GAMES AND ECONOMIC BEHAVIOR 7, 390–405 (1994)

Finitely Many Players with Bounded Recall in Infinitely Repeated Games

EHUD LEHRER[1]

Department of Managerial Economics and Decision Sciences, Kellogg Graduate School of Management, Northwestern University, Evanston, Illinois 60208; and School of Mathematics, Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel 69978

Received November 4, 1988

We study the set of limit points of equilibrium payoffs in n-player repeated games, with bounded recall, when the memory capacities of all the players grow to infinity. Two main issues are explored: (i) whether differential information enables players to play correlatively, and (ii) the extent to which boundedly rational players can learn others' behavior patterns and conceal their own. *Journal of Economic Literature* Classification Number: 026.

INTRODUCTION

The notion of bounded rationality in the context of repeated games was introduced first by Aumann (Aumann, 1981). In his survey, Aumann mentions two ways to model a player with bounded rationality: with finite automata and with bounded recall strategies. Neyman (1985) and Rubinstein (1986) were the first to see the impact of restricting the set of strategies available to a player to those strategies that can be implemented by finite automata.

In this paper we address repeated games played by players with bounded recall. Two types of bounded recall in repeated games are found in the literature. The first one, adopted by the author in Lehrer (1988) and here, is that in which a player, when taking an action, can rely on his opponents' previous actions as well as on his own. The second type (see Aumann and Sorin, 1989) restricts the player to relying solely on his opponents'

[1] The author acknowledges the valuable comments of the anonymous referees of *Games and Economic Behavior*.

0899-8256/94 $6.00

actions. One of the reasons for adopting the second type, proposed by Aumann and Sorin, was the avoidance of the possibility of "trigger" strategies.

This article supports that reason in showing that, even though the recall (of the first type) of a player is bounded, he can play some kind of "trigger" strategy. A player executes this strategy by signaling to himself with his own actions and thereby keeping the deviation alive in his memory. This grim strategy, for instance, cannot be carried out if a player can rely solely on his opponents' actions because the deviation (of the punished player) will disappear from the punisher's memory after a while. However, in a case where the players are capable of memorizing their own actions, the punisher can remind himself by a certain behavior pattern, which is identified with the deviation, that a deviation has occurred and that he should follow the punishing strategy.

We investigate the asymptotic behavior of the Nash equilibrium payoff set, when the memory capacities of all the players grow to infinity. The same problem for two-player games was handled in the context of finite automata by Ben-Porath (1989) and in the context of bounded recall by Lehrer (1988).

These two papers show, in their respective contexts, that as long as the "strong" (the one with the higher computational capacity) player's memory does not improve rapidly (i.e., exponentially) compared to the "weak" player's memory, the latter can conceal his particular behavior pattern. In other words, the "strong" player knows the distribution according to which the "weak" player's strategy is chosen, but he fails to reveal which particular pure strategy was chosen.

This issue of concealing one's own chosen strategy is also addressed here, in the context of finitely many players. It is shown that in order to ensure a certain security level, a player should hide his realized strategy. Otherwise, other players, after a while, learn enough about him and his weak spots to push his payoff below the security level. In order to prove that what is meant to be hidden is indeed concealed, we employ an information-theory technique—the memory capacity of a player is measured by its entropy.

The second main new idea contained in this paper is the correlation between players, based on their differential information. Different players have different memory capacities. Thus, during the game, players base their actions on different information. This results in correlated actions played at every stage. We study the extent to which players can correlate their moves by utilizing the histories as a correlation device. It turns out that, under some assumptions on the speed of growth of memory capacities, players can fully coordinate their strategies and act as if there were one player. This is done when one player, say, player i, is found to be deviating and all other players punish him. If players are ordered and labeled according to their memory capacity, player i is actually punished by i players: $i - 1$ players with shorter memory and all "stronger" players acting as one (by using the histories properly). The phenomenon of correlation by utilizing the histories was treated, in different contexts, by Gilboa and Schmeidler (1989), by Lehrer (1991), and by Kalai and Lehrer (1990).

As a motivation of the topic, imagine a situation where, after playing a long term game, the capabilities of all the participants improve (possibly as a result of an evolutionary development). The question of the asymptotic behavior of the equilibrium payoffs is the question of the worth of developing quickly. Namely, what is the value, in terms of equilibrium payoffs, of a larger information-handling capability? Our main result states that a greater recall affects only the individually rational level. In other words, longer memory players (the "strong" ones) can only ensure themselves a greater *minimal* payoff. On the other hand, the lower security level of the "weak" players ensures, in some cases, more opportunities for others.

The paper is organized as follows. Section 2 presents the model and the main results. In Section 3, the repeated game played with bounded recall strategies is described in detail. Sections 4 and 5 are devoted to proofs and in Section 6 we consider the issue of perfectness. We assert that the same result holds also for perfect equilibrium.

2. Notations and Main Theorems

The game G_k consists of

(i) the set of players $M = \{1, \ldots, n\}$;
(ii) n finite sets of actions $\Sigma_1, \ldots, \Sigma_n$

(Denote $\Sigma = \times_{j=1}^{n} \Sigma_j$);

(iii) n payoff functions $h_1, \ldots, h_n$; $h_i: \Sigma \to \mathbb{R}$, $i \in M$, where $\mathbb{R}$ is the set of the real numbers (without loss of generality, we can assume that $h_i \geq 0$); and
(iv) n integers $l_1(k), \ldots, l_n(k)$.

$l_i(k)$ is interpreted as the memory length of player i in the kth game. In the sequel, $l_i(k)$ will be greater than $l_j(k)$ whenever $i > j$.

The set of $l_i = l_i(k)$ bounded recall strategies of player i, denoted by $S_i^{l_i}$, is $\{(x, \phi) \mid x \in \Sigma^{l_i}, \phi\colon \Sigma^{l_i} \to \Sigma_i\}$. For any tuple $\sigma = (\sigma_1, \ldots, \sigma_n) \in \times_{j=1}^{n} S_j^{l_j}$ and $t \in \mathbb{N}$, define $a_i^t(\sigma)$ = the payoff of player i at stage t, and $H_i(\sigma)$ = the Cesaro limit of $a_i^t(\sigma)$.

For a set A, $\Delta(A)$ denotes the set of all the distributions over A.

The range of the payoff function $H_i(\bullet)$ can be extended naturally to $\times_{j=1}^{n}\Delta(S_j^{l_j})$. Let $H = (H_1, \ldots, H_n)$. The n-player game $G_k = (\Delta(S_1^{l_1}), \ldots, \Delta(S_n^{l_n}); H)$ has, by the Nash theorem, a nonempty set of Nash equilibrium payoffs, denoted by N_k.

In the following notation, p, q, and r are random variables taking actions (or action combinations) as values.

Notation 2.1. (1) Let $i \in M$. Denote $\mathbf{d}_i = \text{Min}_q \text{ Max}_p \text{ Min}_r E_{p,q,r}(h_i)$, where E is the expectation functional, the distributions of q, p, and r are in $\times_{j<i}\Delta(\Sigma_j)$, $\Delta(\Sigma_i)$, and $\Delta(\times_{j>i}\Sigma_j)$, respectively, and, moreover, p is independent of q and r, and r is q-measurable.

The meaning of $\mathbf{d}_i$ is the following. The players with memories shorter than i play q where the actions are not correlated. Player i knows the distribution q but not its particular realization. However, players with memories longer than i ($j > i$) play correlatively and they may know the realization of the action combination q. Thus, r may be q-measurable. These players ($j : j > i$) know the distribution of p but not its realization. To sum up, $\mathbf{d}_i$ is a security level of player i with differential information: all the players "stronger" than i know the particular pure strategy played by the "weaker" players. However, player i himself does not know it.

$\overline{d}_i$ is defined like $\mathbf{d}_i$ with the further restriction that q and r are independent. Thus, the players with memories longer than i do not know the realization of q. In the following it is shown that player i can guarantee himself $\mathbf{d}_i$ and other players can ensure that player i does not get more than $\overline{d}_i$ when he is punished.

(2) Define

$$\mathbf{F} = \{x \in \mathbb{R}^n \mid x \text{ is feasible and } x_i \geq \mathbf{d}_i,\ i \in M\},$$
$$\overline{F} = \{x \in \mathbb{R}^n \mid \text{ is feasible and } x_i \geq \overline{d}_i,\ i \in M\},$$

and

$$\overline{F}^s \text{ is defined like } \overline{F} \text{ with strict inequalities.}$$

$\overline{F}$ and $\mathbf{F}$ are the feasible payoffs which are greater than the respective security level. Note that both are greater than the set of the folk theorem (see Aumann, 1981) in which the individually rational level is the minmax one.

Notation 2.2. Let $\{B_k\}$ be a sequence of sets in $\mathbb{R}^n$.

(1) $\liminf_k B_k = \{x \in \mathbb{R}^n \mid$ there is a sequence $x_k \in B_k$ which converges to $x\}$;

(2) $\limsup_k B_k = \{x \in \mathbb{R}^n \mid x$ is an accumulation point of a sequence $x_k \in B_k\}$.

THEOREM 1. *If* $\lim_{k\to\infty} [\log l_n(k)/l_1(k)] = 0$, *then* $\limsup_{k\to\infty} N_k \subseteq \mathbf{F}$.

THEOREM 2. *If* $\lim_{k\to\infty} [l_{i+1}(k)/l_i(k)] = \infty$ *for every* $1 \leq i \leq n-1$, *then* $\overline{F} \subseteq \liminf_{k\to\infty} N_k$, *provided that* $\overline{F}^s \neq \emptyset$.

3. A DETAILED DESCRIPTION OF G_k

DEFINITION 3.1. Let A be a set.

(1) x is an A-word if x is a finite string of symbols of A; i.e., $x \in A^l$ for some integer l.
(2) If x is an A-word, $|x|$ denotes the length of x.
(3) x is an $(m - A)$-word if x is an A-word and $|x| = m$.
(4) Let α, β be integers such that $1 \geq \alpha \geq \beta \geq |x|$ and $x = (x_1, \ldots, x_m)$, then $x(\alpha, \beta) = (x_\alpha, \ldots, x_{\alpha+\beta-1})$.

Let $\sigma_i = (x_i, \phi_i) \in S_i^{l_i}$ be a pure bounded recall strategy of player i, who has l_i-bounded recall. At the first stage player i plays $\phi_i(x_i)$. Denoting $z_1 = (\phi_1(x_1), \ldots, \phi_n(x_n))$, player i gets the payoff $a_i^1 = h_i(z_1)$ at that stage. At the second stage player i plays $\phi_i(x_i(2, l_i), z_1)$. Denoting $z_2 = (\phi_1(x_1(2, l_1), z_1), \ldots, \phi_n(x_n(2, l_n), z_1))$, player i gets the payoff $a_i^2 = h_i(z_2)$ at that stage. At the third stage player i plays $\phi_i(x_i(3, l_i), z_1, z_2)$, and so on.

For every t and $\sigma = (\sigma_1, \ldots, \sigma_n)$, $a_i^t(\sigma)$ denotes the payoff of player i at stage t, when σ is played, and by $H_i(\sigma)$, the Cesaro limit of $a_i^t(\sigma)$: $H_i(\sigma) = \lim_T (1/T)\Sigma_{t=1}^T a_i^t(\sigma)$. If σ is a tuple of mixed strategies (probability distributions over $S_i^{l_i}$, $i \in M$) then $a_i^t(\sigma)$ denotes the expectation of player i's payoff at stage t, and $H_i^t(\sigma)$ denotes its Cesaro limit.

We know that every n-player game (one-shot), where player i has a finite set of actions, has a nonempty set of Nash equilibria in mixed strategies. The game defined by the sets $S_i^{l_i}(k)$ and by the payoff functions $H_i(\bullet)$ is finite. Therefore, it has a nonempty set, N_k, of Nash equilibrium payoffs.

4. THE PROOF OF THEOREM 1

We show that $\limsup_k N_k \subseteq \mathbf{F}$. For this purpose we define a strategy $\overline{\sigma}_i^k$ of player i in G_k which ensures him at least the payoff $\mathbf{d}_i - \delta$ versus any tuple $(\sigma_1^k, \ldots, \sigma_{i-1}^k, \sigma_{i+1}^k, \ldots, \sigma_n^k) \in \times_{j\neq i} S_j^{l_j(k)}$, for any $\delta > 0$ and sufficiently large k. The strategy $\overline{\sigma}_i^k$ is defined in such a way that the probabilistic propositions of Lehrer (1988), quoted below, can be applied.

Fix an integer k. Player i knows at each stage t the memories of all players $j, j < i$. Thus, if $u \in \Sigma^{l_i}$ is the memory of player i, then he knows the expected strategy of each player $j < i$, denoted by $q_j(u)$. Let $q(u) = (q_1(u), \ldots, q_{i-1}(u))$. Denote

$$V_i(q(u)) = \text{Max}_p \, \text{Min}_r \, E_{p,q(u),r}(h_i),$$

where the minimum is taken over all r with distribution in $\Delta(\times_{j>i}\Sigma_j)$ and the maximum is taken over all p independent of $q(u)$ and with distribution in $\Delta(\Sigma_i)$. Denote a mixed strategy in $\Delta(\Sigma_i)$ which ensures $V_i(q(u))$ (when $q_j(u)$ is played by $j < i$) by $p(u)$.

Let $p^\eta(u)$ be a strategy of player i which assigns to every action at least probability η and which is the closest strategy of this kind to $p(u)$. Precisely, $p^\eta(u)$ is the closest point (with respect to $\|\bullet\|_\infty$) to $p(u)$ in the set $\{p \mid p$ is a mixed strategy of player i, and $p_s \geq \eta$ for every $s \in \Sigma_i\}$.

We now define for every $\omega \in \Omega$ a pure strategy $\overline{\sigma}_i^k(\omega)$ by using $\eta > 0$, which is specified later. The initial memory of player i in $\overline{\sigma}_i^k(\omega)$, $\overline{x}_i^k$, is arbitrary, and for all $u \in \Sigma^{l_i}$ define $\overline{\phi}_i^k(\omega)(u) = Z_u(\omega)$, where Z_u is a random variable, defined on Ω, and ranged to Σ_i, which has the same distribution as $p^\eta(u)$. Furthermore, all Z_u, $u \in \Sigma^{l_i}$, are mutually independent. Let $\overline{\sigma}_i^k = (\overline{x}_i^k, \overline{\phi}_i^k)$, and denote $\overline{\sigma}^k = (\overline{\sigma}_1^k, \ldots, \overline{\sigma}_i^k, \ldots, \overline{\sigma}_n^k)$. Note that σ_i^k is a random variable, defined on Ω and ranged to $S_i^{l_i}$—in other words, a mixed strategy of player i in G_k. In order to show that $H_i(\overline{\sigma}^k) > \mathbf{d}_i - \delta$ for k sufficiently large, we need a few lemmata.

Notation 4.1. Let $x = (x_1, \ldots, x_n)$ be an $(n - A)$-word, and let $l < n$. x has an *l-cycle* if there are $1 \leq g < j \leq n$ s.t. $(x_g, \ldots, x_{g+l-1}) = (x_j, \ldots, x_{j+l-1})$, where an integer $n' > n$ is identified with $n'(\text{mod } n)$.

Denote $s(k) = l_i(k)l_n(k)$.

PROPOSITION 4.2. *Let $X_1, X_2, \ldots X_{s(k)}$ be mutually independent Bernoulli random variables, where* $\text{prob}(X_j = 1) = 1 - \text{prob}(X_j = 0) = p_j$. *Assume furthermore that there is an $\eta > 0$ s.t. $\eta \leq p_j \leq 1 - \eta$ for all k and $1 \leq j \leq s(k)$. Denote $c_i(k) = \text{prob}\{X_1, \ldots, X_s(k)$ has an $l_i(k)$-cycle$\}$. Then*

$$\lim_k s(k) \cdot c_i(k) = 0.$$

This proposition is an extension of Proposition 1 in Lehrer (1988), which we quote here.

PROPOSITION 4.3. *Let $Y_1, Y_2, \ldots$ be a sequence of identically distributed mutually independent nontrivial Bernoulli random variables. Then*

$$\lim_k f(k) \cdot \text{prob}\{Y_1, \ldots, Y_{f(k)} \textit{ has an } l_i(k) \textit{ cycle}\} = 0,$$

provided that $\log f(k)/l_i(k)$ *tends to zero as k goes to infinity.*

In the sequel we make use only of the fact that $c_i(k)$ tends to zero. The proof of the weaker version is not easier and therefore the strong version is provided.

Now we can turn to the proof of Proposition 4.2.

Proof of Proposition 4.2. We show that if $Y_1, Y_2, \ldots$ are i.i.d. Bernoulli random variables with $\text{prob}(Y_1 = 1) = \eta$ then $c_i(k) \leq e_i(k) = s^2(k) \cdot \text{prob}\{Y_1, \ldots, Y_{s(k)}$ has an $l_i(k)$-cycle$\}$. Since $\log s(k)/l_i(k) \to 0$, and since $\text{prob}\{Y_1, \ldots, Y_{s(k)}$ has an $l_i(k)$-cycle$\} \leq \text{prob}\{Y_1, \ldots, Y_{s^2(k)}$ has an $l_i(k)$-cycle$\}$, Proposition 4.3, applied for $f(k) = s^2(k)$, provides a proof for Proposition 4.2.

Define the continuous function β_b on $\Delta^b \times \Delta^b$ (Δ^b is the unit simplex of $\mathbb{R}^b$) as

$$\beta_b(x_1, \ldots, x_b, y_1, \ldots, y_b) = \Sigma\Pi x_j^{\varepsilon_j}(1 - x_j)^{1-\varepsilon_j} y_j^{\varepsilon_j}(1 - y_j)^{1-\varepsilon_j},$$

where the summation is over all the vectors $(\varepsilon_1, \ldots, \varepsilon_b) \in \{0, 1\}^b$. Let $B_{g,j} = \{Y_g = Y_j, \ldots, Y_{g+l_i-1} = Y_{j+l_i-1}\}$. Now let g and j be two integers s.t. $g + l_i - 1 \leq j \leq s(k) - l_i + 1$.

The probability of the event $D_{g,j} = \{X_g = X_j, \ldots, X_{g+l_i-1} = X_{j+l_i-1}\}$ is a union of pairwise disjoint events,

$$\cup\{X_g = X_j = \varepsilon_g, X_{g+1} = X_{j+1} = \varepsilon_{g+1}, \ldots, X_{g+l_i-1} = X_{j+l_i-1} = \varepsilon_{g+l_i-1}\},$$

where the union is taken over all $(\varepsilon_g, \ldots, \varepsilon_{g+l_i-1}) \in \{0, 1\}^{l_i}$. Because of the independence of the random variables, the probability of this union is (denoting $z = j - g$)

$$\Sigma\Pi_{s=g}^{g+l_i-1} p_s^{\varepsilon_s}(1 - p_s)^{1-\varepsilon_s} p_{s+z}^{\varepsilon_s}(1 - p_{s+z})^{1-\varepsilon_s}.$$

Denote $v = (p_g, \ldots, p_{g+l_i-1}, p_j, \ldots, p_{j+l_i-1})$. Note that $\text{prob}(D_{g,j}) = \beta_b(v)$ and $\text{prob}(B_{g,j}) = \beta_b(\eta, \ldots, \eta)$. Since $\beta_b(v) \leq \beta_b(\eta, \ldots, \eta)$ (see Appendix), $\text{prob}(D_{g,j}) \leq \text{prob}(B_{g,j})$ when g, j are as above. As for the other g and j, the event $D_{g,j}$ is defined with overlapping cycles (the string starting at g and the one starting at j overlap). This means that there are some strings (shorter than l_i) that appear more than twice. By defining a multilinear function, similar to β_b above, which takes account of any repetition of symbols in the string $X_g, X_{g+1}, \ldots, X_{g+l_i-1}$, one can show that

$$\text{prob}(D_{g,j}) \leq \text{prob}(B_{g,j}) \qquad \text{for every } g \neq j.$$

Now $c_i(k) \leq \Sigma_{g \neq j}\,\text{prob}(D_{g,j})$ and $\Sigma_{g \neq j}\,\text{prob}(B_{g,j}) \leq e_i(k)$. The second inequality holds because there are less than $s^2(k)$ different selections of g and j $(g \neq j)$. Thus, one obtains

$$c_i(k) \leq e_i(k).$$

This concludes the proof of Proposition 4.2. ■

Some notation from information theory must be introduced before proceeding to the proof. Let X be a finite random variable (f.r.v.) with range $\{x_1, \ldots, x_n\}$. Define the *entropy* of X to be $\mathbf{H}(X) = -\Sigma_i p(x_i) \log p(x_i)$, where $p(x_i) = \text{prob}(X = x_i)$. If Y is another f.r.v. that ranged to $\{y_1, \ldots, y_m\}$, then $\mathbf{H}(X, Y) = -\Sigma_{i,j} p(x_i, y_j) \log p(x_i, y_j)$, where $p(x_i, y_j) = \text{prob}(X = x_i$ and $Y = y_j)$.

Denote by $|X|$ the number of atoms of X. $\mathbf{H}(X)$ is always less or equal to $\log|X|$.

DEFINITION 4.4. A f.r.v. Y is said to be ε-independent of a f.r.v. X (write $Y \overset{\varepsilon}{\perp} X$) if

$$\Sigma_j |\text{prob}(Y = y_j | X = x_i) - \text{prob}(Y = y_j)| < \varepsilon$$

for all i, except for a set of atoms the union of which has a measure less than ε.

Define

$$I(Y|X) = \mathbf{H}(X) + \mathbf{H}(Y) - \mathbf{H}(X, Y).$$

Theorem 4.6 of McEliece (1977) states that $I(Y|X) \geq 0$ with inequality if and only if Y and X are independent.

$I(Y|X)$ is interpreted as the information about Y that is contained in X.

PROPOSITION 4.5 (Smorodinsky, 1971). *Given $\varepsilon > 0$ there exists $\delta(\varepsilon) > 0$ s.t. for any f.r.v. x and Y, $I(Y|X) < \delta$ implies $Y \overset{\varepsilon}{\perp} X$.*

The final statement we need is that if Z is Y-measurable, then $I(Y|X) \geq I(Z|X)$ (McEliece, 1977, p. 26).

Now divide $\mathbb{N}$, the set of stages, into blocks of length $s(k)$ each. Fix a block B. Our objective is to compute the expected average payoff of player i in B. Denote by Y the memory of the player with the longest memory excluding i. Denote by Q_t and X_t the joint action of all players $j \neq i$ and of player i at stage t of B, respectively. Without loss of generality all these random variables are defined on the same sample space. We show that, with high probability, $h_i\ (Q_t, X_t)$ cannot be far below $\mathbf{d}_i$ in a relatively large set of stages.

Recall that in our model a player chooses at the beginning of the game a *pure* bounded recall strategy. He is not allowed to randomize (relying on his memory) at any stage. The subtle point is, therefore, to show that the player with the largest memory can, only with a small probability, unveil the pure strategy that was chosen by player i at the game's beginning. Player n can learn about player i's strategy only if, in his memory, there is an $l_i(k)$-cycle of length $l_n(k)$.

In what follows we show that all players $j \neq i$ can only with a small probability learn the identity of the particular pure strategy chosen by player i. The argument is based on the fact that the random variables Z_u, by which $\overline{\phi}_i^k$ was defined, are mutually independent.

Note that the independence of Z_u does not imply that the variables X_t are independent. This is so because $X_1, \ldots, X_{t-1}$ and $Q_1, \ldots, Q_{t-1}$ determine the distribution of X_t. Fix the distribution of X_t, say to q, and consider only the event, say A, on which the distribution of X_t is q. Contrary to the lack of independence on the whole space, on A, $Q_1, \ldots, Q_{t-1}$ and $X_1, \ldots, X_{t-1}$ are independent of X_t provided that no cycles exist. The event A, just described, is a subset of the sample space which is the range of *all* the random variables under discussion and is defined by $Q_1, \ldots, Q_{t-1}$ and $X_1, \ldots, X_{t-1}$. If we fix not only the distribution of one X_t but the distribution of the all X_t's, we get that on a small event (defined by one distribution per one stage) $X_1, \ldots, X_{t-1}$ are independent of X_t for every t, provided that there are no $l_i(k)$-cycles. In other words, by fixing the distribution of X_t, all its predecessors become almost independent of it.

Precisely, for any sequence of distributions $\overline{q} = \{q_t\}_{t\in B} \subseteq \{p^{\eta}(u)\}_{u\in\Sigma^{l_i(k)}}$ define the event

$$\begin{aligned} A(\overline{q}) &= \{\text{at time } t \in B \text{ player } i \text{ played according to the distribution } q_t\} \\ &= \{\omega \mid p^{\eta}(u_t(\omega)) = q_t\}, \end{aligned}$$

where

$$u_t(\omega) = ((X_{t-l_i(k)}(\omega), Q_{t-l_i(k)}(\omega)), \ldots, (X_{t-1}(\omega), Q_{t-1}(\omega))) \qquad \text{for all } t \in B\}$$

Once again, given $A(\overline{q})$, the distribution according to which player i is playing at any stage is fixed. From now on the event $A(\overline{q})$ is given for a fixed $\overline{q}$. We assume that $\text{prob}(A(\overline{q})) > 0$; namely, that the sequence $\overline{q}$ of distributions is plausible. Since all Z_u are independent one gets that on

$A(\bar{q}), X_1, \ldots, X_{t-1}$ are independent of X_t, provided that $u_t(\omega) \neq u_s(\omega)$ for all $s \in B$ and $s < t$.

Now consider the following perturbation of X_t. For every $\omega \in A(\bar{q})$, if $u_t(\omega) \neq u_s(\omega)$ for all $s \in B$, $s < t$; namely, there was no $l_i(k)$-cycle up to time t, define $\bar{X}_t(\omega) = X_t(\omega)$. However, on the event where there were $l_i(k)$-cycles up to time t (i.e., there is $s \in B$ $x < t$ s.t. $u_t(\omega) = u_s(\omega)$), define $\bar{X}_t$ to be an independent variable of $X_1, \ldots, X_{t-1}$ having the distribution of X_t. By so doing we defined a sequence $(\bar{X}_t)_{t \in B}$ of independent variables with the property that $X_t = \bar{X}_t$ on an event where the sequence (X_t) has no $l_i(k)$-cycles. This event has, by Proposition 4.2, a large probability (within $A(\bar{q})$).

We show that if player i could play according to $\bar{X}_t$, rather than according to X_t, he could ensure himself a close payoff to $\mathbf{d}_i$. Since X_t is close to $\bar{X}_t$ this completes the proof.

The following computation demonstrates the fact that, by memorizing Y, player j ($j \neq i$) is capable of identifying only a small portion of the $\bar{X}_t$'s.

$$
\begin{aligned}
&(1/|B|)\Sigma_{t \in B} I(Q_t \mid \bar{X}_t) \\
&\quad \leq (1/|B|)\Sigma_{t \in B} I(Y, Q_t \mid \bar{X}_t) \qquad (Q_t \text{ is } (Y, \bar{X}_1, \ldots, \bar{X}_{t-1}) \text{ measurable}) \\
&\quad \leq (1/|B|)\Sigma_{t \in B} I(Y, \bar{X}_1, \ldots, \bar{X}_{t-1} \mid \bar{X}_t) \qquad \text{by a direct computation)} \\
&\quad \leq (1/|B|)\Sigma_{t \in B} \mathbf{H}(\bar{X}_t) + \mathbf{H}(Y)/|B| - \mathbf{H}(Y, \bar{X}_1, \ldots, \bar{X}_{|B|})/|B| \\
&\quad \leq (1/|B|)\Sigma_{t \in B} \mathbf{H}(\bar{X}_t) - \mathbf{H}(\bar{X}_1, \ldots, \bar{X}_{|B|})/|B| + \mathbf{H}(Y)/|B|.
\end{aligned}
$$

Since $\bar{X}_1, \ldots, \bar{X}_{|B|}$ are independent, the first two terms cancel each other and the third is bounded by $\log|Y|/|B| \leq l_n(k) \log|\Sigma|/s(k)$, which goes to zero as k goes to infinity. Thus, by Proposition 4.5, $(1/|B|)\Sigma_{t \in B} h_i(Q_t, \bar{X}_t)$ tends to $\mathbf{d}_i - \eta \cdot W \cdot |\Sigma_i|$, where W is the greatest payoff appearing in the game.

Now, since X_t differs from $\bar{X}_t$ on an event with probability of at most $c_i(k)$ (see Proposition 4.2), one obtains

$$
(1/|B|)\Sigma_{t \in B} h_i(Q_i, X_t) \geq (1/|B|)\Sigma_{t \in B} h_i(Q_i, \bar{X}_t) - c_i(k) \cdot W \xrightarrow[k \to \infty]{} \mathbf{d}_i - \eta \cdot W \cdot |\Sigma_i|.
$$

Recall that this computation is confined to one particular $A(\bar{q})$. Since it holds for every vector $\bar{q}$ and for every $\eta > 0$ we get $\lim\inf(1/|B|)\Sigma_{t \in B} a_i^t(\bar{\sigma}^k) \geq \mathbf{d}_i$, as desired. ■

5. The Proof of Theorem 2

For the sake of simplicity we introduce the proof of Theorem 2; i.e., that $\overline{F} \subseteq \liminf_k N_k$ for a private case, $n = 3$ and $l_i(k) = k^i$ for $i = 1, 2, 3$. The proof for the general case follows exactly the same outline.

Suppose that $\overline{F}^s \neq \emptyset$. In order to show the desired result, it is sufficient to prove that $\overline{F}^s \subset \liminf N_k$. Let $\alpha \in \overline{F}^s \neq \emptyset$. α is feasible, so α is a convex combination of four payoffs: $\alpha = \sum_{u=1}^4 \gamma_u h(x^u)$, where $\gamma_u \geq 0$, $\sum \gamma_u = 1$, and $x^u \in \Sigma$ for all u. The strategy of player i consists, as usual in this area, of two plans. The first one is the master plan and the second one is the punishment plan, which is played when a deviation is detected. The strategy of player i, described below, is a function from Ω (some sample space) to $S_i^{l_i(k)}$: it is, obviously, equivalent to a mixed strategy in $\Delta(S_i^{l_i(k)})$.

Denote $e(k) = [(k \cdot \log k)^{1/2}]$. We can find integers $m_1(k), \ldots, m_4(k)$ with total sum $e(k)$, s.t. $|m_u(k)/e(k) - \gamma_u| < 1/e(k)$ for all $1 \leq u \leq 4$. Define

$$x = \Big(\underbrace{x^1, \ldots, x^1}_{m_1(k)}, \underbrace{x^2, \ldots, x^2}_{m_2(k)}, \ldots, \underbrace{x^4, \ldots, x^4}_{m_4(k)}\Big).$$

So, $z \in \Sigma^{e(k)}$. A word ν is *good* if it is a concatenation of z to itself several times: $\nu = (z', z, \ldots, z, z'')$, where z' is a tail of z and $z'' = (z^1, \ldots, z^s)$ is a head of z. For such ν define $\phi_i^k(\nu) = z_i^{s+1}$ (recall $z^{s+1} \in \Sigma$ and $z_i^{s+1} \in \Sigma_i$), where if $s + 1 > e(k)$ it is identified with $s + 1 (\mathrm{mod}\ e(k))$. The initial memory of player i is the element $(z'', z, \ldots, z) \in \Sigma^{l_i}$, where z'' is the appropriate tail of z. This concludes the description of the master plan (m.p.).

In cases where all the players play according to their m.p., at the first m_1 stages x^1 is played by the players. Then x^2 is played for m_2 stages and so on until x^4 is played m_4 times. Then x^1 is played again m_1 times and so on.

The punishment plan (p.p.) is executed when one of the players, say j, did not play z_j^{s+1} after the good word ν terminates with $z^1, \ldots, z^s$. In this case player i ($i \neq j$) notices that player j deviated and he has to play according to the (i, j)-p.p. (i punishes j).

We describe first the $(1, j)$-p.p. for $j > 1$. Let $p_j \in \Delta(\Sigma_1)$ be a mixed strategy of player 1 by which he (with the other players $\neq j$) can ensure that the payoff of player j does not exceed $\overline{d}_j$ (in the one-shot game), i.e., $\overline{d}_j = \mathrm{Min}_q \mathrm{Max}_r\, h_j(p_j, q, r)$, where the minimum is taken over all $q \in \Delta(\Sigma_i)$, $i \neq 1, j$, and the maximum over all $r \in \Delta(\Sigma_j)$.

Let $X_1^j, \ldots, X_{k^{10}}^j$, $j = 2, 3$, be mutually independent random variables (defined on the sample space Ω), which assume values in Σ_1. Furthermore, for all $1 \leq \nu \leq k^{10}$, X_ν^j are identically distributed as p_j. Since $\log k^{10}/e(k)$ tends to zero, by Proposition 4.3, the event $C^j(k) = \{X_1^j, \ldots, X_{k^{10}}^j$ has $e(k)$-cycles$\}$ has a probability $c^j(k)$ which satisfies $k^{10}c^j(k) \xrightarrow[k\to\infty]{} 0$.

There are a lot of $2e(k) - \Sigma_1$ words that do not appear in $\cup_{j>1}\{X_1^j(\omega), \ldots, X_{k^{10}}^j(\omega)) \mid \omega \in \Omega \backslash C^j\}$ and in $z_1 z_1 z_1$ (the corresponding word in $\Sigma_1^{3e(k)}$ to zzz). For example, a lot of $2e(k) - \Sigma_1$ words contain $e(k)$-cycles. Take two of them, denoted by y_1^2 and y_1^3 (y_1^j for player j). Now, player 1 can signal to himself: when he observes y_1^j he knows that he is playing according to the $(1, j)$-p.p. In a precise way, denote for all $\omega \in \Omega \backslash C^j(k)$ by $\nu^j(\omega)$ the word $(X_1^j(\omega), \ldots, X_{k^{10}}^j(\omega))$ with y_1^j inserted at the beginning and between blocks of length $l_1 - 2e(k)$:

$$\nu_j(\omega) = (y_1^j, X_1^j(\omega), \ldots, X_{l_1 - 2e(k)}^j(\omega), y_1^j, \ldots, X_{k^{10}}^j(\omega)).$$

The $(1, j)$-p.p. is to play repeatedly according to $\nu^j(\omega)$ (ignoring other player's actions). Note that for all $\omega \in \Omega \backslash (C^2 \cup C^3)$, any l_1-subword of $\nu^2(\omega)$ is different from any l_1-subword of $\nu^3(\omega)$ (because $y_1^2 \neq y_1^3$).

We are coming to the $(2, j)$-p.p. for $j \neq 2$ (2 punishes j). The $(2, 3)$-p.p. is similar to the $(1, j)$-p.p. described above: $X_1^3, \ldots, X_{k^{10}}^3$ are i.i.d. random variables which have the same distribution as q^3 has, where q^3 is the strategy of player 2, by which (with p^3 of player 1) player 3 is prevented from getting more than $\bar{d}_3$. For all ω, but of a set, say, D^3, $(X_i^3(\omega), \ldots, X_{k^{10}}^3(\omega))$ has no $e(k)$-cycles. Let y_2^3 be chosen as y_1^2 and y_1^3 were chosen. By y_2^3 player 2 codifies (for his own use) that the $(2, 3)$-p.p. is played.

The $(2, 1)$-p.p. and the $(3, 1)$-p.p. are played jointly. This is the place where differences in memories serve as a correlation device. In order to bring player 1 down to $\bar{d}_1$, they have to correlate their strategies. Since there are only three players there is no need for j $(j > 1)$ to coordinate the punishment of player 1 with players whose memory is shorter than player 1's. In general, a punishment should be coordinated in the manner described above. The new idea that should be introduced is the correlation of the "strong" players' strategies.

Let q be a strategy in $\Delta(\Sigma_2 \times \Sigma_3)$ (as if players 2 and 3 were one player) which satisfies $\mathrm{Max}_{p\in\Delta(\Sigma_1)} h_1(p, q) = \bar{d}_1$. Assume that q gives only m' points in $\Sigma_2 \times \Sigma_3$ a positive probability. Denote $m = [\log_2 m'] + 1$. Let $Y_1, \ldots, Y_{[k^2/2m]}$ be i.i.d. random variables that range to $\Sigma_2 \times \Sigma_3$, and distribute as q. Denote by Y_i' and by Y_i'' the projection of Y_i to Σ_2 and Σ_3, respectively. Since $[\log(k^2/2m)/e(k)] \to 0$, the event $D^1(k)$ of those ω for which either $(Y_1'(\omega), \ldots, Y_{k^2/2m}'(\omega))$ or $(Y_1''(\omega), \ldots, Y_{k^2/2m}''(\omega))$ have $e(k)$-cycles has a probability $b^1(k)$ which satisfies $k^2 b^1(k) \to 0$.

By elements in Σ_2^m player 2 can codify each of the m' points in $\Sigma_2 \times$

Σ_3 that have a positive probability according to q. We say that the element $\Psi(Y_t(\omega))$ of Σ_2^m encodes $Y_t(\omega)$.

Let y_2^1 and $\bar{y}_2^1$ be two $2e(k) - \Sigma_2$ words containing $e(k)$-cycles which are different from y_2^3 and from any $2e(k)$-word in $z_2z_2z_2$. The (2, 1)-p.p. has two parts: the first, played immediately after player 1 deviates, starts with $\bar{y}_2^1$ and the second starts with y_2^1. At the first $2e(k)$ stages after the deviation, player 2 plays $\bar{y}_2^1$ (recall $\bar{y}_2^1 \in \Sigma_2^{2e(k)}$). Then he plays $\Psi(Y_1(\omega))$ at the next m stages, then $\Psi(Y_2(\omega))$, $\Psi(Y_3(\omega))$, and so on for $m(k^2/2m) = k^2/2$ stages. The first part of the (2, 1)-p.p. terminates with y_2^1 (for $2e(k)$ stages).

The first part of the (2, 1)-p.p. informs player 3 how to play jointly with player 2 in the future. The second part starts with y_2^1, followed by $Y_1(\omega)$, $\ldots$, $Y_{[k^2/m]}(\omega)$, and again by y_2^1, and so on (ignoring all other players' actions). Note that $\bar{y}_2^1$ does not appear in the memory of player 2, while the player is playing according to the m.p., so once $\bar{y}_2^1$ is observed as a part of the memory, players 2 and 3 know that the punishment of player 1 should be started.

The (3, 2)-p.p. is described as follows. Let $\tilde{X}_1, \ldots, \tilde{X}_{k^{10}}$ be i.i.d. random variables with the same distribution as q_3, which satisfies $\overline{d}_2 = \text{Min}_{p_1 \in \Delta(\Sigma_1)} \text{Max}_{p_2 \in \Delta(\Sigma_2)} h_2(p_1, p_2, q_3)$. The event $E^2 = \{\tilde{X}_1, \ldots, \tilde{X}_{k^{10}}$ has an $e(k)$-cycle$\}$ has a very small probability. Denote by $\bar{y}_3^1, \bar{y}_3^1, \bar{y}_3^2$ $2e(k) - \Sigma_3 -$ words (that do not appear in $z_3z_3z_3$) with $e(k)$ cycles. As in the (2, 3)-p.p., $\bar{y}_3^2$ is inserted between blocks of length of k^2 in $\tilde{X}_1(\omega), \ldots, \tilde{X}_k^{10}(\omega)$.

The (3, 1)-p.p. is described in a different way. After discovering a deviation of player 1, player 3 plays $\bar{y}_3^1$ at the first $2e(k)$ stages, then for $k^2/2$ stages is playing arbitrarily, and then plays y_3^1 (for $2e(k)$ more stages). After y_3^1 he observes his memory and plays according to the m first actions of player 2 appearing immediately after the deviation had occurred (using Ψ^{-1}). Then he observes the next m actions of player 2 and plays accordingly, and so on for $[k^2/m]$ times. Then he plays again y_3^1. From this moment on he plays repeatedly in periods of length $[k^2/m] + 2e(k)$.

Denote the strategy described above by σ_i^k, and $\sigma^k = (\sigma_1^k, \sigma_2^k, \sigma_3^k)$. In what follows we omit k and keep in mind that strategies depend on k.

PROPOSITION 5.1. *$\|H(\sigma) - (\alpha_1, \alpha_2, \alpha_3)\|_\infty \to 0$, when $k \to \infty$.*

Proof. We have to consider only the m.p. Thus,

$$\|H(\sigma) - (\alpha_1, \alpha_2, \alpha_3)\|_\infty \leq W/e(k) \xrightarrow[k\to\infty]{} 0.$$

PROPOSITION 5.2. *σ is an equilibrium, for k sufficiently large.*

Proof. We introduce the argument for player 1. Fix a pure strategy of player 1, $\overline{\sigma}_1$. If $(\overline{\sigma}_1, \sigma_2, \sigma_3)$ determines the same stream of actions as σ, namely, the stream determined by the m.p., then $H_1(\bar{\sigma}_1, \sigma_2, \sigma_3) = H_1(\sigma)$. However, if by playing $\bar{\sigma}_1$, player 1 deviates at some stage (since $\bar{\sigma}_1$ is pure, it happens with probability 1), then from that stage on the p.p.'s are executed. Divide the set of stages that come after that stage into blocks of length $k^2/m + 2e(k)$. Fix a certain block B, and let U_t be the action of player 1 at its tth stage (U_t becomes a random variable). Denote by U_0 the memory of player 1 before B starts. Let also X_t be the joint action of players 2 and 3. On the set $\Omega \backslash D^1$, X_t is identical with $Y_{t-2e(k)}$ for $2e(k) + 1 \leq t \leq 2e(k) + [k^2/m]$ (recall that the first $2e(k)$ stages in B are devoted to signaling). Since $\text{prob}(D^1(k))k \to 0$ and since $(2e(k) + k^2/m)/k \to \infty$ we can get, by using Proposition 4.5, that for at least $k^{1.5}/m$ t's $U_t \perp^{\xi(k)} X_t$, where $\xi(k) \xrightarrow[k\to\infty]{} 0$. Hence the average payoff of player 1 in B is at most

$$\frac{2e(k)W}{2e(k) + k^2/m} + \frac{W}{k} + 2\xi(k)W + \bar{d}_1$$

(the first term is for the stages which are devoted to signaling by $\bar{y}_2^1$ and $\bar{y}_3^1$, the second term stands for those stages t for which it is not true that $U_t \perp^{\xi(k)} X_t$, the third term is for the $\xi(k)$-independence error, and the fourth term is because all Y_t are distributed as q). Since the first three terms tend to zero, and $\alpha \in F^s$, the expected payoff of player 1 in B is less than α_1, for k sufficiently large. The same computation holds for all the blocks, and therefore $H_1(\bar{\sigma}_1, \sigma_2, \sigma_3) < \alpha_1$ for k sufficiently large. It holds for every pure strategy $\bar{\sigma}_1$ and thus for every mixed strategy.

Similar arguments hold for all other possibilities of deviations, and the proof of Theorem 2 is finished. ■

5. Remarks

The punishment plans defined above are of the "trigger" type. We could define it in such a way that the master plan will again be played even though a deviation had occurred. Note first that by a deviation from the master plan sustaining α, players n and $n - 1$ can gain by at most W. Thus their punishment, in which the other players should not be coordinated, can be of a fixed length. That is to say, the length of player n's punishment phase should not increase with k. Hence, from a certain k on, the punishment phase can be much shorter than $l_1(k)$. In that case, all the players can count up to the length of the punishment phase and return simultaneously to the master plan. The same argument holds for player $n - 1$'s punishment.

Returning to the master plan by counting can be done only if the punished player is $n-1$ or n. In all other cases, punishing requires coordination. Namely, the punishment phase should be longer than the memory of the punished player. However, in this case, player n or $n-1$ can send a message, encoded by a certain combination of actions—one for each player. Observing these messages, other players know to return to the master plan. Two points should be emphasized:

1. All the players should return to the master plan simultaneously. Therefore, the message sent by players n and $n-1$ should be of the same length. Immediately after the messages have been sent, all the players should play the master plan.
2. It may be that player j's punishment is beneficial to player n, for instance. This may give incentive to player n not to send the message and to remain in the p.p. In order to construct a perfect equilibrium, we should eliminate that incentive. For this reason we used two players, $n-1$ and n, to send a message. Observing at least one of these messages will lead a player to return to the master plan. Thus, unilateral deviation of player $n-1$ or player n will not be beneficial to them.

We therefore obtain that all the results alluded to above hold for perfect equilibrium as well.

Appendix

Let $\beta_b(x_1, \ldots, x_b, y_1, \ldots, y_b) = \Sigma\Pi x_j^{\varepsilon_j}(1-x_j)^{1-\varepsilon_j} y_j^{\varepsilon_j}(1-y_j)^{1-\varepsilon_j}$, where the summation is taken over all $(\varepsilon_1, \ldots, \varepsilon_b) \in (0, 1)^b$.

LEMMA. *If $\eta \le x_j, y_j \le 1-\eta$ for all j, then $\beta_b(x_1, \ldots, x_b, y_1, \ldots, y_b) \le \beta_b(\eta, \ldots, \eta)$.*

Proof. (Applying the Induction Technique of Holzman *et al.* 1986). The lemma is proved through induction on b. For $b = 1$, $\beta_1(x_1, y_1) = x_1 y_1 + (1-x_1)(1-y_1)$. Since β_1 is linear in x,

$$\beta_1(x_1, y_1) \le \beta_1(\eta, y_1) \qquad \text{if } y_1 \le \tfrac{1}{2} \qquad \text{and, similarly,} \qquad \beta_1(\eta, y_1) \le \beta_1(\eta, \eta).$$

Moreover,

$$\beta_1(x_1, y) \ge \beta_1(1-\eta, y_1) \qquad \text{if } y_1 \le \tfrac{1}{2}$$

and similarly,

$$\beta_1(1-\eta, y_1) \le \beta_1(1-\eta, 1-\eta).$$

A similar argument holds for x_1 and, therefore, the desired inequality for $b = 1$.
Now let $b > 1$:

$$\beta_b(x_1, \ldots, x_b, y_1, \ldots, y_b) = x_b y_b \beta_{b-1}(x_1, \ldots, x_{b-1}, y_1, \ldots, y_{b-1}) + (1 - x_b)(1 - y_b)\beta_{b-1}(x_1, \ldots, x_{b-1}, y_1, \ldots, y_{b-1})$$

(by the induction hypothesis)

$$\leq x_b y_b \beta_{b-1}(\eta, \ldots, \eta) + (1 - x_b)(1 - y_b)\beta_{b-1}(\eta, \ldots, \eta)$$

(by the first step of the induction)

$$\leq \beta_{b-1}(\eta, \ldots, \eta)\beta_1(x_b, y_b) \leq \beta_{b-1}(\eta, \ldots, \eta)\beta_1(\eta, \eta) = \beta_b(\eta, \ldots, \eta). \quad \blacksquare$$

A similar technique would work for any other multilinear function corresponding to overlapping cycles.

References

AUMANN, R. (1981). "Survey of Repeated Games" in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*, 11–42. Mannheim/Vienna/Zurich: Bibliographisches Institut.

AUMANN, R., AND SORIN, S. (1980). "Cooperation and Bounded Recall," *Games Econ. Behav.* **1,** 5–39.

BEN PORATH, E. (1989). "Repeated Games with Bounded Complexity," *Games Econ. Behav.*, in press.

GILBOA, I., AND SCHMEIDLER, D. (1989). "Infinite Histories and Steady Orbits in Repeated Games," mimeograph.

HOLZMAN, R., LEHRER, E., AND LINIAL, N. (1986). "Some Bounds for the Banzhaf Index and Other Semi-values," *Math. Oper. Res.* **13,** 358–363.

KALAI, E., AND LEHRER, E. (1990). "Rational Learning Leads to Nash Equilibrium," mimeograph.

LEHRER, E. (1988). "Repeated Games with Stationary Bounded Recall Strategies," *J. Econ. Theory* **46,** 130–144.

LEHRER, E. (1991). "Internal Correlation in Repeated Games," *Int. J. Game Theory* **19,** 431–456.

MCELIECE, R. I. (1977). *The Theory of Information and Coding.* Reading, MA: Addison-Wesley.

NEYMAN, A. (1985). "Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoners' Dilemma," *Econ. Lett.* **19,** 227–229.

RUBINSTEIN, A. (1986). "Finite Automata Play—The Repeated Prisoners' Dilemma," *J. Econ. Theory* **39,** 83–96.

SMORODINSKY, M. (1971). *Ergodic Theory, Entropy.* New York/Berlin: Springer-Verlag.