Games and Economic Behavior **38,** 362–399 (2002)
doi:10.1006/game.2001.0887, available online at http://www.idealibrary.com on IDEAL®

Correlated Equilibrium in Stochastic Games[1]

Eilon Solan[2]

MEDS Department, Kellogg Graduate School of Management, Northwestern University, and the School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel

and

Nicolas Vieille[3]

Laboratoire d'Econométrie, Ecole Polytechnique and Département Finance et Economie, HEC, 1, rue de la Libération, 78 Jouy-en-Josas, France

Received October 13, 1998

We study the existence of uniform correlated equilibrium payoffs in stochastic games. The correlation devices that we use are either autonomous (they base their choice of signal on previous signals, but not on previous states or actions) or stationary (their choice is independent of any data and is drawn according to the same probability distribution at every stage). We prove that any n-player stochastic game admits an autonomous correlated equilibrium payoff. When the game is positive and recursive, a stationary correlated equilibrium payoff exists. *Journal of Economic Literature* Classification Numbers: C72, C73.

Key Words: stochastic games; correlated equilibrium; positive recursive games.

1. INTRODUCTION

A stochastic game is played in stages. At every stage the game is in some state of the world, and each player, given the whole history (including the current state), chooses an action in his action space. The action combination that was chosen by all players, together with the current state, determines

[1] The first author thanks the Center for Rationality and Interactive Decision Theory at the Hebrew University of Jerusalem and the Laboratoire d'Econométrie de l'Ecole Polytechnique for their hospitality and the Institute of Mathematics at the Hebrew University of Jerusalem for financial support.

[2] To whom correspondence should be addressed. E-mail: eilons@post.tau.ac.il.

[3] E-mail: vieille@poly.polytechnique.fr.

0899-8256/02 $35.00

the daily payoff that each player receives and the probability distribution according to which the new state of the game is chosen.

In a stochastic game the players have two, seemingly contradictory, goals. First, they need to ensure that their future opportunities remain high. At the same time, they should make sure that their daily payoff is also high. When the game is discounted, there is a clear balance between the two goals. When the game is undiscounted, the future opportunities are "infinitely" more important than the daily payoff, but if players do not collect any payoff, they will get nothing.

Stochastic games were introduced by Shapley (1953) who proved the existence of the discounted value in two-player zero-sum games, as well as the existence of stationary optimal strategies. This result was generalized for discounted equilibria in n-player games by Fink (1964).

Recursive games were introduced by Everett (1957). These are stochastic games where the payoff for the players in non-absorbing states is zero (a state is absorbing if the probability to leave it, whatever the players play, is 0). Thus, in recursive games the players collect payoff only once the game is absorbed, and therefore, in non-absorbing states, they play only for future opportunities. A recursive game is positive if the payoff for the players in absorbing states is positive.

Everett proved the existence of the (uniform) value for two player, zero-sum recursive games and the existence of stationary ε-optimal strategies.

Existence of the value in two-player zero-sum stochastic games was proved by Mertens and Neyman (1981). It is well known (see, e.g., Blackwell and Ferguson, 1968) that uniform optimal strategies, as well as stationary uniform ε-optimal strategies, need not exist.

Recently Vieille proved the existence of an equilibrium payoff in every two-player non-zero-sum stochastic game. In his proof Vieille deals separately with daily payoffs and future opportunities. He divides the state space into two subsets—states where the players play for high daily payoff, and states where they play for high future opportunities.

First, Vieille (2000a) identified those states where the players play for high daily payoffs and showed how the players should play in those states. He then turns them into absorbing states. In the other states, players play only for future opportunities, disregarding their daily payoff. Thus, the payoff in those states can be redefined to be 0.

To summarize, Vieille proves that if every two-player recursive game that satisfies some property admits an equilibrium payoff then every two-player stochastic game admits an equilibrium payoff.

Second, Vieille (2000c) shows how the players should play in those states where they play for future opportunities and proves the existence of equilibrium payoffs for this class of recursive games.

This view, that in some states players play for high daily payoffs and in other states they play for good future opportunities, highlights the importance of recursive games in the study of general stochastic games.

For stochastic games with more than two players, very little is known. Even though some special classes were solved (see, e.g., Solan, 1999; Solan and Vieille, 1998), proving, or disproving, the existence of equilibrium payoffs in general seems a daunting task. We study here the existence of *correlated* equilibrium payoffs in n-player stochastic games.

Correlation devices were introduced by Aumann (1974, 1987). A correlation device chooses for every player a private signal before the start of play and sends to each player the signal chosen for him. Each player can base his choice of an action on the private signal that he has received.

For multi-stage games, various generalizations of correlation devices have been introduced (Fudenberg and Tirole, 1991). (i) The most general device receives at every stage some private message from each player. It then chooses for each player a private signal for that stage (*communication device*; Forges, 1986, 1988; Myerson, 1986; Mertens, 1994). (ii) The most restricted device chooses, as in the case of one-shot games, one private signal before the start of play (*correlation device*; Forges, 1986). (iii) In between, there are devices that choose private signals at every stage, but base their choice only on the current state (*weak correlation devices*; Nowak, 1991) or only on previous signals (*autonomous correlation devices*; Forges, 1986).

Solan (2000) proved that every feasible and individually rational payoff in a stochastic game is a correlated equilibrium payoff, where the correlation device chooses at every stage a signal that depends on the previous signal. However, he leaves open the question whether there exists a feasible and individually rational payoff.

In the present paper we study two types of correlation devices: (i) stationary devices that choose at every stage a signal according to the same probability distribution, independent of any data, and (ii) autonomous devices that base their choice of new signal on the previous signal, but not on any other information.

We prove two results: (a) Every stochastic game admits a correlated equilibrium, using an autonomous correlation device. The equilibrium path is sustained using threat strategies, in which players punish a deviator by his min–max value. This means that players need not correlate in the punishment phase. Moreover, punishment occurs only if a player disobeys the recommendation of the device. (b) If the game is positive recursive, then the correlation device can be taken to be stationary.

Both proofs utilize various methods that appeared in the literature. For general stochastic games, we first construct a "good" strategy profile, meaning a profile that yields all players a high payoff, in which no player can profit by a unilateral deviation that is followed by an indefinite punishment. The construction uses the technique of Mertens and Neyman (1981).

We then follow Solan (2000) and define a correlation device that *mimics* that profile: the device chooses for each player an action according to the probability distribution given by the profile and recommends that each player play that action. To make deviations nonprofitable, the device reveals to all players the recommendations in the previous stage. This way, a deviation is detected immediately and can be punished. In particular, the device that we construct is not canonical (Forges, 1988).

For positive recursive games we use a variant of the technique of Vieille (2000c). We define for every $\varepsilon > 0$ a continuous function from the space of fully mixed strategy profiles into itself that should be thought of as an approximate best reply. This function has a fixed point, and by studying the asymptotic behavior of the sequence of fixed points we are able to define a simple stationary correlation device that induces a correlated equilibrium.

Even though our proofs use many known techniques, from Mertens and Neyman (1981) to Vieille (2000b, 2000c) and Solan (2000), the paper is self-contained, and no acquaintance with this literature is assumed.

The paper is arranged as follows. In Section 2 we introduce the model and the main results. After some preliminaries in Section 3 we describe the correlation devices that are used, and we give sufficient conditions for the existence of autonomous and stationary correlated equilibrium payoffs. In Section 4 we prove that every stochastic game admits an autonomous correlated equilibrium payoff, and in Section 5 we prove that every positive recursive game admits a stationary correlated equilibrium payoff.

2. THE MODEL AND THE MAIN RESULTS

A *stochastic game* G is given by (i) a finite set of players N, (ii) a finite set of states S, (iii) for every player $i \in N$, a finite set of available actions A^i (we denote $A = \times_{i\in N} A^i$), (iv) a transition rule $q : S \times A \to \Delta(S)$, where $\Delta(S)$ is the space of all probability distributions over S, and (v) a daily payoff function $r : S \times A \to \mathbf{R}^N$. We assume w.l.o.g. that $|r| \leq 1$.

The game lasts for infinitely many stages. The initial state s_1 is given. At stage n, the current state s_n is announced to the players. Each player i chooses an action $a_n^i \in A^i$; the action combination $a_n = (a_n^i)_{i\in N}$ is publicly announced, s_{n+1} is drawn according to $q(\cdot \mid s_n, a_n)$, and the game proceeds to stage $n+1$.

A state s in a stochastic game is *absorbing* if $q(s \mid s, a) = 1$ for every $a \in A$. We denote by S^* the subset of absorbing states.

The game is *recursive* if $r^i(s, \cdot) = 0$ for every non-absorbing state $s \in S \backslash S^*$ and every player $i \in N$. It is *positive* if $r^i(s, \cdot) > 0$ for every absorbing state $s \in S^*$ and every player $i \in N$.

DEFINITION 2.1. An *autonomous correlation device* is a pair $\mathscr{D} = (((M_n^i)_{i\in N}, d_n)_{n\in\mathbf{N}})$, where (i) M_n^i is a finite set of signals for player i at stage n, and (ii) $d_n : M(n) \to \Delta(M_n)$, where $M_n = \times_{i\in N} M_n^i$ and $M(n) = M_1 \times M_2 \times \cdots \times M_{n-1}$.

A *stationary correlation device* is a pair $\mathscr{D} = ((M^i)_{i\in N}, d)$, where $d \in \Delta(M)$ and $M = \times_{i\in N} M^i$.

Thus, a stationary correlation device is an autonomous correlation device $\mathscr{D} = (((M_n^i)_{i\in N}, d_n)_{n\in\mathbf{N}})$, where M_n^i is independent of n and d_n is a constant function that is independent of n.

Given a correlation device $\mathscr{D}$ we define an extended game $G(\mathscr{D})$. The game $G(\mathscr{D})$ is played exactly as the game G, but at the beginning of each stage n, a signal combination $m_n = (m_n^i)_{i\in N}$ is drawn according to $d_n(m_1, \ldots, m_{n-1})$ (or according to d if the device is stationary) and each player i is informed of m_n^i. Then, each player may base his choice of a_n^i also on previous signals $m_1^i, \ldots, m_n^i$ that he received.

At stage n, player i observes an element of $H_n^i(\mathscr{D}) = \Pi_{k=1}^{n-1}(S \times M_k^i \times A) \times S \times M_n^i$. Therefore, a (behavioral) *strategy for player i* in $G(\mathscr{D})$ is a function σ^i: $H^i(\mathscr{D}) \to \Delta(A^i)$, where $H^i(\mathscr{D}) = \cup_{n\in\mathbf{N}} H_n^i(\mathscr{D})$. We denote by $\sum^i(\mathscr{D})$ the set of all strategies of player i in $G(\mathscr{D})$.

A *profile* $\sigma = (\sigma^i)_{i\in N}$ is a vector of strategies, one for each player. We denote $\sum(\mathscr{D}) = \times_{i\in N} \sum^i(\mathscr{D})$, the space of all profiles in $G(\mathscr{D})$.

The set of *finite histories* is $H(\mathscr{D}) = \cup_{n\in\mathbf{N}}(\Pi_{k=1}^{n-1}(S \times M_k \times A) \times S \times M_n)$. This is the set of all histories which are observed by an observer, who observes the signals that are received by *all* the players. The set of all *infinite histories* is denoted by $H_\infty(\mathscr{D})$. It is defined accordingly. We endow this space with the σ-algebra generated by all the finite cylinders.

Every correlation device $\mathscr{D}$, every profile $\sigma \in \sum(\mathscr{D})$, and every initial state $s \in S$ induce a probability measure $\mathbf{P}_{\mathscr{D}, s, \sigma}$ over $H_\infty(\mathscr{D})$; that is, the probability measure induced by σ and $\mathscr{D}$, given the initial state, is s. We denote expectation w.r.t. this measure by $\mathbf{E}_{\mathscr{D}, s, \sigma}$.

Define for every player $i \in N$ and every $n > 0$ the expected payoff of player i during the first n stages by

$$\gamma_n^i(\mathscr{D}, s, \sigma) = \mathbf{E}_{\mathscr{D}, s, \sigma}\left[\frac{1}{n}\sum_{j=1}^{n} r^i(s_j, a_j)\right]$$

and for every $\lambda \in (0, 1)$ the λ-discounted payoff as

$$\gamma_\lambda^i(\mathscr{D}, s, \sigma) = \mathbf{E}_{\mathscr{D}, s, \sigma}\left[\lambda\sum_{j=1}^{\infty}(1-\lambda)^{j-1} r^i(s_j, a_j)\right].$$

We denote by $G_n(\mathcal{D})$ and $G_\lambda(\mathcal{D})$ the games with strategy sets $(\sum^i(\mathcal{D}))_{i\in N}$ and payoff functions γ_n and γ_λ, respectively.

DEFINITION 2.2. A payoff vector $\gamma \in R^{S\times N}$ is a (uniform) *autonomous* (*resp. stationary*) *correlated equilibrium payoff* if for every $\varepsilon > 0$ there exists an autonomous (resp. stationary) correlation device $\mathcal{D}$ and a profile σ in $\sum(\mathcal{D})$ such that the following two conditions are satisfied:

1. There exists a finite horizon $n_0 \in \mathbf{N}$ such that for every $n \geq n_0$, σ is an ε-equilibrium in the game $G_n(\mathcal{D})$ with length n and

$$\|\gamma_n(\mathcal{D}, s, \sigma) - \gamma_s\| \leq \varepsilon \qquad \forall s \in S.$$

2. There exists $\lambda_0 \in (0, 1)$ such that for every $\lambda \in (0, \lambda_0)$, the profile σ is an ε-equilibrium in the game $G_\lambda(\mathcal{D})$ and

$$\|\gamma_\lambda(\mathcal{D}, s, \sigma) - \gamma_s\| \leq \varepsilon \qquad \forall s \in S.$$

Note that for every $\varepsilon > 0$ a different correlation device may be used. It is known that Conditions 1 and 2 in the above definition are equivalent. We shall deal with the former one.

The main result of the paper is:

THEOREM 2.3. *Every stochastic game admits an autonomous correlated equilibrium payoff.*

The equilibrium path is sustained by threat of punishment by the min–max value, and a player is punished only if he disobeys the recommendation of the device.

When the game is positive and recursive, one can find a correlated equilibrium payoff where the device is stationary:

THEOREM 2.4. *Every positive recursive game admits a stationary correlated equilibrium payoff, where the correlation device is independent of* ε.

In spite of the fact that the device is independent of ε, the profile that is used by the players does depend on ε.

3. THE CORRELATION DEVICES

The purpose of this section is to obtain sets of sufficient conditions for the existence of autonomous and stationary correlated equilibrium payoffs.

3.1. *Preliminaries*

The mixed extension of q to $S \times \Delta(\times_{i\in N} A^i)$ is still denoted by q. The transition in state s for a mixed action x is denoted by $q(\cdot \mid s, x)$ or $q_{s,x}$.

For every finite set K and every probability distribution $\mu \in \Delta(K)$, $\mu[k]$ is the probability of k under μ. Any element $k \in K$ is identified with the probability distribution in $\Delta(K)$ that assigns probability 1 to k. For any probability measure μ and real valued function u defined over S, we denote by $\mu u = \sum_{s\in S} \mu[s]u(s)$ the expectation of u under μ.

Denote by $H_n = (S \times A)^{n-1} \times S$ the space of all histories of length n in G and denote by $H = \cup_{n\in \mathbf{N}} H_n$ the space of all finite histories. $H_\infty = (S \times A)^{\mathbf{N}}$ is the space of all infinite histories. We identify the space of histories of length 1, H_1, with the state space S. For every finite history $h \in H$, s_h denotes its last state.

A *strategy* of player i in G is a function τ^i: $H \to \Delta(A^i)$. A *profile* in G is a vector $\tau = (\tau^i)_{i\in N}$ of strategies, one for each player. A *correlated profile* in G is a function τ: $H \to \Delta(A)$.

In general, the symbol τ denotes a profile in G, $\tilde{\tau}$ denotes a correlated profile in G, and σ denotes a profile in an extended game $G(\mathcal{D})$.

Stationary strategies of player i are strategies that depend only on the current state and not on previous signals, states, or actions. Thus, a stationary strategy of player i can be identified with an element $x^i = (x^i_s)_{s\in S} \in (\Delta(A^i))^S$, with the understanding that x^i_s is the lottery used by player i to select his action in state s. We define $X^i = (\Delta(A^i))^S$ to be the set of stationary strategies of player i, $X = \times_{i\in N} X^i$ to be the set of stationary profiles, and $X^{-i} = \times_{j\neq i} X^i$ to be the set of stationary profiles of players $N\setminus\{i\}$. In particular, for every $x \in X$, every $s \in S$, and every $u : S \to \mathbf{R}$, $q_{s,x} u = \sum_{s'\in S} q(s' \mid s, x)u(s')$ is the expectation of u under $q(\cdot \mid s, x)$.

Every finite history $h \in H$ and every correlated profile $\tilde{\tau}$ in G induce a probability measure $\mathbf{P}_{h,\tilde{\tau}}$ over H_∞—the measure induced by $\tilde{\tau}$ in the subgame starting after the history h. The corresponding expectation operator is denoted by $\mathbf{E}_{h,\tilde{\tau}}$.

For any subset $L \subseteq N$, we denote $A^L = \times_{i\in L} A^i$ and $A^{-L} = \times_{i\notin L} A^i$.

For every $c, c' \in \mathbf{R}^n$, $c \geq c'$ if and only if $c^i \geq c'^i$ for every $i = 1, \ldots, n$.

3.2. *The Min–Max Value*

For every correlated profile $\tilde{\tau}$, every history $h \in H_m$, and every $n \in \mathbf{N}$ define

$$\gamma^i_n(h, \tilde{\tau}) = \frac{1}{n}\mathbf{E}_{h,\tilde{\tau}}\big(r^i(s_m, a_m) + \cdots + r^i(s_{m+n-1}, a_{m+n-1})\big),$$

that is, the expected average payoff in the first n stages following the history h.

For every player i and every initial state s let $v^i(s)$ be the min–max value of player i at state s in G. That is, $v^i(s)$ is the unique real number such that for every $\varepsilon > 0$ there exists $n_0 \in \mathbf{N}$ and a (non-correlated) profile $\tau_{i,\varepsilon}^{-i}$ of players $N \setminus \{i\}$ that satisfy for every strategy τ^i of player i

$$\gamma_n^i\big(s, \tau_{i,\varepsilon}^{-i}, \tau^i\big) \le v^i(s) + \varepsilon \qquad \forall\, n > n_0$$

and for every profile τ^{-i} of players $N \setminus \{i\}$ there exists a strategy τ^i of player i that satisfies

$$\gamma_n^i(s, \tau^{-i}, \tau^i) \ge v^i(s) - \varepsilon \qquad \forall\, n > n_0.$$

Thus, up to an ε, players $N \setminus \{i\}$ can lower the expected payoff of player i down to $v^i(s)$, but they cannot lower it any more. The profile $\tau_{i,\varepsilon}^{-i}$ is an *ε-punishment profile against player i*.

For every $\lambda \in (0,1)$ let $v_\lambda^i(s)$ be the λ-discounted min–max value of player i. Using the method of Bewley and Kohlberg (1976) one can show that the function $\lambda \mapsto v_\lambda^i(s)$ has bounded variation.

LEMMA 3.1. *For every player $i \in N$ and every state $s \in S$, $v^i(s)$ exists. Moreover, $v^i(s) = \lim_{\lambda \to 0} v_\lambda^i(s)$.*

This result was proved by Mertens and Neyman (1981) for two-player stochastic games, and Neyman (1988), using the same method, proved in an unpublished paper that the result holds for n-player games.

3.3. *An Autonomous Correlation Device*

3.3.1. *Mimicking a Strategy*

In this section we define for every correlated profile $\tilde{\tau}$ in G two autonomous correlation devices. The first device $\mathcal{D}^1 = \mathcal{D}^1(\tilde{\tau})$ *mimics* the profile $\tilde{\tau}$. That is, at stage n it sends to each player a vector of actions, one for every history of length n. The action combination for history h is chosen according to $\tilde{\tau}(h)$.

The second device $\mathcal{D}^2 = \mathcal{D}^2(\tilde{\tau})$ is an augmentation of $\mathcal{D}^1$. In addition to sending recommendations, the device reveals its recommendations from the previous stage. This way all players can compare the realized actions of each player to the recommended action and detect deviations.

We shall now define formally these two devices.

Let $\tilde{\tau}$ be a correlated profile in G. Define an autonomous correlation device $\mathcal{D}^1 = (((M_n^i)_{i \in N}, d_n^1)_{n \in \mathbf{N}})$ as follows. $M_n^i = (A^i)^{H_n}$, and $d_n^1(\cdot) = \otimes_{h \in H_n} \tilde{\tau}(h)$ is a product probability distribution, which is independent of $m_1, \ldots, m_{n-1}$.

Define a profile σ_0 in $G(\mathcal{D}^1)$ by

$$\sigma_0^i\big(s_1, m_1^i, a_1, \ldots, s_n, m_n^i\big) = m_n^i(s_1, a_1, \ldots, s_n). \tag{1}$$

In words, since m_n^i is a vector of actions, one for each possible history, in σ_0^i, player i plays the action that corresponds to the realized history.

The probability measure $\mathbf{P}_{s,\tilde{\tau}}$ over H_∞ coincides with the marginal probability measure induced by $\mathbf{P}_{\mathcal{D}^1, s, \sigma_0}$ over H_∞. We say that the device $\mathcal{D}^1$ *mimics* the profile $\tilde{\tau}$.

We now define a second autonomous correlation device $\mathcal{D}^2 = (((\overline{M}_n^i)_{i\in N}, d_n^2)_{n\in\mathbf{N}})$ as follows. $\overline{M}_n^i = M_n^i \times M_{n-1}$ (M_n^i and M_{n-1} are the sets defined for $\mathcal{D}^1$). Denote the signal at stage n by $\bar{m}_n = (m_n, \tilde{m}_n) \in M_n \times (M_{n-1})^N$. We define

$$\begin{aligned} d_n^2((m_1, \tilde{m}_1), \ldots, (m_{n-1}, \tilde{m}_{n-1})) &= d_n^1 \otimes m_{n-1} \otimes m_{n-1} \otimes \cdots \otimes m_{n-1} \\ &\in \Delta\big(A^{H_n} \times (A^{H_{n-1}})^N\big) \\ &= \Delta\big(\times_{i\in N}\big((A^i)^{H_n} \times A^{H_{n-1}}\big)\big) = \Delta(\overline{M}_n). \end{aligned}$$

Thus d_n^2 selects a vector of action combinations according to d_n^1 and reveals to everyone the previous recommended vector of signals.

Note that the marginal probability measure $\mathbf{P}_{\mathcal{D}^2, s, \sigma_0}$ over H_∞ coincides with the probability measure $\mathbf{P}_{s,\tilde{\tau}}$, where σ_0 is the analogue of (1) in $G(\mathcal{D}^2)$.

Remark.[4] The procedure used here in defining $\mathcal{D}^1$ shows that every correlation device can be considered as an autonomous (but not necessarily canonical) correlation device.

3.3.2. *The Sufficient Condition*

DEFINITION 3.2. Let $\gamma = (\gamma(h))_{h\in H}$ be a function defined on finite histories with values in R^N and let $\tilde{\tau}$ be a correlated profile. *Average payoffs under* $\tilde{\tau}$ *converge* to γ if $\lim_{n\to\infty} \gamma_n(h, \tilde{\tau}) = \gamma(h)$, uniformly in h.

In words, in every subgame the sequence of average payoffs induced by $\tilde{\tau}$ has a limit (which may depend on the subgame).

For every correlated probability distribution $y \in \Delta(A)$, every player $i \in N$, and every action $a = (a^{-i}, a^i) \in \text{supp}(y)$ define $(y \mid a^i) \in \Delta(A^{-i})$ to be the conditional probability distribution induced by y, given that the action chosen for player i is a^i. Formally,

$$(y \mid a^i)[b^{-i}] = y[b^{-i}, a^i]/y[A^{-i} \times \{a^i\}].$$

Note that since $a \in \text{supp}(y)$, the denominator does not vanish.

In particular, for every correlated profile $\tilde{\tau}$ and every finite history $h \in H$, $(\tilde{\tau}(h) \mid a^i)$ is the conditional probability over A^{-i}, given that the action chosen for i is a^i.

[4]We thank an anonymous referee for drawing our attention to this point.

DEFINITION 3.3. Let $\varepsilon > 0$, and let $\gamma = (\gamma(h))_{h \in H} \in (\mathbf{R}^N)^H$ be a function that assigns to every finite history a payoff vector. A correlated profile $\tilde{\tau}$ is *ε-individually rational* with respect to γ if for every stage $n \in \mathbf{N}$, every history $h_n \in H_n$ of length n, every action combination $a = (a^i)_{i \in N} \in \operatorname{supp}(\tilde{\tau}(h_n))$, every player $i \in N$, and every action $b^i \in A^i$

$$\mathbf{E}_{s,\tilde{\tau}}[\gamma^i(h_{n+1}) \mid h_n, a^i] \geq q_{s_n, (\tilde{\tau}(h) \mid a^i), b^i} v^i - \varepsilon. \tag{2}$$

In words, a correlated profile $\tilde{\tau}$ is ε-individually rational if given any finite history h_n, any player i who knows the action recommended to him and should choose between the following two alternatives, (i) play the recommended action and get the payoff defined by γ or (ii) play any action b^i and be punished at the min–max level, cannot profit more than ε in any sufficiently long game by choosing the latter.

THEOREM 3.4. *If for every ε there exists $\gamma = (\gamma(h))_{h \in H} \in (\mathbf{R}^N)^H$ and a correlated profile $\tilde{\tau}$ in G such that* (i) *average payoffs under $\tilde{\tau}$ converge to γ and* (ii) *$\tilde{\tau}$ is ε-individually rational w.r.t. γ, then G admits an autonomous correlated equilibrium payoff.*

Proof. This theorem is a weaker version of Proposition 4.6 in Solan (2000). We provide here a sketch of the proof.

We construct an autonomous correlation device in the following manner. At every stage the device recommends an action to each player and reveals to each player the actions that it recommended to *all* the players at the previous stage. This way, a deviation of any player is detected immediately by all other players and can be punished with the min–max value.

Fix $\varepsilon > 0$, and let $\gamma = \gamma_\varepsilon$ and $\tilde{\tau}$ satisfy the conditions for this ε.

Consider the autonomous correlation device $\mathcal{D} = \mathcal{D}^2(\tilde{\tau})$ that was defined in Section 3.3.1. The device sends at every stage n a vector of signals for every player: (i) for every history $h \in H_n$ a recommended action $a^i(h)$ and (ii) the vector of signals it sent to all the players at stage $n - 1$.

Define a profile σ in the extended game $G(\mathcal{D})$ as follows. After history h of length n, player i plays:

- If no deviation was ever detected, play the recommended action $a^i(h)$.
- For every player j, compare his realized action from the previous stage a^j_{n-1} to the recommendation of the device $a^j(h')$, where h' is the $n \to 1$-prefix of h. If some player j has deviated from the recommendation, switch to the punishment strategy $\tau^i_{j,\varepsilon}$ forever. If more than one player deviated at the same stage, ignore this deviation.

The marginal probability measure $\mathbf{P}_{\mathcal{D},s,\sigma}$ over H_∞ coincides with $\mathbf{P}_{s,\tilde{\tau}}$. Hence $\gamma^i_n(\mathcal{D}, h, \sigma) = \gamma^i_n(h, \tilde{\tau})$. Since average payoffs under $\tilde{\tau}$ converge to

γ, there exists $n_1 \in \mathbf{N}$ such that for every $n \geq n_1$ and every finite history $h \in H$,

$$\gamma^i(h) + \varepsilon \geq \gamma_n^i(\mathcal{D}, h, \sigma) \geq \gamma^i(h) - \varepsilon. \tag{3}$$

Since $\tilde{\tau}$ is ε-individually rational and if a player disobeys the recommendation of the device he is punished by his min–max value, it follows that for every $n \geq n_0 + n_1$, where n_0 is defined in the definition of the min–max value in Section 3.2, and for every $\sigma'^i \in \sum^i(\mathcal{D})$,

$$\gamma_n^i(\mathcal{D}, s, \sigma) \geq \gamma_n^i(\mathcal{D}, s, \sigma^{-1}, \sigma'^i) - 3\varepsilon. \tag{4}$$

Finally, for every ε let $\gamma_\varepsilon^\star \in \mathbf{R}^{S\times N}$ be defined by $\gamma_\varepsilon^\star(s) = \gamma_\varepsilon(s)$. By (3) and (4), every accumulation point of the family $\gamma_\varepsilon^\star$, as ε goes to 0, satisfies condition 1 of Definition 2.2; hence it is a correlated equilibrium payoff. Since the state space is finite, such an accumulation point exists. ■

3.4. *Stationary Correlation Device*

In this section, we provide a sufficient condition for the existence of stationary correlated equilibrium payoffs. This requires a more detailed analysis of the structure of the game. The condition we present is stated for general stochastic games; however, we can only prove that it is satisfied for positive recursive games.

3.4.1. *Communicating Sets*

Let $x \in X$ be a stationary profile.

The stationary profile $y \in X$ is a *perturbation* of x if $\text{supp}(x_s^i) \subseteq \text{supp}(y_s^i)$ for every player $i \in N$ and every state $s \in S$.

A set $C \subset S$ is *stable* under x if $q(C \mid s, x) = 1$ for every $s \in C$. The set is *communicating* under x if for every state $s' \in C$ there exists a perturbation y of x such that C is stable under y and

$$\mathbf{P}_y(\exists n \geq 1, s_n = s' \mid s_1 = s) = 1 \qquad \forall s \in C.$$

This is a property of the support of y. In particular, y can be chosen arbitrarily close to x. Let $y_{C,x,\varepsilon,s'}$ be such a perturbation that satisfies $\|y_{C,x,\varepsilon,s'} - x\| < \varepsilon$.

This definition captures the idea that the players can reach any state in C from any other state in C by slightly perturbing the stationary profile x.

We denote by $\mathcal{C}(x)$ the collection of all the sets that communicate under x. Whenever we say that a communicating set $C \in \mathcal{C}(x)$ is maximal, we mean maximal w.r.t. set inclusion. It is not difficult to check (and, since we do not use this property, left to the reader) that maximal communicating sets are disjoint.

3.4.2. *On Exits*

Let $x \in X$ be a stationary profile and let $C \in \mathscr{C}(x)$ be a communicating set under x.

DEFINITION 3.5. An exit from C (w.r.t. x) is a tuple $e = (s, x^{-L}, a^L)$ where $s \in C, \emptyset \neq L \subseteq N, a^L \in A^L$, and $q(C \mid s, x^{-L}, a^L) < 1$ while $q(C \mid s, x^{-L'}, a^{L'}) = 1$ for every strict subset L' of L.

Thus, an exit is a way that enables the players to leave a communicating set by slightly perturbing the profile x.

For simplicity, we sometimes write $e = (s, a^L)$ when no confusion may arise. The set of all exits from C w.r.t. x is denoted by $E(x, C)$. For $e = (s, x^{-L}, a^L) \in E(x, C), L(e) = L$ is the subset of players that need to perturb. If $L(e) = \{i\}$, we say that e is a *unilateral exit* of player i. Otherwise, e is a *joint exit*. We sometimes write q_e instead of q_{s,x^{-L},a^L}.

For any subset $C \subset S$ we define the *exit stage* from C by

$$e_C = \inf\{n \geq 1, s_n \notin C\}.$$

This is the first stage in which the game is not in C.

Every probability distribution over exits $\mu \in \Delta(E(x, C))$ defines a natural probability distribution $\tilde{\mu}$ over the states in $S \backslash C$:

$$\tilde{\mu}[s'] = \frac{\sum_{(s,x^{-L},a^L)\in E(x,C)} \mu(s, a^L) q(s' \mid s, x^{-L}, a^L)}{\sum_{(s,x^{-L},a^L)\in E(x,C)} \mu(s, a^L) q(S\backslash C \mid s, x^{-L}, a^L)} \qquad \forall s' \in S\backslash C. \quad (5)$$

A profile τ in G is an *ε-perturbation* of a stationary profile x if for every finite history $h \in H$, $\|\tau(h) - x\| < \varepsilon$.

The following simple lemma is an immediate consequence of the definition of a communicating set.

LEMMA 3.6. *Let $C \in \mathscr{C}(x)$ be a communicating set and let $\varepsilon > 0$. There exists a profile $\tilde{\tau} = \tilde{\tau}_{C,x,\varepsilon}$ in G such that*

1. *$\tilde{\tau}$ is an ε-perturbation of x.*
2. *For every exit $e = (s, a^L)$ and every $n \in \mathbf{N}$,*

$$\mathbf{P}_{\tilde{\tau}}(\exists m > n, s_m = s, \tilde{\tau}(h_m) = x_s \mid h_n) = 1.$$

Recall that h_m is the finite history up to stage m, and s_m is the last state of this history. In words, the second condition means that for every exit there are infinitely many stages where the play visits s and the players play the mixed action combination x_s.

Proof. Denote $J = |E(x, C)|$ and $E(x, C) = \{(s_1, a_1^{L_1}), \ldots, (s_J, a_J^{L_J})\}$. The profile $\tilde{\tau}$ is defined in rounds. At each round, the players play the following for *every* exit $(s_j, a_j^{L_j}) \in E(x, C)$:

(a) Play the stationary profile y_{C,x,ε,s_j} until the play reaches s_j.

(b) At s_j, play x_{s_j}.

It is easy to verify that $\tilde{\tau}$ satisfies the requirements. ∎

We use Lemma 3.6 to prove that any exit distribution from a communicating set can be induced by some strategy profile.

LEMMA 3.7. *Let a communicating set* $C \in \mathscr{C}(x)$, *let a probability distribution over exits* $\mu \in \Delta(E(x, C))$, *and let* $\varepsilon > 0$ *be given. There exists a profile* $\tau = \tau_{C, x, \varepsilon, \mu}$ *in* G *such that*

1. τ *is an* ε*-perturbation of* x.

2. $\mathbf{P}_{s, \tau}(s_{e_C} = s') = \tilde{\mu}[s']$ *for every* $s \in C$; *that is, the probability distribution of the first state out of* C *that the play visits coincides with* $\tilde{\mu}$, *where* $\tilde{\mu}$ *is given by* (5).

Proof. Denote $J = |E(x, C)|$ and $E(x, C) = \{(s_1, a_1^{L_1}), \ldots, (s_J, a_J^{L_J})\}$. For simplicity, set $\mu_j = \mu[(s_j, a_j^{L_j})]$. Let $\delta \in (0, \varepsilon)$ be sufficiently small, and let $\delta_1, \ldots, \delta_J$ satisfy for each j (i) $\delta_j \in (0, \delta)$ and (ii) $\delta_j = \delta_{j-1}/(1 - \delta_{j-1}\mu_{j-1})$.

Consider the strategy that was defined in the proof of Lemma 3.6, but replace (b) with:

(b′) At s_j, each player $i \notin L_j$ plays x^i, whereas each player $i \in L_j$ plays $(1 - \eta_j)x^i + \eta_j a_j^i$, where $\eta_j = (\delta_j \mu_j)^{1/|L_j|}$.

Thus, if, for example, $s_1 = s_2 = s$, the players follow a strategy that leads the game to $s_1 = s$, at s they use the exit $a_1^{L_1}$ with low probability, then they follow a strategy that leads the game to $s_2 = s$, where they use the exit $a_2^{L_2}$ with low probability, and so on.

If δ is sufficiently small then condition 1 holds.

Let p_j be the probability that exit $(s_j, a_j^{L_j})$ is the first exit to be played. If the players follow τ then the ratio p_j/p_{j-1} is

$$\frac{p_j}{p_{j-1}} = \frac{\left(1 - \eta_{j-1}^{|L_{j-1}|}\right)\eta_j^{|L_j|}}{\eta_{j-1}^{|L_{j-1}|}} = \frac{(1 - \delta_{j-1}\mu_{j-1})\delta_j\mu_j}{\delta_{j-1}\mu_{j-1}} = \frac{\mu_j}{\mu_{j-1}}.$$

It follows that condition 2 holds. ∎

Let $\mu \in \Delta(E(x, C))$ be a probability distribution over exits that is supported by unilateral exits, and let $i \in N$ be a player. Assume that $\mu[e] > 0$ for some unilateral exit e of player i, and let λ_i be the induced probability distribution over the unilateral exits of player i. By the construction of τ and $\tilde{\tau}$ we have

$$\tau_{C, x, \varepsilon, \lambda_i} = (\tilde{\tau}^{-i}_{C, x, \varepsilon}, \tau^{i}_{C, x, \varepsilon, \mu}). \tag{6}$$

Note that this equality holds since in the definition of $\tau^i_{C, x, \varepsilon, \mu}$ we listed *all* exits in $E(x, C)$, rather than only those in $\text{supp}(\mu)$.

Let x be a stationary profile and let $\prod = \{C_1, \dots, C_K, T\}$ be a partition into disjoint communicating sets under x and the set T of remaining transient states under x. For every $k = 1, \dots, K$ let $\mu_k \in \Delta(E(x, C_k))$. Recall that by (5) each μ_k defines a probability distribution $\tilde{\mu}_k$ over $S \backslash C_k$. The triplet $(x, \prod, (\mu_k)_k)$ defines a natural Markov chain over the state space S as follows. Transitions from states in some C_k are defined by $\tilde{\mu}_k$, whereas transitions from transient states $s \in T$ are defined by x_s. Formally,

$$p(s, s') =: \begin{cases} \tilde{\mu}_k[s'] & s \in C_k \quad \text{for some} \quad k = 1, \dots, K \\ q(s' \mid s, x) & s \in T \end{cases} \tag{7}$$

3.4.3. *The Sufficient Condition*

Recall that $S^\star$ is the subset of absorbing states.

PROPOSITION 3.8. *Let $\gamma \in \mathbf{R}^{N \times S}$ be a payoff vector and let $x \in X$ be a stationary profile. Let $\prod = (C_1, \dots, C_K, T)$ be a partition of $S \backslash S^\star$ such that $C_k \in \mathscr{C}(x)$ for every $k = 1, \dots, K$ and every $s \in T$ is transient w.r.t. x. Assume that for every $k = 1, \dots, K$ there exists a probability distribution μ_k over $E(x, C_k)$ such that the following conditions hold*:

1. *The Markov chain over S induced by $(x, \prod, (\mu_k)_k)$ is absorbing*; *that is*, *the probability that the Markov chain eventually reaches an absorbing state is* 1.

2. *For every state s and every player $i \in N$*,

   (a) $q_{s, x} \gamma^i = \gamma^i_s$.

   (b) *For every action $a^i \in A^i$, $q_{s, x^{-i}, a^i} \gamma^i \leq \gamma^i_s$.*

3. $\gamma^i_s = r^i(s)$ *for every state $s \in S^\star$ and every player $i \in N$.*

4. $\gamma^i_s \geq v^i_s$ *for every player i and every state $s \in S$.*

Moreover, *for every $k = 1, \dots, K$ we have*:

5. $\mu_k \gamma^i = \gamma^i_s$ *for every player i and every state $s \in C_k$.*

6. *At least one of the following holds*:

(a) *For every player i, if $e \in E(x, C_k)$ is a unilateral exit of player i with $\mu_k[e] > 0$, then $q_e \gamma^i = \gamma^i_s$ for every $s \in C_k$.*

(b) *For every exit $e \in E(x, C_k)$, if $\mu_k[e] > 0$ then e is a unilateral exit of some player.*

7. *If $e_1, e_2 \in E(x, C_k)$ are two unilateral exits of player i such that $\mu_k[e_1] > 0$, then $q_{e2}\gamma^i \leq q_{e_1}\gamma^i \leq \gamma^i_s$ for every $s \in C_k$.*

Then γ is a stationary correlated equilibrium payoff.

Let us first explain the main ideas of the proof and the role of each of the conditions.

The payoff vector γ is our desired correlated equilibrium payoff. We will construct a strategy τ where the players play x_s in any transient state $s \in T$, and the exit distribution from each communicating set C_k is $\tilde{\mu}_k$.

By condition 5 it follows that γ_s is constant in every set C_k. We denote this common value by γ_{C_k}.

Condition 1 will imply that some absorbing state is reached in finite time. Conditions 2(a) and 5 say that the sequence $(\gamma(s_n))$ is a martingale in the induced Markov chain, and condition 3 says that it coincides with the absorbing payoff in absorbing states; hence the expected payoff by τ is indeed γ.

Let us now check where players can profit by unilaterally deviating from τ and where the correlation device is required.

Condition 4 asserts that γ is above the min–max level. By condition 2(b) no player can increase his continuation payoff by deviating in transient states; hence in those states players do not need a correlation device.

Once the play enters a communicating set C_k, no player can increase his continuation payoff by playing a unilateral exit that is not in the support of μ_k. This fact holds for players who have unilateral exit in supp(μ_k) by condition 7 and for the other players by condition 2(b).

In a communicating set C_k that satisfies condition 6(a), players are indifferent to their unilateral exits that have positive probability under μ. In such a case, players do not need any correlation device to control the exit distribution from C_k. Indeed, one possible definition of τ is to let the players follow $\tau_{C_k, x, \varepsilon, \mu_k}$ until the play leaves C_k. By condition 6(a) no player can profit by deviating at stages in which he should use one of his unilateral exits. Moreover, the players can monitor the behavior of their opponents in stages in which joint exits should be used by using standard statistical tests. This idea was used by Vieille (2000c, Proposition 7) in the setup of two-player recursive games and by Solan (1999, Lemma 5.3) in the setup of n-player absorbing games. Both proofs extend to the present case.

We are left with communicating sets C_k that satisfy condition 6(b); that is, the exit distribution is supported by unilateral exits. This is the only case where correlation between the players is required. The role of the correlation device will be to choose one player and inform him that he should use one of his unilateral exits from C_k. The other players will receive a signal that tells them not to use any of their unilateral exits. Since the game is positive the chosen player cannot profit by not using any of his unilateral exits. By condition 7 all of his unilateral exits that have positive probability under μ_k yield him the same expected continuation payoff; hence he cannot profit by altering the probabilities in which he should use the different unilateral exits. By condition 7 the expected continuation payoff of any player i that was not chosen is at least $\gamma^i(C_k)$; hence he cannot increase his expected continuation payoff by using one of his unilateral exits.

Proof. Let $\varepsilon > 0$ be fixed. Let $K' \subset \{1, \ldots, K\}$ be the subset of indices k for which C_k satisfies condition 6(b). For every $k \in K'$ let α_k^i be the sum of $\mu_k[e]$ over all unilateral exits e of player i. By condition 6(b), $\sum_{i \in N} \alpha_k^i = 1$ for every $k \in K'$.

Define a stationary correlation device $\mathcal{D} = ((M^i)_{i \in N}, d)$ as follows. The signal space of player i is $M^i = \{0, 1\}^{K'}$; that is, for every set $k \in K'$, every player receives a signal either 0 or 1. The product signal space is $M = \prod_{i \in N} M^i = (\{0, 1\}^N)^{K'}$.

At every stage the device chooses for each $k \in K'$ one player, according to the probability distribution $\alpha_k = (\alpha_k^i)$. Each player i then receives a signal in M^i, where the kth coordinate is 1 if i was chosen for C_k and 0 otherwise. Formally, $d = \otimes_{k \in K'} d_k$ is a product distribution, where $d_k[l_i] = \alpha_k^i$ and $l_i \in M$ is defined by $l_i(j) = 1$ if and only if $i = j$.

Define a profile σ in the extended game $G(\mathcal{D})$ as follows.

- When the play is in a transient state $s \in T$, play x_s.
- When the play enters a communicating set C_k that satisfies condition 6(a), follow the profile $\tau_{C_k, x, \varepsilon, \mu_k}$ until the play leaves C_k.
- When the play enters (at stage n) a communicating set C_k that satisfies condition 6(b), denote by i the unique player that received the signal 1 for the set C_k at stage n. Player i plays the strategy $\tau^i_{C_k, x, \varepsilon, \mu_k}$, and each player $j \neq i$ follows the strategy $\tilde{\tau}_{C_k, x, \varepsilon}$. By (6), the strategy that is played is $\tau_{C_k, x, \varepsilon, \lambda_i^k}$, where λ_i^k is the restriction of μ_k over the unilateral exits of player i from C_k.

Note that the device was used *only* in sets C_k that satisfy condition 6(b), and only at stages where the play enters one of those sets.

Assume that the players follow σ in $G(\mathscr{D})$. By Lemma 3.7 the exit distribution from each set C_k that satisfies condition 6(a) is $\tilde{\mu}_k$, and by Eq. (6) the same holds when the set satisfies condition 6(b). By conditions 1, 2(a), 3, and 5, the expected payoff for the players is γ_s, where s is the initial state.

By condition 1 the corresponding Markov chain is absorbing. Hence the expected number of visits to transient states or communicating sets is finite. Therefore, to prove that condition 1 of Definition 2.2 holds, that is, that no player can profit too much by deviating along the play, it is sufficient to prove that no player can profit too much by deviating in any transient state or during one visit to some communicating set.

We are now going to add statistical tests and punishment profiles to deter players from deviating.

(a) If a player plays an action that is not compatible with the above profile, all players switch to an ε-punishment profile against that player. Simultaneous deviations of several players are ignored.

By conditions 2(b) and 4, no player can profit by deviating at transient states.

We now concentrate on communicating sets C_k that satisfy condition 6(a). Recall that the profile $\tau_{C_k, x, \varepsilon, \mu_k}$ that was defined in the proof of Lemma 3.7 uses the constants δ and $(\eta_j)_j$. In every round of the strategy $\tau_{C_k, x, \varepsilon, \mu_k}$, each exit is used with probability $O(\delta)$. By condition 6(a) no player can increase his expected continuation payoff by altering the probability in which he uses one of his unilateral exits. Thus, we assume that players do not deviate in such a way, and we have to deal only with joint exits.

For these joint exits, we add the following tests:

(b) For each exit $(s_j, a_j^{L_j})$, each player $i \in N$ is checked after many rounds that the distribution of his realized actions whenever the play was in s_j is ε-close to x_s^i.

(c) For each exit $(s_j, a_j^{L_j})$, each player $i \in L_j$ is checked whether he plays the action a_j^i with frequency η_j. Formally, after many visits to s_j, the realized probability p that player i plays a_j^i at those stages where the play visits s_j should satisfy $|\frac{p}{\eta_j} - 1| < \varepsilon$.

Two conditions have to be met for such tests to be both reliable and effective: to be reliable, the probability of false detection of deviation should be small, and to be effective, the probability that exit is used before the test is employed should be small even if the tested player deviates.

For test (b) to be reliable, it should be employed after a number of stages which is large compared to $1/x_s^i[a^i]$ for every $a^i \in \text{supp}(x_s^i)$. For test (b) to be effective, this number should be small compared to $1/\delta$ (since at every stage, exit occurs with probability δ). Thus, if δ is sufficiently small the test can be employed in a reliable and effective manner. For a detailed analysis of this test the reader can consult Vrieze and Thuijsman (1989, Case B, p. 302) or Solan (1999, Lemma 5.1).

To be reliable, test (c) should only be performed after a number of rounds which is large compared to $1/\eta_j$. Since the order of magnitude of η_j is $\delta^{1/|L_j|}$, test (c) should be employed only after stage $\delta^{-\nu}$, where $\nu > 1/|L_j|$. If δ is sufficiently small, test (c) is reliable. Since until stage $\delta^{-\nu}$ the expected number of times any single player i should play a^i_j is $\delta^{1/|L_j|-\nu}$, there exists $r > 0$ (independent of δ) such that the probability that an honest player i would play a^i_j until stage $\delta^{-\nu}$ more than $r\delta^{1/|L_j|-\nu}$ times is smaller than ε.

Thus, in test (c) player i is checked, in addition to the test mentioned above, whether the number of times he plays a^i_j until stage $\delta^{-\nu}$ is at most $r\delta^{1/|L_j|-\nu}$. This test is done only between the first and the $\delta^{-\nu}$th visit to s_j. The purpose of this additional test is to prevent player i from playing a^i_j too often prior to stage $\delta^{-\nu}$, in the hope that exit occurs. Therefore it must be that, even if player i chooses to play a^i_j in $r\delta^{1/|L_j|-\nu}$ stages, the probability that exit occurs prior to stage $\delta^{-\nu}$ is small. Whenever player i plays a^i_j, the probability that exit occurs is of the order $\delta^{1-1/|L_j|}$. Thus, it is enough to choose $\nu < 1$.

Finally, we turn our attention to communicating sets C_k that satisfy condition 6(b). Note that in such a set, if $\alpha^i_k > 0$ for at least three players, then no player except the chosen one knows who should use his unilateral exits.

Here only tests (a) and (b) are needed. Let i be the player who should use one of his unilateral exits. By condition 7 he is indifferent between his unilateral exits; hence he cannot profit by altering the probabilities with which he uses any of his unilateral exits. Moreover, he cannot profit by using a unilateral exit that has probability 0 under μ. Since the game is positive recursive, $\gamma_s \geq 0$, and 0 is the payoff player i will get if he does not use any of his unilateral exits. Thus player i cannot profit by deviating.

Consider now a player $j \neq i$. This player does not know which player got the signal 1. All he knows is that unless he deviates, none of *his* unilateral exits will be used. Since in those exits he receives by condition 7 at most $\gamma^j_{C_k}$, his expected payoff conditioned on receiving the signal 0 is at least $\gamma^j_{C_k}$. If player j deviates and uses one of his unilateral exits, his deviation will not be detected by anyone except player i. In particular, player j will not be punished (provided there are at least three players). The probability that players i and j will use one of their unilateral exits at the same stage is small; hence, by condition 7, the expected payoff of player j is at most $\gamma^j_{C_k} + \varepsilon$. ∎

4. GENERAL STOCHASTIC GAMES

In this section we prove Theorem 2.3. By Theorem 3.4 it is sufficient, given $\varepsilon > 0$, to find a vector $\gamma = (\gamma(h))_{h\in H}$ of payoff vectors and a correlated profile $\tilde{\tau}$ in G that satisfy:

- Average payoffs under $\tilde{\tau}$ converge to γ.
- $\tilde{\tau}$ is ε-individually rational with respect to γ.

Using the method of Mertens and Neyman (1981) we construct for every $\varepsilon > 0$ a profile τ in G that is ε-individually rational. Unfortunately, we do not know whether average payoffs under τ converge. Using τ we define a correlated profile $\tilde{\tau}$ that is ε-individual rational and in some sense cyclic. The cyclic nature of $\tilde{\tau}$ will ensure that average payoffs under $\tilde{\tau}$ converge.

4.1. *A Mertens–Neyman Profile*

The result of Mertens and Neyman (1981) can be summarized as follows. Let $g : (0,1) \times S \to \mathbf{R}$ be a function with bounded variation,[5] let $\varepsilon' > 0$ and let $i \in N$ be a player. Since g has bounded variation, the limit $g_0 = \lim_{\lambda\to 0} g_\lambda$ exists.

Recall that s_h is the last state of the history h. Mertens and Neyman (1981) constructed for every $n \in \mathbf{N}$ and $\alpha > 0$ a function λ_n^i: $H \to (0, \alpha)$ that depends both on g and on ε'. Assume that there is a profile τ in G that satisfies for *every* finite history h of length n,[6]

$$g_{\lambda_n^i(h)}(s_h) \leq \mathbf{E}_{s_h, \tau(h)}\Bigg(\lambda_n^i(h) r^i(s_h, a) + (1 - \lambda_n^i(h)) \sum_{s'\in S} q(s' \mid s_h, a) g_{\lambda_n^i(h)}(s')\Bigg). \tag{8}$$

Recall that s_h is the last state of the history h. Mertens and Neyman then prove that the following hold when the players follow τ:

(MN.1) λ_n^i converges to 0 with probability 1 (this fact will not be used here).

(MN.2) For every finite history $h \in H$ of length m and every $n \geq m$, $\mathbf{E}_{h,\tau}[g_0(s_n)] \geq g_0(s_h) - \varepsilon'$.

(MN.3) There exists $n_0 \in \mathbf{N}$ such that for every finite history $h \in H$,

$$\gamma_n^i(h, \tau) \geq g_0(s_h) - \varepsilon' \qquad \forall\, n \geq n_0.$$

[5]That is, for every $s \in S$ the function $g_\lambda(s)$, as a function of λ, has bounded variation.
[6]Here, the expectation is with respect to a, the choices made after history h.

Intuitively, for every discount factor λ and every state, the function g assigns some desired level for player i. If the profile τ guarantees that at any stage n the λ_n^i-level does not fall on average, then the expected payoff of player i is guaranteed to be at least $g_0 - \varepsilon'$.

Recall that $(v_\lambda^i(s))_{s\in S}$ is the λ-discounted min–max value of player i.

For every state $s \in S$ and every vector of discount factors $\vec{\lambda} = (\lambda^i)_{i\in N}$, let $G(s, \vec{\lambda})$ be the one-shot game with (i) player set N, (ii) the action set of each player i is A^i, and (iii) the payoff function is

$$\lambda^i r^i(s, a) + (1 - \lambda^i) \sum_{s'\in S} q(s' \mid s, a) v_{\lambda^i}^i(s').$$

Note that in an equilibrium of $G(s, \vec{\lambda})$ each player receives at least $v_{\lambda^i}^i(s)$. Indeed, let x be such an equilibrium. The meaning of $v_{\lambda^i}^i(s)$ is that if his continuation payoff is given by $v_{\lambda^i}^i$, then for every mixed action combination of his opponents player i has a reply that guarantees him an expected payoff of $v_{\lambda^i}^i(s)$. In particular he can guarantee $v_{\lambda^i}^i(s)$ against x^{-i}.

For every state $s \in S$ and every vector of discount factors $\vec{\lambda} \in (0, 1)^N$, let $x(s, \vec{\lambda})$ be an equilibrium of $G(s, \vec{\lambda})$.

Fix $\varepsilon' > 0$. Let $\alpha \in (0, 1)$ be sufficiently small so that $|v_\lambda^i(s) - v^i(s)| < \varepsilon'$ for every $\lambda \in (0, \alpha)$. For every player $i \in N$ and every $n \in \mathbf{N}$, let $\lambda_n^i\colon H \to (0, 1)$ be the function devised by Mertens and Neyman w.r.t ε', α, and $g_\lambda(s) = v_\lambda^i(s)$.

Define a non-correlated profile τ in G by

$$\tau(h) = x(s_h, (\lambda_n^i(h))_{i\in N}) \tag{9}$$

for every history h of length n. That is, at every stage the players play an equilibrium in the one-shot game $G(s, \vec{\lambda})$, where each player calculates his discount factor separately.

By construction we have

$$v_{\lambda_n^i(h)}^i(s_n) \le \mathbf{E}_{x(s_n,\vec{\lambda}_n(h))}\Big(\lambda_n^i(h) r^i(s_n, a_n) + (1 - \lambda_n^i(h)) v_{\lambda_n^i}^i(s_{n+1})\Big), \tag{10}$$

which is the basic equation (8) needed by Mertens and Neyman.

By (MN.2), (MN.3), and the choice of α, it follows that the profile τ is $3\varepsilon'$-individually rational w.r.t. v.[7] The main problem with this definition of τ is that we do not know whether average payoffs converge under τ.

Remark. Our definition of a correlated equilibrium payoff is a uniform one. Another possible definition is using the limsup of the daily payoffs; that is, a payoff vector $\gamma \in \mathbf{R}^{N\times S}$ is an autonomous correlated equilibrium payoff in the *limsup sense* if for every $\varepsilon > 0$ there exists an autonomous

[7]That is, the function that assigns to each history h the payoff vector $v(s_h)$.

correlation device $\mathcal{D}$ and a strategy profile σ in $G(\mathcal{D})$ such that for every initial state s, every player $i \in N$, and every strategy σ'^i of player i,

$$\gamma^i + \varepsilon \geq \gamma^i(\mathcal{D}, s, \sigma) \geq \gamma^i - \varepsilon \geq \gamma^i(\mathcal{D}, s, \sigma^{-i}, \sigma'^i) - 2\varepsilon, \tag{11}$$

where

$$\gamma^i(\mathcal{D}, s, \sigma) = \mathbf{E}_{\mathcal{D},s,\sigma}\left(\limsup_{n\to\infty} \frac{1}{n}(r^i(s_1, a_1) + \cdots + r^i(s_n, a_n))\right).$$

Let τ be the strategy profile defined in (9), let $\mathcal{D} = \mathcal{D}^2(\tau)$ be the device defined in Section 3.3.1, and let σ be the strategy profile in $G(\mathcal{D})$ that follows the recommendation of the device as long as no deviation is detected and punishes a deviator by his min–max level. Using (MN.2) and (MN.3), one can show that (11) holds.

Thus, the construction we presented here provides a simple proof for the existence of an autonomous correlated equilibrium payoff in the limsup sense.

4.2. *Modifying the Profile*

The conclusion is that the profile τ needs to be modified, so that average payoffs under τ will converge. The goal of this section is to define a modification of τ, namely a correlated profile $\tilde{\tau}$, that is individually rational and in which average payoffs under $\tilde{\tau}$ do converge.

One way to ensure that average payoffs converge is to have the profile cyclic.

A naive way to modify τ in a cyclic manner is to play τ by blocks of some length N_0, namely, to forget the past history at the end of each block and to resume playing according to τ, as if the game was starting anew. Observe that the sequence of the *initial* states of the *successive* blocks follows a Markov chain. Average payoffs therefore converge, but it is not clear whether this profile is ε-individually rational. The reason for that is the following. The min–max value of the initial state may decrease a little from block to block, in expected terms, by at most ε'. Since the past history is forgotten at the beginning of each block, errors may accumulate. Since Mertens and Neyman's result says only that the average payoff over any block is at least the min–max value of the first state in the block (up to ε', provided N_0 is large), we get no estimate on the average payoff received during course of the play.

The solution is to divide the state space into three disjoint subsets.

1. States s such that there exists a mixed action y that satisfies the following. If in state s the players play y, then (i) the expected continuation min–max value does not decrease, and (ii) there is a positive probability that the min–max value changes for at least one player. This set of states is denoted by $\overline{S}$.

2. States s where the probability to reach under τ during the first N_0 stages a state in $\overline{S}$ or to reach a state with a different min–max value is small. N_0 is sufficiently large to satisfy some conditions, including (MN.3). This set of states is denoted by S_1.

3. The set of all remaining states, which is denoted by S_2.

We define $\tilde{\tau}$ as follows:

(a) Whenever the play visits a state $s \in \overline{S}$, the players play a mixed action y that satisfies the condition (1) of the previous paragraph.

(b) Whenever the play reaches a state $s \in S_2$, the players play for N_0 stages as they played in τ.

(c) Whenever the play visits a state $s \in S_1$, the players play for N_0 stages *almost* as they played under τ. The modification is done to make the probability of reaching a state in $\overline{S}$ or of reaching a state with a different min–max value in the first N_0 stages vanish.

The idea behind this definition is the following. The fact that $\tilde{\tau}$ and τ are close has two implications: under $\tilde{\tau}$, (i) $v(s_n)$ would be almost a submartingale, and it will converge $\tilde{\tau}$-a.s., and (ii) the expected average payoff in the first N_0 stages is high (at least the min–max value). The first implication will be used to show that the min–max value changes only finitely many times along the play. One can then bound the number of visits to states in $\overline{S}$ and S_2. In particular, from some point on, the play remains in S_1, and the min–max value remains constant. The second implication will be used to show that once the play remains in S_1, the expected average payoff is high (at least the min–max value).

Thus, each state s in S_1 has the following property: there exists a stationary correlated profile τ' such that, provided the initial state is s, under τ', (i) the min–max level remains fixed, and (ii) the average expected payoff remains high. In particular, states in S_1 are easily solved: provided the initial state is in S_1, one can construct a simple autonomous correlation device such that following the recommendation of the device is ε-optimal.

Two final points before we continue the formal proof. First, the set $\overline{S}$ will be larger than the one defined above. Second, to make sure that $\tilde{\tau}$ is ε-individually rational, given the action chosen for i at state s, the conditional distribution over A^{-i} should be close to $x^{-i}(s, \vec{\lambda})$, for some $\vec{\lambda}$ sufficiently small.

4.2.1. *Partitioning the State Space*

We start by identifying the set $\overline{S}$.

For every state $s \in S$ define $X^*(s)$ to be the set of all accumulation points of $\{x(s, \vec{\lambda})\}$:

$$X^*(s) = \left\{ y \in \Delta(A) \mid \exists(\vec{\lambda}_n) \to 0 \text{ s.t. } y = \lim_{n\to\infty} x(s, \vec{\lambda}_n) \right\}.$$

Since the action sets are finite, $X^*(s)$ is non-empty and compact.

Since $x(s, \vec{\lambda})$ is an equilibrium in the one-shot game $G(s, \vec{\lambda})$, it follows that

$$q_{s,y} v^i \geq q_{s,y^{-i},a^i} v^i \qquad \forall\, s \in S, y \in X^*(s), i \in N, a^i \in A^i. \tag{12}$$

In particular

$$q_{s,y} v^i \geq v^i(s) \qquad \forall\, s \in S, y \in X^*(s), i \in N. \tag{13}$$

We denote by $C_s = \{s' \in S \text{ s.t. } v(s) = v(s')\}$ the set of states where the min–max value is equal to the min–max value at s (for all players).

Define

$$\overline{S}_1 = \{s \in S \mid \exists y \in X^*(s), \text{ s.t. } q(S \setminus C_s \mid s, y) > 0\}.$$

Those are all the states where, by playing some mixed action in $X^*(s)$, the min–max level changes with positive probability.

We can now define iteratively

$$\overline{S}_{n+1} = \{s \in S \mid \exists y \in X^*(s), \text{ s.t. } q((S \setminus C_s) \cup \overline{S}_n \mid s, y) > 0.$$

Clearly $\overline{S}_{n+1} \supseteq \overline{S}_n$, and since the state space is finite, there exists some $n \leq |S|$ such that $\overline{S}_{n+1} = \overline{S}_n$.[8] Define $\overline{S} = \cup_{n=1}^{|S|} \overline{S}_n$.

Thus, from any state $s \in \overline{S}$ there is a positive probability that the min–max level will change during the first $|S|$ stages after the play visits s, while the expected min–max level does not drop. By appropriately defining $\tilde{\tau}$ in $\overline{S}$ we will be able to bound the expected number of visits to those states.

From now on we fix $\varepsilon > 0$ sufficiently small. Since $X^*(s)$ was defined as the set of all accumulation points of sequences $x(s, \vec{\lambda})$ as $\vec{\lambda}$ goes to 0, there exists $\alpha_0 > 0$ such that if $\vec{\lambda} \in (0, \alpha_0)^N$ then $d(x(s, \vec{\lambda}), X^*(s)) < \varepsilon$ ($d(\cdot, \cdot)$ is the supremum distance).

By construction, for every $s \notin \overline{S}$

$$q((S \setminus C_s) \cup \overline{S} \mid s, y) = 0 \qquad \forall\, y \in X^*(s) \tag{14}$$

[8] Actually, by (13), for any selection of X^* there is an ergodic set in the corresponding Markov chain where v is constant; hence this n must be strictly smaller than $|S|$.

and

$$q((S \backslash C_s) \cup \overline{S} \mid s, x(s, \vec{\lambda})) \leq \varepsilon \qquad \forall \vec{\lambda} \in (0, \alpha_0)^N, \tag{15}$$

provided that α_0 is sufficiently close to zero.

We now introduce the sets S_1 and S_2 that were mentioned in the above presentation. Take $\varepsilon' = \varepsilon^{(|N|+1) \times S+1}$, and denote by τ the corresponding Mertens–Neyman's profile with $\alpha = \alpha_0$. By (MN.2) and (MN.3) there exists an integer N_0 such that, for every state $s \in S$,

$$\gamma_{N_0}(s, \tau) \geq v(s) - \varepsilon \quad \text{and} \quad \mathbf{E}_{s,\tau}[v(s_{N_0+1})] \geq v(s) - \varepsilon^{(|N|+1) \times S+1}. \tag{16}$$

DEFINITION 4.1. A state $s \notin \overline{S}$ is *good* if $\mathbf{P}_{s,\tau}(\exists\, n \leq N_0 + 1, s_n \in (S \backslash C_s) \cup \overline{S}) \leq \varepsilon$.

A state s is good if, starting from s, there is a small probability that under τ the min–max level changes or the play reaches a state in $\overline{S}$ in the first N_0 stages.

The set of good states is denoted by S_1 and we set $S_2 = S \backslash (S_1 \cup \overline{S})$.

4.2.2. *The Correlated Distance*

The definition of individual rationality requires that given the action chosen for him, no player can gain by a deviation that is followed by punishment with his min–max level. To fulfill this requirement, the correlated strategy $\tilde{\tau}$ that we are going to define has to be close in some correlated sense to the set X^*. We define this closeness concept now.

Recall that for every correlated probability distribution $y \in \Delta(A)$ and every $a = (a^{-i}, a^i) \in \text{supp}(y)$, $(y \mid a^i)$ is the conditional probability over A^{-i}, given that the action chosen for i is a^i.

DEFINITION 4.2. Let $x, y \in \Delta(A)$ with $\text{supp}(y) \subseteq \text{supp}(x)$. The *correlated distance* between y and x is defined by

$$d_c(y; x) = \max_{i \in N,\, a=(a^{-i}, a^i) \in \text{supp}(y)} \| (y \mid a^i) - (x \mid a^i) \|_\infty .$$

This definition captures the difference in information players have when they know the action chosen for them under y, relative to the same action being chosen under x.

Note that if $\text{supp}(y) \subset \text{supp}(x)$, then $d_c(x; y)$ is not defined. However, if $\text{supp}(y) = \text{supp}(x)$ then $d_c(y; x) = d_c(x; y)$. Moreover, $d_c(z; y) + d_c(y; x) \geq d_c(z; x)$ whenever the two correlated distances on the left-hand side are defined.

Fix a state $s \in S_1$, and call an action combination *a bad* if $q((S \backslash C_s) \cup \overline{S} \mid s, a) > 0$. Those are actions that should not be played in good states. We now define for every mixed action y that gives low weight to bad actions a correlated mixed action $\tilde{y}$ that (i) is close to y in the correlated distance and (ii) gives probability 0 to bad actions.

A naive way to do it would be to normalize y over the non-bad actions. As the following example shows, this may lead to a large correlated distance; hence we may need to eliminate more actions.

Consider a 2×2 matrix where player 1 is the row player, player 2 is the column player, actions of player 1 are $\{T, B\}$, and actions of player 2 are $\{L, R\}$. Let y be the mixed profile where player 1 plays T with probability $1 - \varepsilon$, player 2 plays L with probability $1 - \varepsilon$, and the two play independently. If the action combination (B, R) is the only bad action, define $\tilde{y} = (1-\varepsilon)^2/(1-\varepsilon^2)(T, L) + (1-\varepsilon)\varepsilon/(1-\varepsilon^2)(T, R) + (1-\varepsilon)\varepsilon/(1-\varepsilon^2)(B, L)$. One can verify that $d_c(\tilde{y}, y) = \| (\tilde{y} \mid B) - (y \mid B) \| = \varepsilon$. However, if the action combination (B, L) is the only bad action and we define $\tilde{y} = (1-\varepsilon)^2/(1-\varepsilon+\varepsilon^2)(T, L) + (1-\varepsilon)\varepsilon/(1-\varepsilon+\varepsilon^2)(T, R) + \varepsilon^2/(1-\varepsilon+\varepsilon^2)(B, R)$ then $d_c(\tilde{y}, y) = \| (\tilde{y} \mid B) - (y \mid B) \| = 1 - \varepsilon$. Thus, one has to eliminate also (B, R) and define $\tilde{y} = (1-\varepsilon)(T, L) + \varepsilon(T, R)$.

Let $s \in S$ be a state. Define

$$B_1 = \{a \in A \text{ such that } q((S \backslash C_s) \cup \overline{S} \mid s, a) > 0\}$$

to be the set of all bad actions. Define recursively

$$B_{n+1} = \{a \in A, \text{ such that } y[a] \leq y[b]/\varepsilon^{1/(2|A|)}, \text{ for some } b \in B_n\}.$$

Clearly, the sequence (B_n) is non-decreasing and hence eventually stationary. Denote by B_∞ the limit and observe that $B_{|A|} = B_\infty$. In particular,

$$y[b] \leq (1/\varepsilon^{1/2}) \max_{a \in B_1} y[a] \tag{17}$$

for any $b \in B_\infty$.

We now define for every mixed action profile y a correlated probability distribution $\tilde{y}$ that is close to y in the correlated distance and gives probability 0 to bad actions.

Define

$$y'[a] = \begin{cases} 0 & a \in B_\infty \\ y[a] & \text{otherwise} \end{cases}$$

Finally, $\tilde{y}$ is the normalization of y' if $y'[A] > 0$ and arbitrary otherwise. Let $\zeta = \min\{q(s' \mid s, a) \mid s', s \in S, a \in A, q(s' \mid s, a) > 0\}$ be the minimal positive transition in the game.

LEMMA 4.3. *For every state* $s \in S_1$, *every* $y \in \prod_{i \in N} \Delta(A^i)$ *such that* $q((S \backslash C_s) \cup \overline{S} \mid s, y) \leq \varepsilon$, *and every* ε *sufficiently small the following hold.*

(i) $y'[A] > 0$, *so that the normalization is well defined.*

(ii) $d_c(\tilde{y}; y) \le |A|\varepsilon^{1/(2|A|)}$.

(iii) $q((S\backslash C_s) \cup \overline{S} \mid s, \tilde{y}) = 0$.

Proof. To prove (i), note that the cardinality of B_∞ is at most $|A|$. Since $q((S\backslash C_s) \cup \overline{S} \mid s, y) \le \varepsilon$, one has $y[b] \le \varepsilon/\zeta$, for every $b \in B_1$. By (17), $y[B_\infty] \le |A|\sqrt{\varepsilon}/\zeta$, and (i) follows.

Let us now show that (ii) holds as well. Fix a player $i \in N$ and an action combination $a \in \text{supp}(\tilde{y})$. Since $a \notin B_\infty$, one has $y[a] > y[b]/\varepsilon^{1/(2|A|)}$, for every $b \in B_\infty$.

In particular, by removing any $b \in B_\infty$ from y, the conditional probability given a^i does not change by more than $\varepsilon^{1/(2|A|)}$. Since we remove at most $|A|$ actions, $d_c(\tilde{y}; y) \le |A|\varepsilon^{1/(2|A|)}$.

Finally, (iii) is immediate from the definitions of B_1 and $\tilde{y}$. ■

4.2.3. *Definition of* $\tilde{\tau}$

Let s_1 be a good state, and set $C = C_{s_1} \cap (S\backslash\overline{S})$. We are now going to assign, for every history of length at most N_0 that starts in S_1, a correlated probability distribution $\tilde{\tau}_1(h)$ in $\Delta(A)$. We then show that under $\tilde{\tau}_1$ (i) play remains in C, (ii) the correlated distance of $\tilde{\tau}_1(h)$ from some $x(s, \vec{\lambda})$ is small, and (iii) the average payoff in the first N_0 stages under $\tilde{\tau}_1$ is at least $v(s_1) - \sqrt{\varepsilon}$.

Let h be a finite history of length at most N_0 that starts in s_1 and visits only states in C. Define $\tilde{\tau}_1(h) = \tilde{y}$, where $y = \tau(h)$.

For any other history h of length at most N_0 let $\tilde{\tau}_1(h) = \tau(h)$.

Let $s \in \overline{S}$. Let n be the minimal integer such that $s \in \overline{S}_{n+1}$. Let $y(s) \in X^*(s)$ satisfy $q((S\backslash C_s) \cup \overline{S}_n \mid s, y(s)) > 0$, where by convention $\overline{S}_0 = \emptyset$.

We are now ready to define the correlated profile $\tilde{\tau}$. The profile is defined in blocks of different length—length 1 for states in $\overline{S}$ and length N_0 for all other states. Let s be the initial state of the current block. Then, regardless of the past play

- If $s \in \overline{S}$, the players play $y(s)$.
- If $s \in S_1$, the players follow $\tilde{\tau}_1$ for N_0 stages.
- If $s \in S_2$, the players follow τ for N_0 stages.

As we show below, the correlated profile $\tilde{\tau}$ satisfies the conditions of Theorem 3.4.

4.2.4. *Analysis of* $\tilde{\tau}$

We first prove that if the play starts at S_1, the average payoffs up to stage N_0 under τ and $\tilde{\tau}$ are close.

LEMMA 4.4. *For every $s_1 \in S_1$ one has* $\|\gamma_{N_0}(s_1, \tilde{\tau}) - \gamma_{N_0}(s_1, \tau)\| < |A|\sqrt{\varepsilon}/\zeta$.

Proof. Since s_1 is good, the probability that under τ the play leaves C in the first N_0 stages is at most ε. Thus, the probability that an action combination $a \in B_1$ is ever played before stage N_0 is at most ε/ζ. Therefore, the probability that under τ an action in B_∞ is ever played before stage N_0 is at most $|A|\sqrt{\varepsilon}/\zeta$.

Since the probability induced by τ over histories that never use bad actions coincides with the probability induced by $\tilde{\tau}$ over H_{N_0}, the set of histories of length N_0, and since payoffs are bounded by 1, the result follows. ■

By construction, if $s_1 \in S_1$ then $\mathbf{P}_{s_1,\tilde{\tau}}(s_{N_0+1} \in C_{s_1} \cap (S \setminus \overline{S})) = 1$. Lemma 4.4 and the choice of N_0 shows that

$$\gamma_{N_0}(s_1, \tilde{\tau}) \geq v(s_1) - |A|\sqrt{\varepsilon}/\zeta - \varepsilon > v(s_1) - 2|A|\sqrt{\varepsilon}/\zeta \qquad \forall s_1 \in S_1. \quad (18)$$

We now bound the number of visits to states in S_2. Recall that from every $s \in \overline{S}$ there is a positive probability that if the players follow $\tilde{\tau}$ the play leaves $\overline{S}$ in the first $|S|$ stages. Let ω_1 be a positive lower bound of this probability for $s \in \overline{S}$.

Define a transition function over the state space S by

$$p(s, s') = \mathbf{P}_{s,\tilde{\tau}}(s_{N_0+1} = s') \qquad \text{if } s \notin \overline{S}$$
$$p(s, s') = q(s'|s, y(s)) \qquad \text{if } s \in \overline{S}$$

Fix an initial state s, and denote for simplicity by (s_n) the induced Markov chain over S. We denote by $\mathbf{E}_p$ the corresponding expectation. By construction,

$$\text{If } s_n \in \overline{S} : \mathbf{E}_p[v(s_{n+1})|s_n] \geq v(s_n)$$
$$\text{and } p(v(S_{n+m}) \neq v(s_n)) \geq \omega_1 \text{ for some } m \leq |S|. \quad (19)$$
$$\text{If } s_n \in S_1 : v(s_{n+1}) = v(s_n),\ p\text{-a.s.} \quad (20)$$
$$\text{If } s_n \in S_2 : \mathbf{E}_p[v(s_{n+1})|s_n] \geq v(s_n) - \varepsilon^{(|N|+1)\times|S|+1}$$
$$\text{and } p(v(s_{n+1}) \neq v(s_n)) \geq \varepsilon. \quad (21)$$

Indeed, (19) follows from the definition of $\overline{S}$, (20) holds by Lemma 4.3, and (21) holds by (16) and Definition 4.1.

Let $\rho = \min\{|v^i(s) - v^i(s')|, i \in N, s \in S, v^i(s) \neq v^i(s')\}$.

LEMMA 4.5. *Consider the Markov chain induced by p. If ε is sufficiently small then*

1. *All ergodic sets are subsets of S_1.*
2. *$v(s_n)$ converges p-a.s. to some function v_∞.*

3. *The expected number of visits to states in* S_2 *and* $\overline{S}$ *is at most* $\varepsilon^{-(|N|+1)\times|S|}$.

4. $\mathbf{E}_p[v_\infty] \geq v(s_1) - \varepsilon$.

Proof. Note that (1) follows from (3), since by (3) no state in $S_2 \cup \overline{S}$ can be in any ergodic subset.

Therefore, by (20), $v(s)$ is constant in every ergodic subset. In particular, (2) follows.

Define the number of visits to $S_2 \cup \overline{S}$ by

$$\widetilde{N} = |\{n \geq 1, s_n \in S_2 \cup \overline{S}\}|.$$

By (19), (20), and (21) we have

$$\mathbf{E}_p[v_\infty] \geq v(s_1) - \varepsilon^{(|N|+1)\times S+1}\mathbf{E}_p\big[\widetilde{N}\big]. \tag{22}$$

Therefore (4) follows from (3).

Thus, to prove the lemma it is enough to prove (3). We proceed to get an upper bound on $\mathbf{E}_p\big[\widetilde{N}\big]$.

For every state s assign the vector $V(s) = (v^1(s), \ldots, v^{|N|}(s)) \in [0,1]^{|N|}$. Define $s_1 \succ s_2$ if and only if $V(s_1) > V(s_2)$ in the lexicographic order, $s_1 \succeq s_2$ if and only if $V(s_1) \geq V(s_2)$ in the lexicographic order, and $s_1 \sim s_2$ if and only if $V(s_1) = V(s_2)$.

By (20), if $s_n \in S_1$ then $s_{n+1} \sim s_n$.

Eq. (19) implies that if $s_n \in \overline{S}$ then

$$\mathbf{P}(s_{n+m} \succ s_n) \geq \omega_1 \rho/(1+\rho) > \varepsilon \qquad \text{for some } m \leq |S|. \tag{23}$$

Our next goal is to show that if $s_n \in S_2$ then

$$\mathbf{P}(s_{n+1} \succ s_n) > \varepsilon^{|N|+1}. \tag{24}$$

Let i be the player with minimal index such that $p(v^i(s_{n+1}) \neq v^i(s_n)) \geq \varepsilon^{|N|+1-i}/2$. By (21), $p(v(s_{n+1}) \neq v(s_n)) \geq \varepsilon$; hence such i exists. In particular,

$$\begin{aligned} p(v^i(s_{n+1}) > v^i(s_n)) &= p(v^i(s_{n+1}) \neq v^i(s_n)) - p(v^i(s_{n+1}) < v^i(s_n)) \\ &\geq \varepsilon^{|N|-1-i}/2 - p(v^i(s_{n+1}) < v^i(s_n)). \end{aligned} \tag{25}$$

Now, by (21), since payoffs are bounded by 1, and by the definition of ρ,

$$\begin{aligned} v^i(s_n) - \varepsilon^{(|N|+1)\times S+1} &\leq \mathbf{E}_p(v^i(s_{n+1}) \mid s_n) \\ &= v^i(s_n)p(v^i(s_{n+1}) = v^i(s_n)) + v^i(s_{n+1})p(v^i(s_{n+1}) \\ &\quad > v^i(s_n)) + v^i(s_{n+1})p(v^i(s_{n+1}) < v^i(s_n)) \\ &\leq v^i(s_n)p(v^i(s_{n+1}) = v^i(s_n)) + p(v^i(s_{n+1}) > v^i(s_n)) \\ &\quad + (v^i(s_n) - \rho)p(v^i(s_{n+1}) < v^i(s_n)). \end{aligned}$$

Using (25) one deduces that

$$\begin{aligned} p(v^i(s_{n+1}) > v^i(s_n)) &\geq v^i(s_n)p(v^i(s_{n+1}) > v^i(s_n)) - \varepsilon^{(|N|+1)\times S+1} \\ &\quad + \rho p(v^i(s_{n+1}) < v^i(s_n)) \\ &\geq v^i(s_n)p(v^i(s_{n+1}) > v^i(s_n)) - \varepsilon^{(|N|+1)\times S+1} \\ &\quad + \rho(\varepsilon^{|N|+1-i}/2 - p(v^i(s_{n+1}) > v^i(s_n)). \end{aligned}$$

Using $v^i(s_n) \geq 0$ and $\rho \leq 1$ we get that

$$p(v^i(s_{n+1}) > v^i(s_n)) \geq \frac{\rho\varepsilon^{|N|+1-i}/2 - \varepsilon^{(|N|+1)\times S+1}}{1 - v^i(s_n) + \rho} \geq \rho\varepsilon^{|N|+1-i}/5.$$

Since for every player $j < i$, $p(v^j(s_{n+1}) \neq v^j(s_n)) < \varepsilon^{|N|+1-j}/2$, it follows that

$$\begin{aligned} p(V(s_{n+1}) > V(s_n)) &> \rho\varepsilon^{|N|+1-i}/5 - \sum_{j=1}^{i-1} \varepsilon^{|N|+1-j}/2 \\ &> \rho\varepsilon^{|N|+1-i}/6 > \varepsilon^{|N|+1}, \end{aligned}$$

and (24) follows.

By (23) and (24) and since $|S_2 \cup \overline{S}| \leq |S| - 1$ it follows that with probability at least $\varepsilon^{(|N|+1)\times|S|}$, the states $s_n, s_{n+1}, \ldots, s_{n+|S|^2}$ include a state in S_1.[9] Since $s_{m+1} \sim s_m$ whenever $s_m \in S_1$, (3) follows. ∎

We may now conclude the proof of Theorem 2.3. Since $\tilde{\tau}$ is defined by blocks, average payoffs under $\tilde{\tau}$ do converge uniformly over h. Denote the limit by $\gamma = (\gamma(h))_{h\in H}$.

By Lemma 4.5(1), any ergodic set w.r.t. p is a subset of S_1, and therefore $v(s)$ is constant in any such set. Fix a finite history $h \in H$. By (18), if s_h is in an ergodic set under p, then $\gamma(h) \geq v(s_h) - 2|A|\sqrt{\varepsilon}/\zeta$. By Lemma 4.5(4), $\gamma(h) \geq v(s) - 2|A|\sqrt{\varepsilon}/\zeta - \varepsilon$ for every $s \in S$.

Finally, for any history h, the correlated distance between $\tilde{\tau}(h)$ and $\tau(h)$ is at most $|A|\varepsilon^{1/(2|A|)}$. Since $\tau(h) = x(s, \vec{\lambda})$ for some $\vec{\lambda} \in (0, \alpha_0)^N$, it follows by (12) that $\tilde{\tau}$ is $2|A|\varepsilon^{1/(2|A|)}$-individually rational with respect to γ. This concludes the proof of Theorem 2.3.

Remark. For simplicity, one can change the definition of $\tilde{y}$ so that $\tilde{y}$ is a non-correlated mixed action that satisfies Lemma 4.3. By a limit argument, one can show the existence of a solvable set in the sense of Vieille (1998). Thus, our proof is an alternative proof for the existence of solvable sets in N-player stochastic games.

[9] Actually, using the definition of $\overline{S}$ instead of (23) one can show that with probability at least $\varepsilon^{(|N|+1)\times|S|}$ one of the states $s_n, s_{n+1}, \ldots, s_{n+|S|-1}$ is in S_1.

5. POSITIVE RECURSIVE GAMES

In this section we restrict ourselves to positive recursive games, and we prove Theorem 2.4. Actually we prove that if the game is positive recursive, then the conditions of Proposition 3.8 hold.

The method that we use is similar to the one used in Vieille (2000c). For every $\varepsilon > 0$ we define a continuous function, which should be thought of as an approximate best reply, from a compact convex subset of the space of fully mixed stationary profiles into itself. This function has a fixed point x_ε (which is a stationary profile) that yields the players a payoff vector γ_ε. By studying the asymptotic behavior of the family (x_ε), we are able to show that the payoff vector $\lim \gamma_\varepsilon$ satisfies the condition of Proposition 3.8 w.r.t. the stationary profile $\lim x_\varepsilon$.

We begin with some preliminary results, that are essentially a review of Vieille (2000b). We then define for every $\varepsilon > 0$ a continuous function from the space of fully mixed stationary profiles into itself and study the asymptotic behavior of a sequence of fixed points.

5.1. *Preliminaries*

For every stationary profile x, $\gamma(x) = (\gamma_s(x))_{s\in S}$ is the expected undiscounted payoff for the players if they follow x. A stationary profile $x \in X$ is *absorbing* if for every initial state, the probability of reaching an absorbing state is 1, provided the players follow x.

For any stationary profile $x \in X$ and pure action combination $a \in A^S$ we define $x(a) = \prod_{s\in S, i\in N} x_s^i[a_s^i]$.

It is of special interest to know how the distribution of exit from a set depends upon the (stationary) strategies used by the players.

For $B \subseteq S\backslash S^\star$, define a B-graph to be a set g of arrows $[s, a \to s']$, where $s \in B$, $a \in A$, $s' \in S$, such that:

1. For each $s \in B$, there is a unique pair (a, s'), such that $[s, a \to s'] \in g$; moreover, $q(s'|s, a) > 0$.
2. For each $s \in B$, there is a path $(s_0, a_0) \to (s_1, a_1) \to \cdots \to s_N$, such that $s = s_0$, $s_N \notin B$, and $[s_n, a_n \to s_{n+1}] \in g$.

The path in condition 2 is unique. We call it the g-path starting from s. Graphically, a B-graph is a collection of disjoint (directed) paths that end outside B, and visit all states in B.

G_B is the set of B-graphs and, for every $s \in B$ and every $s' \notin B$, $G_B(s \to s')$ is the set of $g \in G_B$, such that the g-path starting from s ends up in s'.

For $x \in X$ and $g \in G_B$, we set

$$w_x(g) = \prod_{[s,a\to s']\in g} x_s(a)q(s' \mid s, a). \tag{26}$$

$w_x(g)$ should be interpreted as the *weight* of g under x. Note that since B-graphs must not have cycles, $\sum_{g\in G_B} w_x(g)$ need not be equal to 1.

Recall that $e_B = \inf\{n \geq 1, s_n \notin B\}$ is the first exit stage from B. If B is transient under x, $e_B < +\infty$, $\mathbf{P}_{s,x}$-a.s. for every initial state $s \in B$.

For a state $s \in B$ set $Q_{s,x}(s' \mid B) = \mathbf{P}_{s,x}(s_{e_B} = s')$. This is the exit distribution from B; that is, the probability that if the initial state is $s \in B$ and the players follow x, the first state that they visit outside B is s'.

A very useful formula that relates $Q_{s,x}$ to the weights of the B-graphs is the following (Freidlin and Wentzell, 1984, Chapter 6, Lemma 3.3). If B is transient under x then

$$\mathbf{P}_{s,x}(s_{e_B} = s') = \frac{\sum_{g\in G_B(s\to s')} w_x(g)}{\sum_{g\in G_B} w_x(g)}. \tag{27}$$

5.2. *Asymptotic Analysis*

The main difficulty in the study of the uniform equilibrium is that the limit payoff $\gamma(s, \tau) = \lim_n \gamma_n(s, \tau)$ is not continuous over the strategy space. A classical approach in the study of the uniform equilibrium in n-player stochastic games is to define for every $\varepsilon > 0$ an ε-approximating game where the payoff function is continuous over the strategy space and moreover admits a stationary equilibrium x_ε. By studying asymptotic properties of the sequence (x_ε) one tries to understand the behavior of the play in the original stochastic game. In the results that appear in the literature, the function $\varepsilon \mapsto x_\varepsilon$ could always be chosen to be a Puiseux function (that is, it has an expansion to a Taylor series in fractional powers of ε) and therefore has useful properties.

In our proof this function need not have such an expansion; hence we must impose these properties in other ways.

For the rest of the section we fix a sequence $(x_\varepsilon)_{\varepsilon>0}$ of absorbing stationary profiles. From (27) and (26), one sees that relevant quantities to the calculation of the exit distribution are the ratios $x_\varepsilon(a_1)/x_\varepsilon(a_2)$, for $a_1, a_2 \in A^S$. By taking a subsequence, we assume w.l.o.g. that $\operatorname{supp}(x_\varepsilon)$ is independent of ε and

$$\theta_{a_1 a_2} = \lim_{\varepsilon\to 0} \frac{x_\varepsilon(a_1)}{x_\varepsilon(a_2)} \text{ exists} \qquad \forall\, a_1 \in A^S, a_2 \in \operatorname{supp}(x_\varepsilon). \tag{28}$$

This implies that $\lim_{\varepsilon\to 0} x_\varepsilon$ exists in X. Moreover, this limit depends only on $(\theta_{a_1a_2})_{a_1a_2}$ and not on the exact sequence (x_ε). We denote this limit by $x(\theta)$. One derives from (27) that $\lim_{\varepsilon\to 0} Q_{s,x_\varepsilon}(\cdot \mid B)$ exists in $\Delta(S\backslash B)$ for every subset $B \subseteq S\backslash S^\star$. The limit is denoted by $Q_{s,\theta}(\cdot \mid B)$.

It follows that $\lim_{\varepsilon\to 0} \gamma(x_\varepsilon)$ exists. This limit is denoted by $\gamma(\theta)$. Clearly $\gamma(\theta)$ need not be equal to $\gamma(x(\theta))$.

DEFINITION 5.1. A set $C \subseteq S\backslash S^\star$ *communicates* under θ if $Q_{s,\theta}(s' \mid C\backslash \{s'\}) = 1$ for every $s \in C$ and every $s' \in C\backslash\{s\}$.

This captures the property that, starting anywhere in C, the play visits "infinitely" many times every state in C before leaving this set (as $\varepsilon \to 0$). Denote by $\mathscr{C}(\theta)$ the collection of sets which communicate under θ.

Note that if $C_1, C_2 \in \mathscr{C}(\theta)$ are maximal w.r.t. set inclusion, then $C_1 = C_2$ or $C_1 \cap C_2 = \emptyset$. Indeed, assume that $s' \in C_1 \cap C_2$, and let $s_1 \in C_1$ and $s_2 \in C_2$. Then $Q_{s_1,\theta}(s_2 \mid C_1 \cup C_2\backslash\{s_2\}) \geq Q_{s_1,\theta}(s' \mid C_1\{s'\}) \times Q_{s',\theta}(s_2 \mid C_2\backslash\{s_2\}) = 1$. Therefore $Q_{s_1,\theta}(s_2 \mid C_1 \cup C_2\backslash\{s_2\}) = 1$. By symmetry, $Q_{s_2,\theta}(s_1 \mid C_1 \cup C_2\backslash\{s_1\}) = 1$. Since s_1 and s_2 are arbitrary, $C_1 \cup C_2$ is communicating under θ. Since C_1 and C_2 are maximal, $C_1 = C_2$.

The following two properties hold for every communicating set $C \in \mathscr{C}(\theta)$:

1. The exit distribution $Q_{s,\theta}(\cdot \mid C)$ is independent of $s \in C$ (Vieille, 2000b, Lemma 5). We denote this limit by $Q_\theta(\cdot \mid C)$.
2. C is communicating under $x(\theta)$ (Vieille, 2000b, Lemma 24).

Let $C \in \mathscr{C}(\theta)$ and $g = \{[s, a_s \to s'], s \in C\}$ be a C-graph. By (28) and (26), it follows that $\lim_{\varepsilon\to 0} w_{x_\varepsilon}(g')/w_{x_\varepsilon}(g)$ exists for every $g, g' \in G_C$, but it may be equal to $+\infty$. A graph $g \in G_C$ is *maximal* if $\lim_{\varepsilon\to 0} w_\varepsilon(g')/w_\varepsilon(g) < +\infty$ for every $g' \in G_C$. Intuitively, a graph g is maximal if there is no other graph that weighs "infinitely" more than g (as $\varepsilon \to 0$).[10]

Let $G_C^{\max}$ be the set of all maximal C-graphs, and let $G_C^{\max}(s \to s') = G_C(s \to s') \cap G_C^{\max}$ be the set of all maximal C-graphs where the path starting from s ends up in s'.

By the definition of maximal graphs and (27), the limit exit distribution satisfies:

LEMMA 5.2. *For every $C \in \mathscr{C}(\theta)$, every $s \in C$ and every $s' \notin C$,*

$$Q_\theta(s' \mid C) = \lim_{\varepsilon\to 0} Q_{s,x_\varepsilon}(s' \mid C) = \lim_{\varepsilon\to 0} \frac{\sum_{g\in G_C^{\max}(s\to s')} w_\varepsilon(g)}{\sum_{g\in G_B} w_\varepsilon(g)}.$$

We now relate the exit distribution $Q_\theta(\cdot \mid C)$ to the set of maximal graphs $G_C^{\max}$.

Let $e = (s, x^{-L}(\theta), a^L)$ be an exit from a communicating set $C \in \mathscr{C}(\theta)$, and let g be a C-graph. We say that g *uses* e if g includes an arrow $[s, (a^{-L}, a^L) \to s']$ for some $a^{-L} \in \operatorname{supp}(x^{-L}(\theta))$ and some $s' \notin C$.

For every communicating set $C \in \mathscr{C}(\theta)$ define

$$e_C^\star = \inf\{n \geq 1, (s_n, a_n) \in E(x(\theta), C)\},$$

[10] Our definition of maximal graphs corresponds to minimal graphs in Vieille (2000b, 2000c).

that is, the first stage that an exit is used. Since using an exit does not necessarily mean leaving C and since the play might leave C not through an exit, we cannot compare $e_C^\star$ and e_C. However, as ε goes to 0 the probability that the play leaves C through an exit converges to 1; hence

$$\lim_{\varepsilon\to 0} \mathbf{P}_{s,x_\varepsilon}(e_C^\star \le e_C) = 1 \qquad \forall s \in C. \tag{29}$$

Eq. (29) implies that $e_C^\star$ induces a probability distribution over the exits from C: for every exit $e \in E(x(\theta), C)$, $\mu_{\theta,C}(e)$ is the limit of the probability that under x_ε the first exit from C that is used is e.

As described earlier by (5), every probability distribution μ on exits $E(x(\theta), C)$ induces a probability distribution $\tilde{\mu}$ on the states outside C. One can verify that $\tilde{\mu}_{\theta,C} = Q_\theta(\cdot \mid C)$.

It follows that an exit $e = (s, a^L)$ from C has a positive probability in $\mu_{\theta,C}$ (that is, $\mu_{\theta,C}[e] > 0$) if and only if there is a maximal C-graph g that uses e. In such a case it follows that the $C\setminus\{s\}$-graph g' that is obtained from g by deleting the arrow $[s, a \to s']$ is maximal. Indeed, otherwise there exists a maximal $C\setminus\{s\}$-graph g'' that satisfies $\lim_{\varepsilon\to 0} w_{x_\varepsilon}(g'')/w_{x_\varepsilon}(g') = +\infty$. But then the C-graph $\tilde{g}$ that is defined by adding the arrow $[s, a \to s']$ to g'' satisfies $\lim_{\varepsilon\to 0} w_{x_\varepsilon}(\tilde{g})/w_{x_\varepsilon}(g) = +\infty$, which contradicts the maximality of g.

5.3. *Proof of Theorem* 2.4

We assume w.l.o.g. that if $x \in X$ is fully mixed then the only ergodic sets of the corresponding Markov chain are the absorbing states. If this was not true then, by turning all ergodic sets w.r.t. such x in $S\setminus S^\star$ into absorbing states with payoff 0, one would get a game with the same set of correlated equilibrium payoffs and with the desired property.

We now prove Theorem 2.4. The proof goes as follows. The space of stationary profiles where each player must play any action with probability at least ε^2 is compact, and the undiscounted payoff is continuous over this space. For every $\varepsilon > 0$ sufficiently small we define a continuous function f from this space into itself. This map should be thought of as an approximate best reply. Since the function is continuous, it has a fixed point x_ε. We define $\theta = (\theta_{a_1a_2})$ as the limit (up to a subsequence) of $\left(\frac{x_\varepsilon(a_1)}{x_\varepsilon(a_2)}\right)_{a_1,a_2}$. We then analyze the asymptotic behavior of the sequence (x_ε) and prove that the triplet $(\gamma(\theta), x(\theta), \Pi(x(\theta)))$ satisfies the conditions of Proposition 3.8.

5.3.1. *Constrained Games*

For $\varepsilon > 0$, define

$$X_\varepsilon = \{x \in X \mid x_s^i[a^i] \ge \varepsilon^2 \quad \forall i \in N, s \in S, a^i \in A^i\}.$$

Let $x \in X_\varepsilon$ and $i \in N$. Define the *continuation cost* of playing a^i in state s against x as

$$c_s(a^i; x) = \max_{\tilde{a}^i \in A^i} q_{s,x^{-i},\tilde{a}^i} \gamma^i(x) - q_{s,x^{-i},a^i} \gamma^i(x). \tag{30}$$

If the continuation payoff of player i is $\gamma^i(x)$, this is the amount that player i gives up by playing a^i, rather than his best reply.

Notice that $0 \leq c_s(a^i; x) \leq 1$, $c_s(a^i; x) = 0$ for each a^i which attains the maximum in (30), and $x \mapsto c_s(a^i; x)$ is continuous over X_ε.

We now define a continuous map f from X_ε into itself. For $a^i \in A^i$ and $s \in S$, set

$$f_s^i(x)[a^i] = \frac{\varepsilon^{c_s(a^i;x)}}{\sum_{\tilde{a}^i \in A^i} \varepsilon^{c_s(\tilde{a}^i;x)}} \tag{31}$$

and $f(x) = (f_s^i(x))_{i\in N, s\in S}$. Observe that $\varepsilon/|A^i| \leq f_s^i(x)[a^i] \leq 1$, and $\sum_{a^i \in A^i} f_s^i(x)[a^i] = 1$. Therefore for ε sufficiently small $f(x) \in X_\varepsilon$. The continuity of f follows from the continuity of the continuation cost. By Brouwer's theorem, f has a fixed point. We denote it by x_ε.

Intuitively, in this fixed point each player i plays the action a^i with a probability that depends on its cost—the higher the cost, the smaller probability it receives. As ε goes to 0, the ratio between the probabilities in which two actions whose cost differ by a constant goes to infinity. Thus, if, as ε goes to 0, two actions are played with "comparable" probabilities, then their cost is the same.

5.3.2. *Asymptotic Analysis*

We study here the asymptotic properties of x_ε as ε goes to 0. Up to a subsequence, we may assume that $\mathrm{supp}(x_\varepsilon)$ is independent of ε and for every $a_1 \in A^S$ and every $a_2 \in \mathrm{supp}(x_\varepsilon)$, $\theta_{a_1 a_2} = \lim_{\varepsilon\to 0} x_\varepsilon(a_1)/x_\varepsilon(a_2)$ exists (it is possibly infinite). In particular, the limits $x = x(\theta) = \lim_{\varepsilon\to 0} x_\varepsilon$ and $\gamma = \gamma(\theta) = \lim_{\varepsilon \to 0} \gamma(x_\varepsilon)$ exist. Thus, $c_s(a^i; x_\varepsilon)$ has a limit $c_s(a^i; \theta)$ that satisfies

$$c_s(a^i; \theta) = \max_{\tilde{a}^i \in A^i} q_{s,x^{-i},\tilde{a}^i} \gamma^i - q_{s,x^{-i},a^i} \gamma^i \qquad \forall\, i \in N, a^i \in A^i, s \in S.$$

Note that $c_s(a^i; \theta) = 0$ does *not* imply that $a_s^i \in \mathrm{supp}(x_s^i)$.

Write $\Pi(\theta) = (C_1, \ldots, C_K, T)$ the decomposition of $S\backslash S^\star$ into maximal communicating sets under θ and the remaining states. For each k, denote $\mu_k = \mu_{\theta, C_k}$. Since $(C_k)_{k=1}^K$ are maximal, the Markov chain induced by $(x, \Pi(x), (\mu_k)_k)$ reaches an absorbing state in finite time.

We shall prove that conditions 1 through 7 of Proposition 3.8 are satisfied. We start with the simplest.

- Condition 1 follows from the definition of $\Pi(\theta)$.
- For every $\varepsilon > 0$, and $s \in S^\star$, $\gamma_s(x_\varepsilon) = r(s)$. Condition 3 follows.
- For every $\varepsilon > 0$ and $s \in S$,

$$q_{s,x_\varepsilon}\gamma(x_\varepsilon) = \gamma_s(x_\varepsilon). \tag{32}$$

Condition 2(a) follows by taking the limit $\varepsilon \to 0$. On the other hand, (32) means that, for all s, the sequence $(\gamma_{s_n}(x_\varepsilon))$ is a martingale under $\mathbf{P}_{s,x_\varepsilon}$ (for the filtration $(\mathscr{H}_n)$, where $\mathscr{H}_n$ is the σ-algebra over the space of infinite histories induced by H_n). By the optional sampling theorem, for every k and every $s \in C_k$,

$$\mathbf{E}_{s,x_\varepsilon}[\gamma_{s_{e_B}}(x_\varepsilon)] = \gamma_s(x_\varepsilon).$$

Condition 5 follows by letting $\varepsilon \to 0$.

- By construction, $x_s^i[a^i] > 0$ implies that $\lim_{\varepsilon\to 0} c_s(a^i; x_\varepsilon) = 0$; therefore $q_{s,x^{-i},a^i}\gamma^i = \max_{a^i\in A^i} q_{s,x^{-i},a^i}\gamma^i$. By summation over $a^i \in \text{supp}(x_s^i)$, one gets

$$q_{s,x}\gamma^i = \max_{a^i\in A^i} q_{s,x^{-i},a^i}\gamma^i.$$

Since $q_{s,x}\gamma^i = \gamma^i(s)$ and the continuation cost is non-negative, condition 2(b) is established.

LEMMA 5.3. *Condition* 4 *holds*; *that is*, $\gamma \geq v$.

Proof. Assume that the inequality $\gamma^i \geq v^i$ does not hold for player i. Let $S_0 \subseteq S$ contain the states s where $v^i(s) - \gamma^i(s) > 0$ is maximal. Since $v^i(s) = \gamma^i(s)$ for $s \in S^\star$, $S_0 \cap S^\star = \emptyset$. Since $\gamma^i(s) \geq 0$ for every s, $v^i(s) > 0$ for $s \in S_0$. Let $S_1 \subseteq S_0$ contain the states of S_0 where v^i is maximized. There exist $s \in S_1$, $a^i \in A^i$, with

$$q(S_1 \mid s, x^{-i}, a^i) < 1 \tag{33}$$

and

$$q_{s,x^{-i},a^i}v^i \geq v^i(s) = \max_{s'\in S_0} v^i(s') \tag{34}$$

(otherwise players $N\backslash\{i\}$ could bring player i's payoff below v^i by playing x^{-i} on S_1, and punishing him if the play leaves S_1). By construction of S_1, (33) and (34) imply that $q(S_0 \mid s, x^{-i}, a^i) < 1$.

On the other hand, $q_{s,x^{-i},a^i}\gamma^i \leq \gamma^i(s)$ by condition 2. Thus

$$q_{s,x^{-i},a^i}v^i - q_{s,x^{-i},a^i}\gamma^i \geq v^i(s) - \gamma^i(s) = \max_{s'\in S}(v^i(s') - \gamma^i(s')).$$

This implies $q(S_0 \mid s, x^{-i}, a^i) = 1$—a contradiction. ∎

To prove conditions 6 and 7 we need the following definitions. For every subset $C \subseteq S$ and every C-graph $g \in G_C$, define the *overall cost* of g by

$$c(g; \theta) = \sum_{i \in N} \sum_{[s,a \to s'] \in g} c_s(a^i; \theta).$$

Since x_ε is a fixed point of (31), it follows that if $c(g_1; \theta) < c(g_2; \theta)$ then $\lim_{\varepsilon \to 0} w_{x_\varepsilon}(g_1)/w_{x_\varepsilon}(g_2) = \lim_{\varepsilon \to 0} \varepsilon^{c(g_1;\theta)-c(g_2;\theta)} = +\infty$. Note that if $c(g_1; \theta) = c(g_2; \theta)$, then we have no information on this limit.

For every communicating set $C \in \mathscr{C}(\theta)$ and every exit $e = (s, a^L)$ from C define $c(e; \theta)$, the *cost of the exit e,* as the overall cost of a maximal C-graph that uses e (as the proof of Lemma 5.4 below shows, it is irrelevant which maximal C-graph is chosen).

LEMMA 5.4. *Let $e_1, e_2 \in E(x, C_k)$ be two exits. If $\mu_k[e_1], \mu_k[e_2] > 0$, then $c(e_1; \theta) = c(e_2; \theta)$.*

Proof. Let $e_1, e_2 \in E(x, C_k)$ be two exits such that $c(e_1; \theta) < c(e_2; \theta)$. Let g_1 be a maximal C-graph that uses e_1, and let g_2 be an arbitrary C-graph that uses e_2. By assumption, $c(g_1; \theta) = c(e_1; \theta) < c(e_2; \theta) \leq c(g_2; \theta)$, which implies that $\lim_{\varepsilon \to 0} w_{x_\varepsilon}(g_1)/w_{x_\varepsilon}(g_2) = +\infty$. Since g_2 is arbitrary the result follows. ∎

LEMMA 5.5. *Let $C \in \mathscr{C}(\theta)$ be a communicating set and let $s_0 \in C$ be arbitrary. There exists a $C \backslash \{s_0\}$-graph g such that* (i) *every g-path ends at s_0, and* (ii) *the overall cost of g is* 0.

Proof. Since $C \in \mathscr{C}(\theta)$, it follows that $C \in \mathscr{C}(x(\theta)) = \mathscr{C}(x)$ as well. Hence, under the stationary profile y_{C,x,ε,s_0} the set C is stable and, starting from C, the play reaches s_0 in finite time.

It follows that there exists a graph $g = \{[s, a_s \to s']\} \in G_{C\backslash\{s_0\}}$ such that a_s^i belongs to the support of $y^i_{C,x,\varepsilon,s_0}$ for every $s \in C\backslash\{s_0\}$ and every player $i \in N$. Indeed consider the directed graph (in the sense of graph theory) whose vertices are the states in C and includes the edge $s \to s'$ if and only if there exists a mixed action combination $a_s \in \text{supp}(y_{C,x,\varepsilon,s_0})$ such that $q(s' \mid s, a_s) > 0$. Since from each vertex s there is a directed path to s_0, it follows that there is a spanning tree with the same property. The spanning tree defines a natural $C\backslash\{s_0\}$-graph g. Since g was defined by a spanning tree, any g-path ends at s_0.

Since C is stable under $y_{C,x,\varepsilon,s}$, it follows that $q(C \mid s, x^{-i}a_s^i) = 1$. Since $\gamma(\theta)$ is constant over C, $c_s(a_s^i; \theta) = 0$. Summing over all arrows in g we get $c(g; \theta) = 0$. ∎

COROLLARY 5.6. *Let $C \in \mathscr{C}(\theta)$ be a communicating set and let $e \in E(x, C)$ be an exit. If e is a joint exit then $c(e; \theta) = 0$, whereas if $e = (s, a^i)$ is a unilateral exit then $c(e; \theta) = \gamma_s(\theta) - q_{s,x^{-i}(\theta),a^i}\gamma(\theta)$.*

Proof. Let $e = (s, a^L)$ be a joint exit. Then $|L| \geq 2$. Let g be the $C\setminus\{s\}$-graph that satisfies Lemma 5.5 for $s_0 = s$. Let g' be the C-graph that is defined by adding to g the arrow $[s, (a^{-L}, a^L) \to s']$ for some $a^{-L} \in \text{supp}(x^{-L}(\theta))$ and $s' \notin C$. Since $|L| \geq 2, q(C \mid s, x^{-i}(\theta), a^i) = 1$ for every $i \in L$. Since $\gamma(\theta)$ is constant on C, $c(a^i; \theta) = 0$ for every $i \in L$. It follows that $c(g; \theta) = 0$; hence $c(e; \theta) = 0$ as well.

Now let $e = (s, a^i)$ be a unilateral exit. Any C-graph that uses e must have an arrow of the form $[s, (a^{-i}, a^i) \to s']$ for some $a^{-i} \in x^{-i}(\theta)$ and $s' \notin C$. Hence the cost of any such graph is at least $\gamma_s(\theta) - q_{s,x^{-i}(\theta),a^i}\gamma(\theta)$. But as in the case of a joint exit, we can construct a C-graph that uses e and whose overall cost is exactly $\gamma_s(\theta) - q_{s,x^{-i}(\theta),a^i}\gamma(\theta)$, and therefore this is the cost of e. ∎

We now show how Lemma 5.4 and Corollary 5.6 imply that conditions 6 and 7 hold.

If condition 6(b) does not hold then there is a joint exit e in $\text{supp}(\mu_k)$. By Corollary 5.6 $c(e; \theta) = 0$. By Lemma 5.4 $c(e'; \theta) = 0$ for every exit $e' \in \text{supp}(\mu_k)$, which by Corollary 5.6 establishes condition 6(a).

Condition 7 also follows. Let $e_1, e_2 \in E(x, C_k)$ be two unilateral exits of player i such that $\mu_k[e_1] > 0$. We will show that $q_{e_2}\gamma^i \leq q_{e_1}\gamma^i \leq \gamma^i_{C_k}$. By condition 2(b), $q_{e_1}\gamma^i \leq \gamma^i_{C_k}$. Assume to the contrary that $q_{e_2}\gamma > q_{e_1}\gamma$. Since γ is constant over C_k, it follows by Corollary 5.6 that $c(e_2; \theta) < c(e_1; \theta)$. In particular, there exists a C_k-graph g_2 that uses e_2 such that for every C_k-graph g_1 that uses g_1, $\lim_{\varepsilon\to 0} w_\varepsilon(g_2)/w_\varepsilon(g_1) = +\infty$, which implies that $\mu_k[e_1] = 0$, a contradiction.

REFERENCES

Aumann, R. J. (1974). "Subjectivity and Correlation in Randomized Strategies," *J. Math. Econ.* **1**, 67–96.

Aumann, R. J. (1987). "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica* **55**, 1–18.

Bewley, T., and Kohlberg, E. (1976). "The Asymptotic Theory of Stochastic Games," *Math. Oper. Res.* **1**, 197–208.

Blackwell, D., and Ferguson, T. S. (1968). "The Big Match," *Ann. Math. Statist.* **39**, 159–163.

Everett, H. (1957). "Recursive Games," in *Contributions to the Theory of Games*, Vol. 3, pp. 47–78, Annals of Mathematical Studies, AM-39, Princeton, NJ: Princeton Univ. Press.

Fink, A. M. (1964). "Equilibrium in a Stochastic n-Person Game," *J. Sci. Hiroshima Univ.* **28**, 89–93.

Forges, F. (1986). "An Approach to Communication Equilibria," *Econometrica* **54**, 1375–1385.

Forges, F. (1988). "Communication Equilibria in Repeated Games with Incomplete Information," *Math. Oper. Res.* **13**, 77–117.

Freidlin, M., and Wentzell, A. (1984). *Random Perturbations of Dynamical Systems*. Berlin: Springer-Verlag.

Fudenberg, D., and Tirole, J. (1991). *Game Theory*. Cambridge, MA: MIT Press.

Mertens, J. F. (1994). "Correlated and Communication Equilibria," in *Game Theoretic Methods in General Equilibrium Analysis* (J. F. Mertens and S. Sorin, Eds.). Dordrecht/Norwell, MA: Kluwer Academic.

Mertens, J. F., and Neyman, A. (1981). "Stochastic Games," *Int. J. Game Theory*, **10**, 53–66.

Myerson, R. B. (1986). "Multistage Games with Communication," *Econometrica* **54**, 323–358.

Neyman, A. (1988). "Stochastic Games," preprint.

Nowak, A. S. (1991). "Existence of Correlated Weak Equilibria in Discounted Stochastic Games with General State Space," in *Stochastic Games and Related Topics* (T. E. S. Raghavan, *et al.*, Eds.). Dordrecht/Norwell, MA: Kluwer Academic.

Shapley, L. S. (1953). "Stochastic Games," *Proc. Nat. Acad. Sci. U.S.A.* **39**, 1095–1100.

Solan, E. (1999). "Three-Player Absorbing Games," *Math. Oper. Res.* **24**, 669–698.

Solan, E. (2000). "Rationality and Extensive From Correlated Equilibria in Stochastic Games," D.P. 1298. The Center for Mathematical Studies in Economics and Management Science, Northwestern University.

Solan, E., and Vieille, N. (1998). "Quitting Games," *Math. Oper. Res.,* to appear.

Vieille, N. (1998). "Solvable Sets in N-Player Stochastic Games," *SIAM J. Control Opt.* **38**, 1794–1804.

Vieille, N. (2000a). "Equilibrium in 2-Person Stochastic Games. I. A Reduction," *Israel J. Math.* **119**, 55–91.

Vieille, N. (2000b). "Small Perturbations and Stochastic Games," *Israel J. Math.* **119**, 127–144.

Vieille, N. (2000c). "Equilibrium in 2-Person Stochastic Games. II. The Case of Recursive Games," *Israel J. Math.* **119**, 93–126.

Vrieze, O. J., and Thuijsman, F. (1989). "On Equilibria in Repeated Games with Absorbing States," *Int. J. Game Theory* **18**, 293–310.