Satisficing Leads to Cooperation in Mutual Interests Games*

by

Amit Pazgal+

May 1995

Abstract

We study the play of mutual interests games by satisficing decision makers. We show that, for a high enough initial aspiration level, and under certain assumptions of "tremble," there is a high probability (close to unity) of convergence to the Pareto dominant cooperative outcome. Simulations indicate that the theoretical result is robust with respect to the "trembling" mechanism.

---

* I am grateful to Itzhak Gilboa for his guidance. I also wish to thank Eddie Dekel, Ehud Kalai, and David Schmeidler for comments and suggestions.

+ Department of Managerial Economics and Decision Sciences, Kellogg Graduate School of Management, Northwestern University, Evanston IL, 60208. E-mail: pazgal@nwu.edu

1. Introduction

A central question in game theory is under what conditions will agents who act solely in their own interests evolve to cooperate. One reason for cooperation failure is the inability of agents to agree on a preferred outcome. It is therefore natural to focus first on games with a Pareto dominant outcome, and ask whether agents would cooperate to obtain it.

Games with Mutual Interests[1] (*MI*) are games in which there exists an outcome which is strictly preferred by every player to any other outcome. Even though the efficient outcome is always a Nash equilibrium outcome, there are reasons to doubt that it would actually be chosen. For instance, Harsanyi and Selten's [1988] risk dominance criterion might select a different equilibrium. Similarly, Aumann [1990] argues that in the "stag hunt" game the cooperative outcome might not be a reasonable prediction.

However, if the *MI* game is repeated, the Pareto dominant outcome is more compelling. Indeed, many authors have tried to derive it as the unique prediction for the repeated game. Some authors have appealed to players' rationality. For instance, Osborne [1990] used forward induction to achieve the pareto optimum for a class of repeated pure cooperation two by two games. Balkenborg [1993] studies two player common interests games and uses strictness and stability concepts to guarantee that players achieve the efficient outcome.

Following the derivation of the cooperative outcome in the prisoner's dilemma from bounded rationality (Neyman [1985], Rubinstein [1986], and Abreu and Rubinstein [1988]), several authors have attempted to employ bounded rationality arguments in *MI* games as well. Aumann and Sorin [1989] single out cooperation in a two player repeated common interests game by using players with bounded memory and "trembling hand" perturbations. Binmore and Samuelson [1992] show that a modified evolutionary stable strategy in a repeated symmetric two person game must maximize the sum of the payoffs to both players. Anderlini and Sabourian [1990] derive payoffs which are arbitrarily close to the efficient payoffs in games were players are restricted to choose strategies which are implementable by Turing machines. Kandori, Mailath and Rob [1993] and Young [1993] assume noisy imitations of successful strategies in games between randomly matched

---

[1] Aumann and Sorin [1989] differentiate between games with Mutual Interests and games with **Common Interests** (*CI*), where the unique pareto optimal payoff can be achieved via several different strategy profiles.

players chosen out of a large population. On the sub-class of pure coordination games their results imply that as the "mutation" rate tends to zero, the limit frequency of the *MI* outcome tends to unity.

Furthermore, several papers have shown that the application of an evolution inspired solution concept to a cheap talk game would result in the Pareto dominant outcome, if such exists. For example, Kim and Sobel [1991] consider a subclass of *CI* games where the pure strategy Nash Equilibria are Pareto ranked. Similar results were obtained by Wärneryd [1990] (sender receiver games) , Matsui [1991] (pure cooperation), and Schlag [1993] (symmetric partnership games).

In this paper we take the bounded rationality approach to an extreme. Players are so naïve that they do not even realize that they are participating in a game; all they know is that, at every stage, they choose an action and get a payoff. We assume that each player has an aspiration level which can be interpreted as a payoff level she would like to get. At every stage each player adopts an action which has the highest cumulative payoff relative to her aspiration level. If a certain action was never tried by the player, she evaluates it by the aspiration level. Thus she is "satisficing" à la Simon [1955] and March and Simon [1958]. Namely, she will not attempt to optimize as long as the "familiar" actions obtain the aspiration level. On the other hand, among the actions that were chosen, the player does "optimize" by choosing a maximizer of the cumulative payoff. The latter is re-scaled relative to the aspiration level. (I.e., the action is evaluated by the sum of the differences between experienced payoffs and the current aspiration level.) Thus, our notion of an "aspiration level" can also be viewed as a "reference point."[2]

Cumulative payoff is arguably one of the simplest possible evaluation rules. It may be viewed as an index which summarizes a player's memory of the action's performance. It should be noted that this rule is not invariant with respect to "shifts" of the utility function. By contrast, traditional game theory chooses the "zero" on the utility scale arbitrarily. It therefore makes sense to enrich the model by the inclusion of "aspiration levels" or "reference points"—which might differ from the arbitrarily chosen "zero." On a more conceptual level, the impression that an action forms in the player's mind depends on her implicit expectations.

Our decision rule pre-supposes less information and less "rationality" than the existing learning/evolution "myopic best response" rules or even "fictitious play." In those

---

[2] Note, however, that it differs from Kahneman and Tversky's [1984] "reference point" in that the latter refers to an agent's most recent wealth or endowment.

models players know the game, observe the past path of play, form implicit beliefs about the other players strategies, and myopically choose a best response relative to these beliefs. Under our rule players need only remember the cumulative payoff they achieved from each of their actions, the number of times they played each one, and their current aspiration level.

We further assume that players are ambitious at the beginning of the game and "hope" to achieve high payoffs, but they become more and more realistic as the game progresses. They update their initial aspiration level towards the maximum payoff they have received in the past. Thus our agents are case-based decision makers (Gilboa and Schmeidler [1992]) and their aspiration level updating rule is similar to the case-based optimization one (Gilboa and Schmeidler [1993]). In contrast, Bendor, Mookherjee and Ray [1994] explore long-run equilibrium notions where players use the "satisficing" rule with fixed aspiration levels, which turn out to be the long run payoffs. Shoham and Tennenholtz [1993] introduce the notion of co-learning to the artificial intelligence literature. Co-learning refers to a process in which several agents simultaneously try to adapt to each other's behavior so as to produce desirable global system properties. Their agents use the "Highest Cumulative Reward" rule, which is equivalent to our rule with a fixed aspiration level of zero. They prove convergence to the desirable outcome in a class of symmetric two by two games, with random matching and finite memories.

In the following we show that in a mutual interests game, if players are sufficiently ambitious at the beginning, and if they "tremble" whenever they switch an action, then, after a finite stage of experimentation, they will settle down and play the Pareto dominant action profile forever after (with probability which is arbitrarily close to one). Specifically, we assume that at every stage each player computes the cumulative payoff of each action relative to the current aspiration level. If the "current" (most recently played) action has the best past performance, the player keeps playing it. If, however, it is not optimal in the above sense, the player will switch to a maximizer of the cumulative payoff with probability $(1-\varepsilon)$, whereas with probability ε, she will "tremble" and choose a random action.

One may imagine our players as implementing their strategies by machines that always have one button pressed. Sticking to the same choice in the stage game does not involve an explicit action, while switching does. More generally, this assumption of an asymmetric noise may be justified by "inertia" or by preference for the "status quo." (See Anderlini and Ianni [1993].) Gilboa and Schmeidler [1993] do not assume any "trembles" to get their optimization result. Rather, the "noise" in their model is generated by a high

aspiration level. One might have conjectured that this would suffice in a game situation as well. This is not the case since a high aspiration level only guarantees that each player tries all her strategies; it does not imply that the players will try all strategy combinations. At any rate, the assumption that some noise may exist appears rather realistic.

While our particular assumption regarding the noise is crucial to the proof of our result, computer simulations indicate that convergence might be possible under a variety of alternative assumptions. In particular, we examine the case of a "uniform tremble," that is, a tremble which may occur at any stage (regardless of switching), and the absence of noise altogether. In both cases, convergence to the *MI* profile is not guaranteed. However, when the noise is small or non-existent, convergence appears to be the rule rather than the exception. Specifically, games with randomly generated payoffs yield high frequency of the *MI* outcome. For instance, one may reject the hypothesis that the frequency is less than 95%.

In summary, this paper may be viewed as an exercise in modeling extremely bounded rationality. It shows that cooperation may evolve even if players are not aware of the interactive nature of the situation. Ironically, it appears that the more "rational" players are assumed to be, the stronger are the assumptions needed to explain cooperation.

The rest of the paper is organized as follows. Section 2 presents the formal model. In section 3 we prove the main result. Section 4 is devoted to some numerical simulations. Section 5 concludes.

2. Model and Results

We study n-player strategic form games. The set of players is denoted by $I = \{1, \cdots, n\}$. Let A^i be a finite set of pure actions for player i and let $A = A^1 \times A^2 \times \cdots \times A^n$ be the set of pure action profiles. To avoid trivial cases it will be assumed throughout that there are at least two players and every player has at least two pure actions. We further assume that players are only allowed to choose pure actions. Let $u^i : A \to \Re$ specify the (stage) payoff, $u^i(a)$, for player i when the action profile $a \in A$ is chosen. Let $\overline{u}^i$ denote the highest payoff that player i can obtain in the stage game. Assume that the stage game is one of *mutual interests*. Thus there exists a unique pure action profile which gives the highest possible payoff to all the players.

Let $a^* \in A$ be the Mutual Interests Nash Equilibrium action profile,

$$u^i(a^*) \equiv \overline{u}^i > u^i(a) \quad \forall a \in A;\ a \neq a^*$$

Imagine that the players play the game repeatedly, at dates t=1,2,.... We describe the set of possible finite game paths by:

$$S = \left\{ \left(t, \left(a_1, \cdots, a_t \right) \right) \middle|\ t \geq 0,\ a_\tau \in A\ \ \tau = 1, \cdots, t \right\}$$

Thus a game path is a list of the action choices of the players for every period. Every player is assumed to have an initial aspiration level H_0^i, which describes the payoff she hopes to get at every stage of the game. As the game progresses the players become more and more "realistic," and at every stage they adopt a new aspiration level which is a weighted average of their previous aspiration level and the maximum payoff they encountered. Formally, let $\alpha^i \in (0,1)$ be the adjustment rate at which player *i* updates her aspiration level.

Given a specific finite history $s = \left(t, \left(a_1, \cdots, a_t \right) \right)$, we can now define (recursively) the aspiration level of player *i* at the beginning of period t given history s:

$$\begin{aligned} &For \quad t = 0 \qquad H^i(s) = H_0^i \\ &For \quad t > 0 \qquad M^i(s) = \underset{\tau \leq t}{Max}\ u^i(a_\tau) \\ &\qquad\qquad\quad H^i(s) = \alpha^i \cdot H^i(s') + \left(1 - \alpha^i\right) \cdot M^i(s) \end{aligned}$$

where $s' = \left(t-1, \left(a_1, \cdots, a_{t-1} \right) \right)$ and. $\alpha^i \in (0,1)$

After every history $s = \left(t, \left(a_1, \cdots, a_t \right) \right)$ player *i* evaluates each of her pure actions according to her cumulative payoff from playing the action in the previous $(t-1)$ stages relative to her current aspiration level. We thus define the functional $U^i(s, a^i)$:

$$\begin{aligned} &For \quad t = 0 \quad U^i(s, a^i) = 0 \\ &For \quad t > 0 \quad U^i(s, a^i) = \sum_{\tau=1}^{t} I_{a_\tau^i = a^i} \cdot \left(u^i\left(a_\tau\right) - H^i(s) \right) \end{aligned}$$

where $I_{a_\tau^i = a^i}$ is the indicator function that gets the value 1 if player *i* played a^i in period τ (and zero otherwise).

Thus, for every pure action, player i adds up the difference between her payoff playing the action and her *current* aspiration level. At the beginning of the game, when the player was very ambitious, her payoffs may seem very unsatisfactory but as she becomes more realistic she re-evaluates past performance relative to today's aspirations.

Being of bounded rationality, player i would like to myopically choose (at period t) the pure action with the highest $U^i(s,a^i)$. In our model we include some trembles. If the player decides to choose the same action again she will play it with certainty; however, if she would like to switch to a different action there is an ε chance that she may tremble and end up with another one. Thus, effectively, player i plays a mixed action, σ^i, at every stage.

It will prove convenient to define for every $i \in I$ and $\varnothing \neq B \subseteq A^i$:

$$\delta^i_B(a^i) = \begin{cases} \frac{1}{|B|} & if \ \ a^i \in B \\ 0 & otherwise \end{cases}$$

(note that δ^i is an element of the $|A^i|$ dimensional simplex)

At the initial state $s_0 \equiv (0,(\ ))$ player i randomizes uniformly among all of her actions :

$$\sigma^i(s_0) = \delta^i_{A^i}$$

Given a finite history $s = (t,(a_1,\cdots,a_t))$ with $t > 0$ player i's mixed action would be:

$$\sigma^i(s) = \begin{cases} \delta^i_{\{a^i_t\}} & a^i_t \in \arg\max U^i(s,\cdot) \\ (1-\varepsilon)\cdot\delta^i_{\arg\max U^i(s,\cdot)} + \varepsilon\cdot\delta^i_{A^i} \end{cases}$$

where $\varepsilon \in (0,1)$.

On the set of all possible game paths we define a Markov chain with the following transitional probabilities:

$$P(s,s') = \begin{cases} \prod_{i=1}^{n} \sigma^i(s)(a^i_{t+1}) & if \quad s = (t,(a_1,\cdots,a_t)) \ and \ s' = (t+1,(a_1,\cdots a_t,a_{t+1})) \\ 0 & otherwise \end{cases}$$

Define the set of all possible infinite game histories:

$$\Omega \equiv \left\{ \omega = (s_0,s_1,\cdots,s_t,\cdots) \middle| \ P(s_t,s_{t+1}) > 0 \quad \forall t \right\}$$

For $\omega \in \Omega$, ω_t denotes the t-th component of ω, i.e., for $\omega = (s_0, \cdots, s_t, \cdots)$, $\omega_t = s_t$.

Endow Ω with the σ-algebra, Σ, generated by finite histories and let μ be the probability measure on Ω induced by the transition matrix P.

Consider the set of states of the world, ω, at which the mutual interest profile, a^*, is the only one played after a certain time:

$$G \equiv \left\{ \omega = (s_0, s_1, \cdots, s_t, \cdots) \in \Omega \,\middle|\, \exists T \;\; \forall t > T,\; \omega_t = \left(t, (a_1, \cdots, a_t)\right) \; has \;\; a_t = a^* \right\}$$

(Note that it is Σ-measurable)

We can now formulate our main result:

Theorem : Given $\eta > 0$ there exists $M \in \Re$ such that for all $H_0 = \left(H_0^1, \cdots, H_0^n\right)$ with $H_0^i > M \;\; \forall i \in I$, the induced Markov chain satisfies $\mu(G) \geq 1 - \eta$.

Observe that the definition of the Markov chain, and therefore of (Ω, Σ, μ), depends on H_0. The theorem thus applies to all Markov chains whose H_0 is high enough. (To simplify notation, we suppress H_0 from the symbols P, Ω, Σ, μ.)

The theorem guarantees that, using our myopic rule, with high enough initial aspiration levels, the players will play only the mutual interests equilibrium (after a finite period of time) with a probability which is as close as we want to unity. Note that the result is independent of the size, ε, of the tremble as long as it is strictly positive.

3. Proof

Given a state $s = \left(t, (a_1, \cdots, a_t)\right)$, define the number of times player *i* played action a^i:

$$K^i(s, a^i) = \#\left\{ \tau \leq t \,\middle|\, a_\tau^i = a^i \right\}$$

In the same way we define: $K(s, a) = \#\left\{ \tau \leq t \,\middle|\, a_\tau = a \right\}$

Let $\eta > 0$ be given.

Lemma 1: There exist $M \in \Re_+$ and an integer k_1 such that for all $H_0 = \left(H_0^1, \cdots, H_0^n\right)$ with $H_0^i > M \;\; \forall i \in I$, we have $\mu\left(\left\{ \omega \in \Omega \,\middle|\, K\left(\omega_{k_1}, a^*\right) > 0 \right\}\right) \geq 1 - \eta$.

Thus if all players start with high aspiration levels the probability that they will play the mutual interest profile at least once during the first k_1 periods is arbitrarily close to unity.

Proof : Let k_0 be a large enough integer such that, if all the players switch actions simultaneously at least k_0 times, the probability that they have played a^* is bigger than $1-\eta$. Specifically, let k_0 satisfy:

$$\left(1-\prod_{i\in I}\frac{\varepsilon}{|A^i|}\right)^{k_0} < \eta \quad .$$

(When all players switch actions at the same period, there is a probability of at least $\prod_{i\in I}\frac{\varepsilon}{|A^i|}$ that they will play a^*)

We wish to find a lower bound, M, on the initial aspiration level, such that the players will choose to switch actions almost every period. Roughly, when the aspiration level is very high, all the stage game payoffs are similarly dissatisficing, and each player will choose an action that was played a minimal number of times. This implies that every player chooses to switch an action often enough. The only times that a player can play the same action twice in a row is just after she played all of her actions the same number of times. Thus during every $1+\prod_{i\in I}|A_i|$ periods there must be at least one period where all the players want to switch actions simultaneously.

Formally, let $k_1 = k_0 \cdot \left(1+\prod_{i\in I}|A_i|\right)$. We will now find M_1 such that whenever $H_t^i > M_1$ (for all i) we have :

$$\left|K^i\left(\omega_t, a^i\right) - K^i\left(\omega_t, b^i\right)\right| \le 1 \quad \forall t \le k_1 \ \forall \omega \in \Omega \ \forall i \in I \ \forall a^i, b^i \in A^i .$$

Let u^* and u_* be (respectively) the highest and lowest payoffs to any player in the stage game. It is enough to show that $\forall t \le k_1 \ \forall \omega \in \Omega \ \forall i \in I \ \forall a^i, b^i \in A^i$,

$$K^i\left(\omega_t, a^i\right) \ge K^i\left(\omega_t, b^i\right) + 1 \quad \Rightarrow \quad U^i\left(\omega_t, b^i\right) > U^i\left(\omega_t, a^i\right).$$

Set $M_1 = k_1 \cdot u^* - \left(k_1 - 1\right)\cdot u_*$. As long as $H_t^i > M \ \ \forall i \in I$, we get

$$U^i\left(\omega_t, b^i\right) > \left(t-1\right)\cdot\left(u_* - H_t^i\right) > t\cdot\left(u^* - H_t^i\right) > U^i\left(\omega_t, a^i\right)$$

$$\forall t \le k_1 \ \forall \omega \in \Omega \ \forall i \in I \ \forall a^i, b^i \in A^i .$$

We finally turn to choose M such that, if $H_t^i > M \ \forall i \in I$, then $H_t^i > M_1 \ \forall i \in I, t \le k_1$.

Note that $\left(H_t^i - u_*\right) \ge \alpha^i \cdot \left(H_{t-1}^i - u_*\right) \ge \left(\alpha^i\right)^t \cdot \left(H_0^i - u_*\right) \ge \left(\alpha^i\right)^{k_1} \cdot \left(H_0^i - u_*\right)$

We wish this expression to be greater then $M_1 - u_*$; thus we let

$$M = \frac{M_1 - u_*}{\left(\alpha_*\right)^{k_1}} + u_* = \frac{k_1 \cdot \left(u^* - u_*\right)}{\left(\alpha_*\right)^{k_1}} + u_*.$$

where $\alpha_* = \min_{i \in I} \alpha^i$.

Hence we found M such that if $H_0^i > M \ \forall i \in I$, for the first k_1 periods the number of times each player played each of her different actions can not differ by more than one. Thus, during those k_1 periods the players simultaneously switch at least k_0 times, and therefore play a^* with probability $1 - \eta$ at least.

☺[3]

Consider a particular Markov chain for $H_0^i > M \ \forall i \in I$, where M is provided by lemma 1.

Let $C \equiv \left\{\omega \in \Omega \middle| K\left(\omega_{k_1}, a^*\right) > 0\right\}$ be the set of histories where the *MI* profile was played at least once during the first k_1 stages (where k_1 is chosen by lemma 1). Lemma 1 yields $\mu(C) \ge 1 - \eta$. In the following we will show that $\mu\left(G \middle| C\right) = 1$, which will complete the proof of the theorem.

Let $H_{k_1}^i$ be player i's aspiration level after k_1 stages (note that it is still above u^*). We wish to find a number of periods $k_2 > k_1$ such that, on C, after k_2 periods, every player will be "practically satisficed" with $\overline{u}^i$. Specifically, we wish to guarantee that, if at any stage $k \ge k_2$, $U^i\left(\omega_k, a^{*i}\right) > U^i\left(\omega_k, b^i\right) + 1 \ \forall b^i \ne a^{*i}, \forall i \in I$, then for all $t \ge k$ every player i will choose a^{*i}, namely, $U^i\left(\omega_t, a^{*i}\right) > U^i\left(\omega_t, b^i\right) \ \forall b^i \ne a^{*i}, \forall t \ge k$.

Consider player i and $k \ge k_2$. On C, player i has experienced the maximal stage payoff $\overline{u}^i$, and therefore her aspiration level is $\overline{u}^i + \left(H_{k_1}^i - \overline{u}^i\right) \cdot \left(\alpha^i\right)^{k-k_1}$.

---

[3] See Aumann and Sorin [1989] for the precise definition of ☺.

Choose $k_2(i) > k_1 + \dfrac{\ln(1-\alpha^i) - \ln(H^i_{k_1} - \overline{u}^i)}{\ln(\alpha^i)}$ and let $k_2 \equiv \underset{i \in I}{Max}\ k_2(i)$. We will show now that it satisfies the above condition: let $k \geq k_2$, and assume that $U^i(\omega_k, a^{*i}) > U^i(\omega_k, b^i) + 1 \quad \forall b^i \neq a^{*i}$. Then, for all $t \geq k$, $U^i(\omega_t, a^{*i}) > U^i(\omega_t, b^i)$ $\forall b^i \neq a^{*i}$.

The proof is by induction. Consider $t \geq k \geq k_2$ and assume that the players chose a^* for $k \leq \tau \leq t-1$ at ω. Then we have:

$$U^i(\omega_t, a^{*i}) = U^i(\omega_k, a^{*i}) + \sum_{\tau=k}^{t-1}\left(\overline{u}^i - H^i_\tau\right) = U^i(\omega_k, a^{*i}) + \sum_{\tau=k}^{t-1}\left(\overline{u}^i - \left(H^i_{k_1} - \overline{u}^i\right)\cdot\left(\alpha^i\right)^{\tau-k_1} - \overline{u}^i\right) \geq$$

$$\geq U^i(\omega_k, a^{*i}) - \sum_{\tau=k_2(i)}^{\infty}\left(\left(H^i_{k_1} - \overline{u}^i\right)\cdot\left(\alpha^i\right)^{\tau-k_1}\right) = U^i(\omega_k, a^{*i}) - \left(\left(H^i_{k_1} - \overline{u}^i\right)\cdot\left(\alpha^i\right)^{k_2(i)-k_1}\right)\cdot\frac{1}{1-\alpha^i}$$

By the selection of $k_2(i)$:

$$U^i(\omega_t, a^{*i}) > U^i(\omega_k, a^{*i}) - 1 > U^i(\omega_t, b^i) = U^i(\omega_k, b^i) \quad \forall b^i \neq a^{*i}$$

when the last equality follows from the fact that, by the induction hypothesis, player *i* did not choose b^i.

Lemma 2: Given a state $s = (t, (a_1, \cdots, a_t))$ with $t \geq k_2$ (in the Markov chain) and $K(s, a^*) > 0$, there is a positive probability to reach a state r where for every player i, $U^i(r, a^{*i}) > U^i(r, b^i) + 1 \quad \forall b^i \neq a^{*i}$.

Proof: For a state *r*, let $PMI(r)$ be the set of players who, at r, would like to play their *MI* action. Assume first that $PMI(s)$ is a strict subset of I, the set of players. (The case $PMI(s) = I$ will be dealt with later.)

The idea of the proof is as follows: for every $i \notin PMI(s)$, there is a positive probability that *i* will not choose her *MI* action for a sufficiently long time, i.e., that every time she wants to switch to it, the noise would prevent her from playing her part of the *MI* profile. For every $j \in PMI(s)$ the player chooses a^{*j}. Since she will not get her *MI* payoff, this action will look less and less appealing to her until finally she would want to choose a different one. We graciously allow her to do so. Once she chose a different action, there is a positive probability that the noise will keep her from playing a^{*j} again. Thus, with a positive probability, after a finite time no player would like to play the *MI* action profile. Given that a^{*i} is not a maximizer of $U^i(\cdot,\cdot)$ for any i, there is a positive probability that none of them will play a^{*i}, but will play any other action long enough, until the condition is satisfied.

Formally, for $j \in PMI(s)$, let $\widetilde{u}^j$ be the second highest payoff player j can achieve in the stage game, and let $\delta^j \equiv \overline{u}^j - \widetilde{u}^j$. Since player j would like to play her MI action we know that $U^j\left(s, a^{*j}\right) \geq U^j\left(s, b^j\right) \quad \forall b^j \neq a^{*j}$ and we can define $\Delta^j \equiv U^j\left(s, a^{*j}\right) - U^j\left(s, \widetilde{a}^j\right)$ where $U^j\left(s, \widetilde{a}^j\right)$ is the second highest $U^j(s,\cdot)$ value. Thus in every subsequent period the value of $U^j\left(\cdot, a^{*j}\right)$ decreases by δ^j at least (recall that the aspiration level is always above $\overline{u}^j$).

After additional $k_3(j) = \left\lfloor \frac{\Delta^j}{\delta^j} \right\rfloor + 1$ periods, the MI action will cease to have the highest $U^j(\cdot,\cdot)$ value and player j would choose some other action. Therefore after $k_3 \equiv \underset{j \in PMI(s)}{Max}\ k_3(j)$ periods there is a positive probability that no player would like to choose to play the MI action. (Note that with probability greater than $\left(\prod_{i \in I}\left(1 - \frac{\varepsilon}{|A^i|}\right)\right)^{k_3}$ all the players that are not in $PMI(\cdot)$ will not choose their MI actions during these k_3 periods.)

Choose a sequence $B = \left(b_1, b_2, \cdots, b_p\right)$ of action profiles such that no player uses her MI action and all of the other actions of every player are represented in the sequence. Formally,

(i) $\quad b_v^i \neq a^{*i} \quad \forall v \leq p \ \forall i \in I$

(ii) $\quad \forall i \in I, \forall a^i \in A^i \setminus \left\{a^{*i}\right\} \ \exists v \leq p \ \ s.t. \ \ b_v^i = a^i$.

Without loss of generality, assume that b_1 is the profile the players are playing after the appropriate k_3 periods. Let $r^1 = \left(t + k_3, \left(\cdots, b_1\right)\right)$ be a state arrived at, from s, as required. I.e., $\forall i \in I, \exists b^i \ \ s.t. \ U^i\left(r^1, b^i\right) > U^i\left(r^1, a^{*i}\right)$. We argue that there is a positive probability that the play will follow a path:

$s = \left(t, \left(a_1, \cdots, a_t\right)\right) \rightarrow \cdots \rightarrow r^1 = \left(t + k_3, \left(a_1, \cdots, a_t, a_{t+1}, \cdots, a_{t+k_3} = b_1\right)\right) \rightarrow$

$r^2 = \left(t + k_3 + l_1, \left(a_1, \cdots, b_1, \cdots, b_1, b_2\right)\right) \rightarrow, \cdots, \rightarrow$

$r^p = \left(t + k_3 + l_1 + \cdots + l_{p-1}, \left(a_1, \cdots a_t, \cdots, b_1, \cdots, b_1, \cdots, b_{p-1}, \cdots, b_{p-1}, b_p\right)\right)$

$r^{p+1} = \left(t + k_3 + l_1 + \cdots + l_p, \left(a_1, \cdots a_t, \cdots, b_1, \cdots, b_1, \cdots, b_p, \cdots, b_p, a^*\right)\right) = r$

Such that for every $v \leq p$, at r^v, b_v^i has a low enough $U^i(\cdot,\cdot)$ value for every i i.e. $U^i\left(r^v, b^i\right) < U^i\left(r^v, a^{*i}\right) - 1$.

It takes at most $l_v(j) = \left\lceil \frac{U^j\left(r^v, b_v^j\right) - U^j\left(r^v, a^{*j}\right) - 1}{\delta^j} \right\rceil$ consecutive plays of b_v to lower b_v^j's cumulative utility to the desired level. Let p_v be the conditional probability to get to r^v given that the current state is r^{v-1}. To see that $p_v > 0$, note that $p_v \geq \left(\prod_{i \in I} \frac{\varepsilon}{|A^i|} \right)^{l_v+1}$ where $l_v \equiv \underset{j \in I}{Max}\ l_v(j)$. After going through the entire sequence of stages, with probability $\prod_v p_v > 0$, we are at state r in which we have $U^i\left(r, a^{*i}\right) > U^i\left(r, b^i\right) + 1 \quad \forall b^i \neq a^{*i} \ \forall i \in I$.

If $PMI(s) = I$ to begin with, we distinguish between two cases: if the players never wish to switch from a^*, we are done. Otherwise after a finite number of periods a player would like to choose some other action. In this case we find a path as above.

☺

The following straightforward lemma about Markov chains is essential for the proof:

Lemma 3: Let S be the state space of a countable Markov chain, $\{X_t\}_{t=1}^{\infty}$ and let $F \subseteq S$ be such that $\forall t > 1 \ P\left(X_t \in F \,|\, X_{t-1} \in F\right) = 1$ and $\forall j \notin F$ $P\left(\exists s > t;\ X_s \in F \,|\, X_t = j\right) > 0$. Then $P\left(\exists T, \forall t > T;\ X_t \in F\right) = 1$.

Proof : Consider the following Markov process $\{Y_t\}_{t=1}^{\infty}$ over the state space $\tilde{S} = \{f\} \cup (S / F)$: $Y_t = X_t$ *if* $X_t \notin F$, and $Y_t = f$ *if* $X_t \in F$.

This process satisfies

(1) $P_{i,j} = P\left(X_t = i \,|\, X_{t-1} = j\right) = P\left(Y_t = i \,|\, Y_{t-1} = j\right) = \tilde{P}_{i,j} \quad \forall i, j \in S / F$

(2) $\tilde{P}_{i,f} = \sum_{j \in F} P_{i,j} \quad \forall i \in S / F$

(3) $\tilde{P}_{f,f} = 1$ (which implies that $\tilde{P}_{f,i} = 0 \quad \forall i \neq f$)

Clearly f is the unique essential state in the new Markov process. Thus, by a standard theorem about Markov chains, the unique stationary distribution, π, of the chain satisfies $\pi(f) = 1$. (See for example Shirayayev [1984] chapter VIII).

Observe that $P(X_t = j) = P(Y_t = j) \quad \forall t; \forall j \in S / F$ and $P(Y_t = f) = P(X_t \in F) \quad \forall t$, from which the result follows.

☺

Let G_M be the set of states in the Markov chain such that the players play the *MI* profile at the last stage and for every player j, $U^j(\cdot, a^{*i})$ is sufficiently high such that she would like to choose to play her *MI* actions forever:

$$G_M \equiv \left\{ s = \left(t, \left(a_1, \cdots, a_t = a^*\right)\right) \;\middle|\; \begin{array}{l} K(s, a^*) > 0 \;,\; t > k_2 \\ U^i(s, a^{*i}) - U^i(s, b^i) > 1 \quad \forall i \in I; \forall b^i \neq a^{*i} \end{array} \right\}$$

Lemma 2 shows that from every state $s = \left(t, \left(a_1, \cdots, a_t\right)\right) \; t \geq k_2$ with $K(s, a^*) > 0$ outside G_M there is a positive probability to move into a state in G_M. For every $s = \left(t, \left(a_1, \cdots, a_t = a^*\right)\right) \in G_M$, with probability 1 the next state will be $s' = \left(t+1, \left(a_1, \cdots, a_t = a^*, a_{t+1} = a^*\right)\right) \in G_M$. Thus by Lemma 3, G_M is an absorbing set of states. Hence the players will play only the *MI* profile after a finite time. In other words we proved that $\mu(G \mid C) = 1$.

☺

4. Simulation

The proof of our main theorem hinges on the specific "trembling" assumption. Further, it might give the impression that convergence occurs only after unreasonably long time, and only for very large initial aspiration levels. However, it turns out that this is not the case. Computer simulations show that convergence to the *MI* outcome is both faster and more robust than the theoretical result would have us believe. We start with the famous Harsanyi and Selten's [1988] risk dominance example:

	L	R
T	**9,9**	**0,8**
B	**8,0**	**7,7**

The play of this game was simulated using initial aspiration levels of 50 for both players varying the noise level and the aspiration updating parameter. Table 1 shows that after the first 50 stages or so the players settled on playing only the *MI* outcome.

Table 1:

	Contingency table for (TL,TR,BL,BR):		
Stage	$\varepsilon = 0.05$ $\alpha = 0.95$	$\varepsilon = 0.1$ $\alpha = 0.95$	$\varepsilon = 0.1$ $\alpha = 0.9$
50	25, 2, 2, 21	23, 4, 3, 20	24, 4, 4, 18
100	60, 5, 5, 30	52, 7, 7, 34	74, 4, 4, 18
150	110, 5, 5, 30	102, 7, 7, 34	124, 4, 4, 18
200	160, 5, 5, 30	152, 7, 7, 34	174, 4, 4, 18

The entries in the table specify the number of times the players played each action profile. For example at stage 100 (where the noise level was 10% and their aspiration level updating parameter was $\alpha = 0.95$) they played TL 52 times, TR 7 times, BL 7 times, and BR 34 times.

Next a "uniform tremble," was used i.e., one that may occur at any stage regardless of the choice. Table 2 shows that as the noise level tends to zero the players choose the *MI* outcome with a frequency approaching unity. (Initial aspiration levels were 100 for both players.)

Table 2:

	Contingency table for (TL,TR,BL,BR) $\alpha = 0.95$:			
Stage	$\varepsilon = 0.05$	$\varepsilon = 0.01$	$\varepsilon = 0.005$	Without Noise
500	319, 30, 30, 121	467, 3, 4, 26	487, 2, 2, 9	436, 11, 11, 42
1,000	637, 61, 61, 241	897, 17, 17, 69	969, 5, 5, 21	936, 11, 11, 42
3,000	1949, 175, 175, 701	2645, 59, 59, 237	2885, 19, 19, 77	2936, 11, 11, 42
5,000	3271, 288, 288, 1153	4441, 93, 93, 373	4825, 29, 29, 117	4936, 11, 11, 42
7,500	4949, 425, 425, 1701	6701, 133, 133, 533	7253, 41, 41, 165	7436, 11, 11, 42
10,000	6606, 566, 566, 2262	8919, 180, 180, 721	9657, 57, 57, 229	9936, 11, 11, 42

In the next simulation the game was chosen randomly. One thousand two-player seven-by-seven games with a *MI* outcome of (10,10) were analyzed using the "inertia" and uniform trembles. The initial aspiration levels was 300 and the updating parameter was $\alpha = 0.95$ for both players. Table 3 shows the average frequency with which the *MI* outcome was played after several stages and its standard deviation.

Table 3:

	Frequency (Standard Deviation) for:			
Stage	"Inertia" $\varepsilon = 0.05$	"Uniform" $\varepsilon = 0.005$	"Uniform" $\varepsilon = 0.0005$	Without Noise
1,000	.78068 (.07041)	.55177 (.13845)	.75013 (.13693)	.88498 (.06969)
2,000	.89034 (.03522)	.61092 (.11785)	.84808 (.09505)	.94249 (.03484)
3,000	.92689 (.02339)	.63007 (.10299)	.88238 (.06784)	.96166 (.02323)
4,000	.94517 (.01409)	.64002 (.09802)	.90119 (.05461)	.97124 (.01742)
5,000	.95614 (.01200)	.64634 (.09498)	.91145 (.04749)	.97700 (.01394)
10,000	.98011 (.00838)	.65047 (.08935)	.91145 (.03191)	.98850 (.00697)

Next, some games with more players were studied. The first one is the "stag hunt game." In this game a group of I hunters simultaneously choose to hunt a stag or hares. It requires all of them to hunt a stag, but each one can hunt a hare by herself. Thus the payoff to a hunter that goes for the hare is 1 regardless of the other hunters' action. If all hunt a stag, each one gets a payoff of 2. But if only some try to hunt a stag they fail and get 0. The game has two pure strategy Nash Equilibria: all hunting hares or all hunting the stag. The game was simulated by using 7 players with initial aspiration levels of 500 and updating parameter $\alpha = 0.95$ for each player. Table 4 demonstrates that play will converge to the "all-stag" equilibrium.

Table 4:

	Count of the all **stag** (all **hare**) equilibrium:			
Stage	"Inertia" $\varepsilon = 0.05$	"Uniform" $\varepsilon = 0.05$	"Uniform" $\varepsilon = 0.005$	Without Noise
50	16 (15)	19 (10)	37 (6)	14 (15)
100	39 (34)	50 (19)	47 (12)	46 (38)
500	429 (40)	274 (82)	441 (49)	424 (47)
1000	929 (40)	534 (167)	930 (49)	924 (47)
5000	4929 (40)	2756 (785)	4669 (130)	4924 (47)
10000	9929 (40)	5418 (1589)	9339 (242)	9924 (47)

Our final example is a variation of the "stag hunt" game. The game involves 5 players, each having three strategies, Up, Middle, and Down. The payoff to a player j is : $u^j(Up) \equiv 6 \times \left(\#\left\{i \middle| i \ \ chose \ \ Up\right\}\right)$, $u^j(Middle) \equiv 4 \times \left(\#\left\{i \middle| i \ \ chose \ \ Middle\right\}\right)$ and $u^j(Down) \equiv 2 \times \left(\#\left\{i \middle| i \ \ chose \ \ Down\right\}\right)$.

This game has three pure strategy Nash Equilibria where all the players choose the same action and the *MI* profile is for all to play Up. Table 5 shows the number of times the players played each of the Nash Equilibria.

Table 5:

	Count of the all **Up, Middle, Down** equilibrium plays:				
Stage	"Inertia" $\varepsilon = 0.05$	"Uniform" $\varepsilon = 0.05$	"Uniform" $\varepsilon = 0.005$	"Uniform" $\varepsilon = 0.001$	Without Noise
100	32, 11, 8	41, 5, 4	44, 8, 8	56, 9, 8	60, 11, 12
500	424, 12, 9	211, 10, 7	432, 11, 12	445, 11, 12	448, 11, 12
1000	924, 12, 9	241, 11, 8	926, 11, 12	944, 11, 12	948, 11, 12
2000	1924, 12, 9	312, 14, 9	1905, 11, 12	1942, 11, 12	1948, 11, 12
5000	4924, 12, 9	611, 21, 10	4487, 12, 12	4931, 11, 12	4948, 11, 12
10000	9924, 12, 9	1296, 27, 12	7765, 16, 15	9921, 11, 12	9948, 11, 12
50000	49924, 12, 9	6237, 44, 24	41887, 19, 20	48121, 11, 12	49948, 11, 12

5. Concluding Remarks

1. The main theorem has a few obvious extensions. For instance, the aspiration level adjustment rule need not be the weighted average one. Similarly, the noise, ε, may depend on the player and on the stage, as long as it is bounded away from zero often enough.

2. The simulation results give rise to the conjecture that a uniform noise, which converges to zero, would lead to a limit frequency of one for the *MI* action profile, in the spirit of Kandori, Mailath and Rob [1993] or Young [1993]. (It is, however, easy to construct examples in which convergence fails in the absence of noise.) By contrast, under our assumption convergence is guaranteed for any $\varepsilon > 0$.

References

Abreu, D. and A. Rubinstein (1988): "The structure of Nash Equilibrium in Repeated Games with finite Automata," *Econometrica*, **56**, 1259-1281.

Anderlini, L. and A. Ianni (1993): "Path Dependence and Learning From Neighbors," *Mimeo*, Cambridge University.

Anderlini, L. and H. Sabourian (1990): "Cooperation and Effective Computability," *Mimeo*, Cambridge University.

Aumann, R.J. (1990): "Communication Need Not Lead to Nash Equilibrium," *Mimeo*, Hebrew University of Jerusalem.

Aumann, R.J. and S. Sorin (1986): "Cooperation and Bounded Recall," *Games and Economic Behavior*, **1**, 5-39.

Balkenborg, D. (1993): "Strictness, Evolutionary stability and Repeated Games with Common Interests," *Mimeo*, University of Pennsylvania.

Bendor J., D. Mookherjee and D. Ray (1994): "Aspirations, Adaptive Learning and Cooperation in Repeated Games," *Mimeo*, Boston University.

Binmore, K. and L. Samuelson (1992): "Evolutionary Stability in Repeated Games Played by Finite Automata," *Journal of Economic Theory*, **57**, 278-305.

Binmore, K. and L. Samuelson (1992): "Evolutionary Stability in Repeated Games Played by Finite Automata," *Journal of Economic Theory*, **57**, 278-305.

Gilboa, I. and D. Schmeidler (1992): "Case-Based Decision Theory," forthcoming, *Quarterly Journal of Economics*.

Gilboa, I. and D. Schmeidler (1993): "Case-Based Optimization," forthcoming, *Games and Economic Behavior* .

Kahneman, D. and A. Tversky (1984): "Choices, Values and Frames," *American Psychologist*, **39**, 341-350 .

Kandori, M., Mailath G. and R. Rob (1993): "Learning, Mutations and Long Run Equilibria in Repeated Games," *Econometrica*, **61**, 27-56.

Kim, Y.G. and J. Sobel (1991): "An Evolutionary Approach to Pre-play Communication," *Mimeo*.

March, J.G. and H.A. Simon (1958): "Organizations," New York, John Wiley and Sons.

Matsui, A. (1991): "Cheap Talk and Cooperation in Society," *Journal of Economic Theory*, **54**, 245-258.

Osbourne, M.J. (1990): "Signaling, Forward Induction, and Stability in Finitely Repeated Games," *Journal of Economic Theory*, **50**, 22-36.

Neyman, A. (1985): "Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoner's Dilemma," *Economic Letters*, **19**, 227-229.

Rubinstein, A. (1986): "Finite Automata Play the Repeated Prisoner's Dilemma," *Journal of Economic Theory*, **39**, 83-96.

Shirayayev, A.N. (1984): "Graduate Texts in Mathematics: Probability," New York, Springer Verlang.

Shoham, Y. and M. Tennenholtz (1993): "Co-Learning and the Evolution of Social Activity," *Mimeo*, Stanford University.

Simon, H.A. (1955): "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, **69,** 99-118.

Schlag, K. (1993): "Cheap Talk and Evolutionary Dynamics," *Mimeo*, University of Bonn.

Wärneryd, K. (1990): "Cheap Talk, Coordination and Evolutionary Stability," *Mimeo*, Stockholm School of Economics.

Young P.H. (1993): "The Evolution of Conventions," *Econometrica*, **61**, 57-84.