there exists $a_{j_2}^2$ ($a_{j_1}^2 \ne a_{j_2}^2$) for which $(a_{j_2}^2)^k \in E^2$. □

A special case that should be considered differently in the proof of Proposition 8 is the one where each player has exactly two actions in the original game. The following lemma shows that this simple case does not cause a problem; i.e., there is no z-loop for $z \in IR$.

Lemma 19. For every game G such that $|A^1| = |A^2| = 2$, for every $k \in \mathbb{N}$, and for every $z \in IR$, there are no z-loops.

Proof. Assume by way of contradiction that there exists a z-loop $L = E^1 \vee E^2$. Let $s^1 \in S^1 \setminus E^1$ and $s^2 \in S^2 \setminus E^2$. For $i = 1, 2$ denote by x_i number of times that the action a^1_i is played in the sequence s^1 and by y_j number of times that the action a^2_j is played in the sequence s^2. By Lemma 18, $(a^1_1)^k, (a^1_2)^k \in E^1$ and $(a^2_1)^k, (a^2_2)^k \in E^2$. Therefore for $j = 1, 2$, $\sum_{i=1,2} \frac{x_i}{k} u_1(a^1_i, a^2_j) \leq z_1$. And for $i = 1, 2$ $\sum_{j=1,2} \frac{y_j}{k} u_1(a^1_i, a^2_j) > z_1$. From this it follows that

$$z_1 \geq \sum_{j=1,2} \frac{y_j}{k} \sum_{i=1,2} \frac{x_i}{k} u_1(a^1_i, a^2_j) = \sum_{i=1,2} \frac{x_i}{k} \sum_{j=1,2} \frac{y_j}{k} u_1(a^1_i, a^2_j) > z_1,$$

which is a contradiction. □

We can now prove the proposition.

Proof of Proposition 8. Recall that α is the minimal angle that is formed by 3 payoffs in the game G, and δ is the minimal distance between two payoff profiles in G. We take $k_0 = \frac{8}{\alpha\delta})$. Let us show that G^k has no z-loop for $z \in IR$.

If $|A^1| = |A^2| = 2$, then by Lemma 19, G^k has no z-loop.

In the other case where at least one player has at least 3 actions, assume without loss of generality that it is player 2 ($|A^2| \geq 3$). By way of contradiction assume that $L = E^1 \vee E^2$ is a z-loop. By Lemma 18 there exist two different actions $a^1, c^1 \in A^1$ such that $(a^1)^k, (c^1)^k \in E^1$. Denote by $B := \{u((a^1)^k, s^2) | s^2 \in S^2\} \subset u(S)$ the set of all payoffs the strategy $(a^1)^k$ yields in the game G^k.

The average z should be in the rectangle

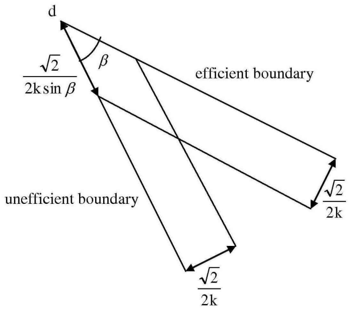
$$\{(x', y') | \min_{(x,y)\in B}(x) \leq x' \leq \max_{(x,y)\in B}(x), \min_{(x,y)\in B}(y) \leq y' \leq \max_{(x,y)\in B}(y)\}.$$

To see this, note that since $(a^1)^k \in E^1$ the set of payoffs that $(a^1)^k$ yields should include payoffs for Player 1 that are both higher and lower than z^1, and similarly for Player 2.

Let $K = PO(B)$ be the set of Pareto-efficient points with respect to $conv(B)$. Denote by $H = -PO(-B) \subset conv(B)$ the set of inefficient points in B. Every point $b \in B$ is a payoff in some state in the loop L; therefore it cannot be the case that either $b \geq\geq z$ or $b \leq\leq z$. Hence, by considerations similar to those of Lemma 6, the distance of z from both the Pareto-efficient boundary and the the inefficient boundary of $conv(B)$ is at most $\frac{\sqrt{2}}{2k}$.

Let d be a point in the intersection of the efficient boundary and the inefficient boundary. Since d is a vertex of $conv(B)$, one has $d \in u(A)$. Let β be the angle between the efficient boundary and the inefficient boundary in the point d. Note that β is an angle between some three payoffs in the game G, so $\beta \geq \alpha$. By geometric considerations it may be seen (Figure 5) that the distance between z and d is at most $\frac{\sqrt{2}}{k \sin \beta}$.

Figure 5:



Now we can apply the same considerations to the other constant action $c^1 \neq A^1$ and get the existence of some other point $e \in u(A)$ such that the distance between z and e is also at most $\frac{\sqrt{2}}{k \sin \gamma}$, where γ is an angle between some other three payoffs, and so $\gamma \geq \alpha$.

Therefore

$$\delta \leq \|d-e\|_2 \leq \|d-z\|_2 + \|e-z\|_2 \leq \frac{\sqrt{2}}{k \sin \beta} + \frac{\sqrt{2}}{k \sin \gamma} \leq \frac{2\sqrt{2}}{k \sin \alpha} < \frac{4}{k \frac{\alpha}{2}} \leq \frac{4}{\frac{8}{\alpha\delta} \frac{\alpha}{2}} = \delta$$

which is a contradiction. □

5.2 The Proof of Convergence

Proof of Lemma 10. Let i_0 be i's *maxmin* pure strategy. We note first that

$$\mathbb{P}(\exists n \geq n_0 \text{ s.t. } x_n^i \geq v^i - \varepsilon | s_{n_0}^i = i_0) = 1.$$

If $x_n^i \geq v^i - \varepsilon$, there is nothing to show; if, however, $x_n^i < v^i - \varepsilon$, then since the strategy i_0 yields only payoffs that are greater than or equal to v^i, player i will play i_0 at least until x_n^i rises above $v^i - \varepsilon$.

It is immediate that $\mathbb{P}(u^i(s_n) - x_n^i \leq \varepsilon \ i.o.) = 1$. And since we have that $\mathbb{P}(s_{n+1} = i_0 | u^i(s_n) - x_n^i \leq \varepsilon) = \frac{1}{|S^i|}$, we can deduce that $\mathbb{P}(x_n^i \geq v^i - \varepsilon \ i.o.) = 1$. □

Proof of Lemma 11. Define a sequence of events A_n by

$$A_n = \{v^i - \epsilon - \frac{\delta}{2} \leq x_n^i < v^i - \epsilon\},$$

and a sequence of stopping times $\{k_n\}_{n=1}^{\infty}$ by

$$k_n = \min\{m \geq n : x_m^i < v^i - \varepsilon - \delta \vee x_m^i \geq v^i - \epsilon\}.$$

Define B_n by

$$B_n = \{x_{k_n}^i < v^i - \delta - \epsilon\}.$$

Using the Borel-Canteli Lemma we show, that

$$\mathbb{P}(A_n \cap B_n \ i.o.) = 0.$$

We first try to bound $\mathbb{P}(B_n | A_n)$ from above. Since $|x_{n+1}^i - x_n^i| < \frac{1}{n}$ we can deduce that if $|x_m^i - x_n^i| \geq \frac{\delta}{2}$, then $m - n > \frac{n\delta}{2}$ for $m > n$. Therefore, given A_n, B_n occurrence caused by at least $\lfloor \frac{n\delta}{2} \rfloor$ periods $n \leq m < k_n$ in which $u_i(s_m) < x_m^i < v^i - \epsilon$. So we have $\lfloor \frac{n\delta}{2} \rfloor$ periods in which player i randomly chooses a strategy. Note that if in one of these periods player i chooses the *minmax* strategy i_0, then B_n occurs with probability 0. Because, if she chooses i_0, then all of her subsequent payoffs are above v^i, that will cause his average increase above $v^i - \epsilon$. Therefore,

$$\mathbb{P}((\exists n \leq m < k_n \text{ s.t. } s^i_m = i_0) \cap B_n | A_n) = 0.$$

Therefore, given A_n, B_n can occur only if in at least $\lfloor \frac{n\delta}{2} \rfloor$ periods player i chooses a random strategy that is different from i_0. Let $c = \frac{|S^i|-1}{|S^i|} < 1$, which represents the probability of randomly choosing a strategy that is different from i_0. Therefore,

$$\mathbb{P}(B_n | A_n) \leq c^{\lfloor \frac{n\delta}{2} \rfloor}.$$

Therefore,

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n \cap B_n) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n) \cdot \mathbb{P}(B_n | A_n) \leq \sum_{n=1}^{\infty} c^{\lfloor \frac{n\delta}{2} \rfloor} < \infty.$$

Using Borel-Canteli Lemma, one has $\mathbb{P}(A_n \cap B_n \ i.o.) = 0$. Every average's down crossing of the interval $[v^i - \varepsilon - \delta, v^i - \varepsilon]$ results in an occurrence of $A_n \cap B_n$ for some n. And since by Lemma 10 the average, x^i_n, is infinitely often above $v^i - \varepsilon$ we have,

$$\mathbb{P}(\{x^i_n < v^i - \varepsilon - \delta \ i.o.\}) \leq \mathbb{P}(A_n \cap B_n \ i.o.) = 0,$$

which proves the lemma.

□

Proof of Lemma 12. The event $\{x_n \in IR^\varepsilon \ i.o\}$ is a tail event, so we can assume throughout the proof that $n > \frac{16}{\varepsilon}$.

Let $l_{1,n}$ be the line that connects points (v^1, v^2) and $(v^1 - \varepsilon, v^2 - \varepsilon + \frac{2}{n})$, and $l_{2,n}$ be the line that connects points (v^1, v^2) and $(v^1 - \varepsilon + \frac{2}{n}, v^2 - \varepsilon)$. These two lines define three disjoint areas (see Figure 3):

$$\begin{aligned}
B_{1,n} &: = \{(y^1, y^2) \in C(\Gamma) | \varepsilon(v^2 - y^2) \geq (\varepsilon - \frac{2}{n})(v^1 - y^1) \text{ and } (\varepsilon - \frac{2}{n})(v^2 - y^2) \leq \varepsilon(v^1 - y^1)\} \\
B_{2,n} &: = \{(y^1, y^2) \in C(\Gamma) | \varepsilon(v^2 - y^2) < (\varepsilon - \frac{2}{n})(v^1 - y^1) \text{ and } y^1 \leq v^1 - \varepsilon\} \\
B_{3,n} &: = \{(y^1, y^2) \in C(\Gamma) | (\varepsilon - \frac{2}{n})(v^2 - y^2) > \varepsilon(v^1 - y^1) \text{ and } y^2 \leq v^2 - \varepsilon\}
\end{aligned}$$

By Lemma 10

$$\mathbb{P}(\exists\{n_i\}_{i=1}^{\infty},\{m_i\}_{i=1}^{\infty} \text{ s.t. } n_i < m_i < n_{i+1},$$
$$x^1_{n_i} > v^1 - \varepsilon \text{ and } x^2_{m_i} > v^2 - \varepsilon) = 1.$$

Consider the segment of time $[n_i, m_i]$, and assume by contradiction that $x_n \notin IR^{\varepsilon} \cup B_{1,n}$ for every $n \in [n_i, m_i]$. Then $x_n \in B_{2,n} \cup B_{3,n}$, $x_{n_i} \in B_{2,n_i}$ and, $x_{m_i} \in B_{3,m_i}$. Hence, there exists time n such that $x_{n-1} \in B_{2,n-1}$ and $x_n \in B_{3,n}$. But the distance between the sets $B_{2,n-1}$ and $B_{3,n}$ is at least $\frac{\sqrt{8}}{n}$, whereas the maximal distance between x_{n-1} and x_n is at most $\frac{\sqrt{2}}{n}$, which is a contradiction.

Now let $\delta = \min(\frac{\varepsilon}{2}, \frac{1}{2}\min_{i=1,2}(\min_{s_1,s_2\in S}|u^i(s_1) - u^i(s_2)|))$. We define another area $D_n \subset F$ (see Figure 3 on page 14):

$$D_n := (B_{1,n} \cap IR^{\varepsilon+\delta})\backslash IR^{\varepsilon}.$$

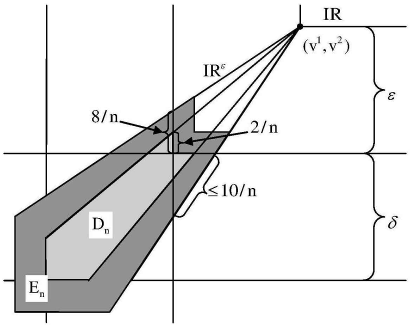
On the one hand, $\mathbb{P}(x_n \in IR^{\varepsilon} \cup B_{1,n}\ i.o) = 1$, and on the other hand, by Lemma 11 $\mathbb{P}(\exists n_0 \text{ s.t. } \forall n > n_0\ x_n \in IR^{\varepsilon+\delta}) = 1$; therefore $\mathbb{P}(x_n \in IR^{\varepsilon} \cup D_n\ i.o) = 1$. We will prove that for every $x_n \in IR^{\varepsilon} \cup D_n$ there exists a constant positive probability that $x_{n+f(n)} \in IR^{\varepsilon}$, where $f(n) \geq 0$, and this will complete the proof.

Let us define the new area E_n (see Figure 6):

$E_n := \{(y^1, y^2) \in C(\Gamma) | \varepsilon(v^2 - y^2) \geq (\varepsilon - \frac{8}{n})(v^1 - y^1) \text{ and } (\varepsilon - \frac{8}{n})(v^2 - y^2) \leq \varepsilon(v^1 - y^1)\} \cap IR^{\varepsilon+2\delta}$.

For every average $x_n \in IR^{\varepsilon+2\delta}$, if player i is satisfied with her payoff $u^i(s_n)$, then $u^i(s_n) > v^i$, because $\delta \leq \frac{1}{2}\min_{i=1,2}(\min_{s_1,s_2\in S}|u^i(s_1) - u^i(s_2)|)$. If $x_n \in D_n \subset IR^{\varepsilon+\delta}$, then $x_{n+1}, x_{n+2} \in IR^{\varepsilon+2\delta}$, and so with probability of at least $\frac{1}{|S|^2}$ in steps $n+1$ and $n+2$, both players will randomize their *maxmin* action every time when they are not satisfied, and in this scenario at step $n+2$ the players will play some action $s_{n+2} \in S$ such that $u(s_{n+2}) \in IR$. $||x_{n+2} - x_n||_2 < \frac{2\sqrt{2}}{n}$, so $d(x_{n+2}, D_n) < \frac{2\sqrt{2}}{n}$, and it follows that $x_{n+2} \in E_n$.

Figure 6:



For convenience set $m = n+2$. Let $b = \max\{\min_{i=1,2}(v^i - x^i_m - \varepsilon), 0\}$; we know that $b \leq \varepsilon$, because $\delta < \frac{\varepsilon}{2}$. Note that $\lceil \frac{mb}{\varepsilon} \rceil$ steps after m, s_m will be played with probability 1, because for $0 \leq l < \lceil \frac{mb}{\varepsilon} \rceil$ the difference between $u^i(s_m)$ and the average at step $m+l$ is

$$u^i(s_m) - \frac{m}{m+l}x^i_m - \frac{l}{m+l}u^i(s_m) = \frac{m}{m+l}(u^i(s_m) - x^i_m) \geq$$

$$\geq \frac{m}{m+l}(v^i - x^i_m) \geq \frac{m}{m+l}(b+\varepsilon) > \frac{m}{m+\frac{mb}{\varepsilon}}(b+\varepsilon) = \frac{b+\varepsilon}{1+\frac{b}{\varepsilon}} = \varepsilon.$$

Therefore with probability of at least $\left(\frac{1}{|S|}\right)^{\frac{32}{\varepsilon}}$ the action s_m will be played $\lceil \frac{mb}{\varepsilon} \rceil + \frac{32}{\varepsilon}$ steps after step m.

By the definition of E_n every payoff $y = (y^1, y^2) \in D_n$ satisfies for $i \neq j$ $i,j = 1,2$:

$$\frac{v^i - y^i}{v^j - y^j} \leq \frac{\varepsilon}{\varepsilon - \frac{8}{n}} < \frac{1}{1 - \frac{8}{m\varepsilon}},$$

which yields

$$v^i - y^i \leq \frac{1}{1-\frac{8}{m\varepsilon}} \min_{j=1,2}(v^j - y^j) \leq \frac{1}{1-\frac{8}{m\varepsilon}}(b+\varepsilon).$$

Now let us compute the difference between v^i and the average at step $m + \lceil \frac{mb}{\varepsilon} \rceil + \frac{32}{\varepsilon}$:

$$v^i - x^i_{m+\lceil \frac{mb}{\varepsilon} \rceil + \frac{32}{\varepsilon}} = v^i - \frac{m}{m+\lceil \frac{mb}{\varepsilon} \rceil + \frac{32}{\varepsilon}}x^i_m - \frac{\lceil \frac{mb}{\varepsilon} \rceil + \frac{32}{\varepsilon}}{m+\lceil \frac{mb}{\varepsilon} \rceil + \frac{32}{\varepsilon}}u^i(s_m) \leq$$

$$\leq v^i - \frac{m}{m+\lceil\frac{mb}{\varepsilon}\rceil+\frac{32}{\varepsilon}}x^i_m - \frac{\lceil\frac{mb}{\varepsilon}\rceil+\frac{32}{\varepsilon}}{m+\lceil\frac{mb}{\varepsilon}\rceil+\frac{32}{\varepsilon}}v^i = \frac{m}{m+\lceil\frac{mb}{\varepsilon}\rceil+\frac{32}{\varepsilon}}(v^i - x^i_m) \leq$$

$$\leq \frac{m}{m+\frac{mb}{\varepsilon}+\frac{32}{\varepsilon}}(v^i - x^i_m) = \varepsilon\frac{(v^i - x^i_m)}{\varepsilon+b+\frac{32}{m}} \leq \varepsilon\frac{\frac{1}{1-\frac{8}{m\varepsilon}}(b+\varepsilon)}{\varepsilon+b+\frac{32}{m}} = \varepsilon\frac{\frac{1}{1-\frac{8}{m\varepsilon}}}{1+\frac{32}{m(b+\varepsilon)}} \leq$$

$$\leq \varepsilon\frac{1}{(1-\frac{8}{m\varepsilon})(1+\frac{32}{m\cdot 2\varepsilon})} = \varepsilon\frac{1}{1+\frac{8}{m\varepsilon}(1-\frac{16}{m\varepsilon})} \leq \varepsilon.$$

So if s_m is played $\lceil\frac{mb}{\varepsilon}\rceil+\frac{32}{\varepsilon}$ steps after step m, then $x^i_{m+\lceil\frac{mb}{\varepsilon}\rceil+\frac{32}{\varepsilon}} \in IR^\varepsilon$, and it occurs with positive constant probability $\left(\frac{1}{|S|}\right)^{\frac{32}{\varepsilon}}$. □

Proof of Lemma 13. Recall that at time n the process behaves like the Markov chain $\mathbb{P}_{\alpha^n}$, where $\alpha_n = x_n + \varepsilon$. For every average $x_n \in IR^\varepsilon \backslash E$ there is no α_n-loop; therefore every invariant set of the Markov chain $\mathbb{P}_{\alpha_n}$ includes an action $s \in S$ such that $u(s) - (\varepsilon, \varepsilon) >> x_n$; i.e., there is a positive probability of at least $\frac{1}{|S|^{|S|}}$ of achieving such an action in at most $|S|$ steps.

Assume $x_n \in IR^\varepsilon \backslash E$, and consider the event where at each time $m > n$ where $x_m \notin E$ and not $u(s_m) - (\varepsilon, \varepsilon) >> x_m$, the players reach in at most $|S|$ steps an action $s \in S$ such that $u(s) - (\varepsilon, \varepsilon) >> x_m$; [6] note that clearly $s \neq s_m$. This event occurs with a probability of at least $\frac{1}{(|S|^{|S|})^{|S|}}$, and subsequently $x_m \in E$.

By Lemma 12 $P(x_n \in IR^\varepsilon\ i.o) = 1$, and as we proved above, if $x_n \in IR^\varepsilon$ then with a probability of at least $\frac{1}{|S|^{(|S|^2)}}$ there exists $m > n$ such that $x_m \in E$; therefore $P(x_n \in E\ i.o.) = 1$. □

Proof of Lemma 14. We will prove it for $\delta < \frac{1}{2}\min\limits_{s_1,s_2 \in S}|u^i(s_1) - u^i(s_2)|$, and then clearly it holds also for every $\delta' > \delta$ because $E^\delta \subset E^{\delta'}$.

[6] During the $|S|$ steps, the average x_m is changed. So there could be a situation where the Markov chain $\mathbb{P}_{\alpha_m}$ changed in steps $m+1, m+2, ..., m+|S|$, and a path to the desired action s no longer exists in the new chain. To avoid this problem we cam assume that the action s_m continues to be played $\frac{|S|}{\varepsilon}$ steps more from the moment that it is no longer the case that $u(s_m) - (\varepsilon, \varepsilon) >> x_m$. This happens with a probability of at least $\left(\frac{1}{|S|}\right)^{\frac{|S|}{\varepsilon}}$.

By the choice of δ, we know that for every $x_n \in E^{\delta} \backslash E$ and action $s \in S$ such that $u(s) - (\varepsilon, \varepsilon) >> x$ satisfies $u(s) - (\varepsilon, \varepsilon) \in E$, which means that if the current average is $x_n = x$, and the players play an action s such that they are both satisfied, then they will play it with probability 1, until the average enters the set E.

From here on, the proof will be very similar to the proof of Lemma 11.

We define a sequence of events A_n by

$$A_n = \{0 < ||E - x_n||_{\infty} \leq \frac{\delta}{2}\}.$$

Define a sequence of stopping times $\{k_n\}_{n=1}^{\infty}$ by

$$k = \min\{m \geq n : x_m \in A_m\}.$$

Define B_n by

$$B_n = \{||E - x_{k_n}|| > \delta\}.$$

Let us bound from above $\mathbb{P}(B_n | A_n)$. Given A_n, B_n occurrence caused by at least $\lfloor \frac{n\delta}{2} \rfloor$ periods $n \leq m < k_n$ in which no action s such that $u(s) - (\varepsilon, \varepsilon) >> x_m$, has been played (because otherwise $||E - x_{k_n}|| = 0$). But for any average $x_m \notin E$, there is a positive probability of $c := \frac{1}{|S|^{|S|}}$ to reach such an action s in $|S|$ steps. So, for $n > \frac{4|S|}{\delta}$, the probability that such an action s will not be played in steps $n + |S|, n + |S| + 1, ..., n + |S| + \lfloor \frac{n\delta}{4} \rfloor$ is at most $(1 - c)^{\lfloor \frac{n\delta}{4} \rfloor}$; i.e., $\mathbb{P}(B_n | A_n) \leq (1 - c)^{\lfloor \frac{n\delta}{4} \rfloor}$. Therefore:

$$\sum_{n=\frac{4|S|}{\delta}}^{\infty} \mathbb{P}(A_n \cap B_n) = \sum_{n=\frac{4|S|}{\delta}}^{\infty} \mathbb{P}(A_n)\mathbb{P}(A_n | B_n) \leq \sum_{n=\frac{4|S|}{\delta}}^{\infty} (1 - c)^{\lfloor \frac{n\delta}{4} \rfloor} < \infty$$

and by the Borel-Cantely lemma $P(A_n \cap B_n \ i.o.) = 0$. Whenever $x_n \in E$ and $x_m \notin E^{\delta}$ $m > n$, $A_n \cap B_n$ occurs for some n. By Lemma 13 we have

$$\mathbb{P}(x_n \notin E_{\delta} \ i.o.) = \mathbb{P}(x_n \notin E_{\delta} \ i.o. | x_n \in E \ i.o.) \leq \mathbb{P}(A_n \cap B_n \ i.o.) = 0.$$

□

References

[1] Aumann, R.J. and Sorin, S. (1989) "Cooperation and Bounded Recall." *Games and Economic Behavior* 1, 5–39.

[2] Börgers, T. and Sarin, R. (2000) "Naive Reinforcement Learning with Endogenous Aspirations." *International Economic Review* 4, 921–950.

[3] Cho, I. and Matsui, A. (2005) "Learning Aspiration in Repeated Games." *Journal of Economic Theory* 2, 171–201.

[4] Foster, D. and Young P.H. (2006) "Regret Testing: Learning to Play Nash Equilibrium without Knowing you have an Opponent." *Theoretical Economics* 1, 341–367.

[5] Motro, O. and Shmida A. (1995) "Near-far search: an evolutionarily stable foraging strategy." *Journal of Theoretical Biology*, 173, 15-22.

[6] Pradelski, B. and Young, H.P. (2010) "Efficiency and Equilibrium in Trial and Error Learning." University of Oxford. *Economics Series Working Papers* 480.

[7] Hart, S. and Mas-Colell, A. (2006) "Stochastic Uncoupled Dynamics and Nash Equilibrium." *Games and Economic Behavior* 2, 286–303.

[8] Karandikar, R. Mookherjee, D. Ray, D. and Vega-Redondo, F. (1998) "Evolving Aspirations and Cooperation." *Journal of Economic Theory* 2, 292–331.

[9] Posch, M. Pichler, A. and Sigmund, K. (1999) "The Effciency of Adapting Aspiration Levels." *Proceedings of Biological Science* 226, 1427–1435.

[10] Simon, H.A., (1955) "A Behavioral Model of Rational Choise." *The Quarterly Journal of Economics* 69, 99–118.

[11] Svenson O., (1981) "Are we all less risky and more skillful than our fellow drivers?" *Acta Psychologica* 47, 143–148.