Solan, Eilon

Working Paper

Rationality and Extensive Form Correlated Equilibria in Stochastic Games

Discussion Paper, No. 1298

Provided in Cooperation with:
Kellogg School of Management - Center for Mathematical Studies in Economics and Management Science, Northwestern University

Suggested Citation: Solan, Eilon (2000) : Rationality and Extensive Form Correlated Equilibria in Stochastic Games, Discussion Paper, No. 1298, Northwestern University, Kellogg School of Management, Center for Mathematical Studies in Economics and Management Science, Evanston, IL

This Version is available at:
https://hdl.handle.net/10419/221654

Rationality and Extensive Form Correlated Equilibria in Stochastic Games *

Eilon Solan[†‡]

June 14, 2000

Abstract

We define the notion of rational payoffs in stochastic games. We then prove that the set of rational payoffs coincides with the set of extensive form correlated equilibrium payoffs; those are equilibrium payoffs in an extended game that includes an autonomous correlation device: a device that sends at every stage a private signal to each player, which is independent of the play, but may depend on previous signals. In particular, it follows that communication between the players and/or between the players and the correlation device cannot increase the set of equilibrium payoffs.

*This is a thorough revision of "Extensive Form Correlated Equilibria", discussion paper #175 , Center for Rationality and Interactive Decision Theory, The Hebrew University of Jerusalem.

[†]Department of Managerial Economics and Decision Sciences, Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston IL 60208. e-mail: e-solan@nwu.edu

[‡]I thank Rakesh Vohra, the associate editor of the *International Journal of Game Theory* and two anonymous referees for their helpful comments.

1 Introduction

The folk theorem characterizes the set of feasible and individually rational payoffs in an infinitely repeated game by means of the one-shot game. As long as the outcome of the coins used by the players is private, a payoff vector that is individually irrational for at least one player cannot be the expected outcome of the game.

There are two features of infinitely repeated games that enable such a characterization. First, the individually rational level of all players remain the *same* along the play, and second, the *same* payoffs are available at each stage.

Aumann (1987) proved that in one-shot games, if the players are Bayesian rational and utility maximizers, then the expected outcome of the game must be a correlated equilibrium, and, given a suitable information structure, every correlated equilibrium can be the expected outcome of the game.

In the present paper we are interested in stochastic games. In particular, the features mentioned above are missing. Since the state of the game changes, the individually rational level of the players changes as well. Moreover, a payoff that is sustainable at some stage may not be sustainable at the next stage, if the state of the game has changed. Since the game is played in stages, players may learn new information as the game proceeds, in which case Aumann's characterization is not valid.

Our first task is to define rational outcomes in a dynamic setup. In one shot games, Aumann (1992) measures "expected irrationality" for each player i in the following way:

> "At each of i's information states, one multiplies the probability of that state by the difference between i's expected payoff there and the maximum expected payoff that he could have gotten by changing his strategy; then one sums over all of i's information states."

Thus, the expected irrationality depends on the strategy profile that is to be played, and it measures the maximum (in expected utility terms) that a player can profit by deviating from his prescribed strategy. Any payoff that corresponds to a strategy profile whose expected irrationality is 0, is a rational outcome of the game.

When the setup is dynamic, there are two points we should take into account. First, the information state changes along the play. Hence, at some

information state, a player may be aware that even though he may profit by deviating today, he may profit even more if he waits few more stages and deviates in more favorable circumstances. Second, since players may acquire additional information along the play, a player who deviates from his prescribed strategy cannot disregard the possibility that his opponents will learn about his deviation in subsequent stages, and, once they learn about it, will also deviate from *their* prescribed strategies, and punish him as severely as they can.

We represent each player's decision to deviate by a stopping time: at every information state, a player may either follow his prescribed strategy, or deviate. A deviation corresponds, then, to a stopping time, supplemented with a new strategy that should be followed whenever the stopping time indicates to stop following the original strategy.

Thus, we define the expected irrationality of some strategy profile w.r.t. player i as the maximum that player i can profit by deviating from his prescribed strategy, provided his deviation is followed by an indefinite punishment. Note that the worst is assumed from the point of view of player i: his deviation is immediately detected and punished.

In general, players may have asymmetric information, in which case the play follows a correlated strategy profile; that is, the evolution of the game is equivalent to a situation where at every stage an action combination is chosen according to some joint probability distribution over the space of action combinations. We then say that player i deviates if, when some action combination was chosen, and each player was informed of his action in this combination, player i decides to play some other action.

One way to define rational payoffs would be as payoffs that correspond to some correlated strategy profile whose expected irrationality is 0 w.r.t. every player. In other words, no player can profit by deviating at any stage from his recommended action, provided his deviation is followed by an indefinite punishment.

Under this definition, the set of rational payoffs may be empty even for simple games (e.g., the "Big Match" with the limsup evaluation, studied by Blackwell and Ferguson (1968)). We therefore define the set of rational payoffs in a more robust way.

We say that a payoff vector is ϵ-rational if it is the expected payoff of some correlated strategy profile whose expected irrationality, w.r.t. every player, is less than ϵ. Any limit of ϵ-rational payoffs, as ϵ goes to 0, is a rational payoff vector.

Note that this definition is the weakest possible: if a payoff vector is irrational than it is irrational if deviations are not detected immediately. Indeed, if a player can profit by a deviation that is punished immediately, he can clearly profit if he deviates but punishment is delayed. Thus, a payoff vector that is irrational cannot be the expected outcome of the game, whatever be the information structure of the game.

Our main result is that in stochastic games, the set of rational outcomes coincides with the set of extensive form correlated equilibrium payoffs. Those are equilibria in an extended game that includes an autonomous correlation device: a device that sends, at *every* stage, a private signal to each player, which may depend on past signals but not on past play. The devices we use are not canonical (Forges, 1988a): the signal each player receive is not a recommended action, but a vector of recommended actions, one for each possible history, as well as the vector of actions that were recommended at the previous stage.

This result is a folk theorem for stochastic games. Moreover, it shows that in stochastic games introducing communication between the players and an autonomous correlation device cannot enlarge the set of equilibrium payoffs. This phenomenon was found by Forges (1988a) for a class of repeated games with incomplete information on one side. Since repeated games with incomplete information on one side can be represented as a stochastic game, where the state variable is the posterior distribution of the uninformed player, our result generalizes the result of Forges (1988a). Recall that when the information is asymmetric, communication between the players and the correlation device may enlarge the set of equilibrium payoffs (Forges, 1986, Example 1). Our characterization result can also be viewed as a generalization of Aumann's (1987) result to stochastic games.

Since our result characterizes the set of extensive form correlated equilibrium payoffs, it may be used to prove non-emptiness of this set. Indeed, Solan and Vieille (1998) proved that every undiscounted stochastic game with finite state and action spaces admits a rational payoff, hence, by our result, an extensive form correlated equilibrium payoff.

When the punishment level of the players is independent of the history, a stronger result can be obtained. In this case, the set of rational outcomes coincides with the set of correlated equilibrium payoffs. In particular, in this class of games the set of correlated equilibrium payoffs coincides with the set of extensive form correlated equilibrium payoffs.

The two characterization results hold in a very general setup. The state

space is an arbitrary measurable separable space, the action spaces of the players are arbitrary complete separable metric spaces, and the payoff function may be any measurable function from histories to payoffs.

Since we deal with a general setup, we represent strategy profiles using a countable sequence of i.i.d. r.v.s. This representation has its own merit, and may be useful elsewhere.

When the number of available actions is finite, and there are at least four players, then a simple application of a result due to Bárány (1992) shows that the set of rational payoffs coincides with the set of direct communication equilibrium payoffs; that is, equilibrium payoffs in an extended game where players may send private messages to each other at every stage. We elaborate on this result in section 6.

Our work is related to that of Myerson (1986), who studies multi-stage games, and characterizes the set of sequential communication equilibria using codominated actions. Nevertheless, there are some important differences. First, Myerson's equilibria are sequential, while in our equilibria players may be required to punish a deviator, which may be irrational for some of the players (though punishment never occurs on the equilibrium path). Second, Myerson is concerned with finite multi-stage games, whereas in our model the game may last infinitely many stages.

The paper does not address the question of whether the set of rational payoffs is empty or not. Moreover, existence of a rational payoff is known only in special cases, and in all but one, the existence is asserted by proving the existence of an equilibrium or of a correlated equilibrium (see, e.g., Mertens, Sorin and Zamir (1994) for the existence of equilibria in finite-stage games and discounted games, and Nowak (1991) for the existence of correlated equilibrium in discounted games). The only exception is, as mentioned above, stochastic games with finitely many states and actions, and the lim sup evaluation, where the existence of a rational payoff is proved directly (and the existence of an equilibrium payoff or of a correlated equilibrium payoff has not been proved yet).

The paper is arranged as follows. The model is presented in section 2, and rational payoffs in section 3. In section 4 we define autonomous correlation devices and extensive form correlated equilibria, and we state two equivalence theorems. These theorems are proved in section 5. In section 6 we explain how communication can substitute correlation, provided there are at least four players.

2 The Model

For every measurable space Y we denote by $\mathcal{P}(Y)$ the space of probability measures over Y. If $\mu \in \mathcal{P}(Y)$ and $C \subseteq Y$ is a measurable set, then $\mu[C]$ is the measure of C under μ. A function $f : X \to \mathcal{P}(Y)$ is measurable if for every measurable subset $C \subseteq Y$ the function $g : X \to [0,1]$ defined by $g(x) = f_x[C]$ is measurable. A product (*resp.* union) of measurable spaces is always endowed with the product (*resp.* union) σ-algebra. Finally, a correspondence is a set-valued function, and a correspondence $\phi : X \to Y$ is measurable if the set $\{x \in X \mid \phi(x) \cap C \neq \emptyset\}$ is X-measurable for every closed subset $C \subseteq Y$.

A *stochastic game* G is given by:

1. A finite set of players I.
2. A measurable space of states S.
3. An initial state $s_1 \in S$.
4. For every player $i \in I$, a complete separable metric space of pure actions A_0^i. We denote $A_0 = \times_{i \in I} A_0^i$.
5. For every player $i \in I$, a measurable correspondence $A^i : S \to A_0^i$. $A^i(s)$ is the set of actions available for player i in state s. We denote $A(s) = \times_{i \in I} A^i(s)$. The space of *infinite histories* is denoted by H_∞:

   $$H_\infty = \{(s_1, a_1, s_2, a_2, \ldots) \in \{s_1\} \times (A \times S)^{\mathbf{N}} \mid a_n \in A(s_n) \qquad \forall n \in \mathbf{N}\}.$$

   We endow H_∞ with the σ-algebra generated by all the finite cylinders.
6. A measurable transition rule q that assigns for each $(s,a) \in \mathrm{Gr}(A)$ a probability measure in $\mathcal{P}(S)$.
7. For every player $i \in I$, a measurable bounded utility function $u^i : H_\infty \to [-R, R]$, where $R \in \mathbf{R}$.

The game is played in stages. At stage n each player is informed of past play $h_n = (s_1, a_1, \ldots, s_n)$, and chooses an action $a_n^i \in A^i(s_n)$, independently of his opponents. The action combination $a_n = (a_n^i)$ that was chosen and the current state s_n determine a new state s_{n+1}, according to the probability measure $q(s_n, a_n)$.

The payoff for each player $i \in I$ is determined by the *infinite* path that has occurred, and is equal to $u^i(s_1, a_1, s_2, a_2, \ldots)$ Note that our definition of a utility function, which follows Maitra and Sudderth (1998), is more general than the standard approach of using daily payoffs.

2.1 Strategies

We denote the space of histories of length n by H_n. The last state of a history h_n of length n is denoted by s_n. The history (s_1) is denoted by s_1. The space of all finite histories is denoted by $H = \cup_{n \in \mathbf{N}} H_n$. Whenever we say that $h_n \in H$, we implicitly mean that h_n has length n.

DEFINITION 2.1 *A* strategy *of player i is a measurable function $\sigma^i : H \to \mathcal{P}(A_0^i)$ such that $\sigma^i(h_n)[A^i(s_n)] = 1$ for every $h_n \in H$. A* profile *is a vector of strategies $\sigma = (\sigma^i)_{i \in I}$. A* correlated profile *is a measurable function $\sigma : H \to \mathcal{P}(A_0)$ such that $\sigma(h_n)[A(s_n)] = 1$ for every $h_n \in H$.*

Note that every profile is a correlated profile. We denote by Σ^i the space of profiles of player i, by $\Sigma_\star$ the space of correlated profiles, and by $\Sigma_\star^{-i}$ the space of correlated profiles of players $N \setminus \{i\}$; that is, the space of measurable functions $\sigma^{-i} : H \to \mathcal{P}(A_0^{-i})$ such that $\sigma^{-i}(h_n)[A^{-i}(s_n)] = 1$ for every $h_n \in H$, where $A^{-i}(s_n) = \times_{j \neq i} A^j(s_n)$.

By Ionescu-Tuclea Theorem (see, e.g., Neveu (1965), Proposition V.1.1), every finite history $h_n \in H$ and every correlated profile σ induce a probability measure $\mathbf{P}_{h_n,\sigma}$ over H_∞; that is, the probability measure induced by σ in the subgame beginning with h_n. We denote expectation w.r.t. this probability measure by $\mathbf{E}_{h_n,\sigma}$.

2.2 Payoffs

For every correlated profile σ and every finite history $h_n \in H$ we denote

$$\gamma^i(h_n, \sigma) = \mathbf{E}_{h_n,\sigma} u^i(s_1, a_1, \ldots).$$

The *payoff* of a correlated profile σ is defined by $\gamma(\sigma) = (\gamma^i(s_1, \sigma))_{i \in I}$.

For every player $i \in I$ and every finite history $h_n \in H$ we define the *punishment level* of player i by:

$$v_{h_n}^i = \inf_{\sigma^{-i} \in \Sigma_\star^{-i}} \sup_{\sigma^i \in \Sigma^i} \gamma^i(h_n, \sigma).$$

$v^i_{h_n}$ is the punishment level that players $N \setminus \{i\}$ can inflict on player i when they act as a single player. A correlated strategy profile σ^{-i} that approximates this infimum up to ϵ is called an *ϵ-punishment* strategy profile.

We assume that for every $n \in \mathbf{N}$, every $\epsilon > 0$ and every player $i \in I$ there exists a correlated profile $\tilde{\sigma}^{-i}_{\epsilon} \in \Sigma^{-i}_{\star}$ such that

$$\sup_{\sigma^i \in \Sigma^i} \gamma^i(h_n, (\tilde{\sigma}^{-i}_{\epsilon}, \sigma^i)) < v^i_{h_n} + \epsilon \qquad \forall h_n \in H_n,$$

and for every correlated profile $\sigma^{-i} \in \Sigma^{-i}_{\star}$ there is a strategy $\sigma^i \in \Sigma^i$ such that

$$\gamma^i(h_n, (\sigma^{-i}, \sigma^i)) > v^i_{h_n} - \epsilon.$$

We do not know under which conditions such strategy profiles exist. However, in various special cases such a correlated profile is known to exist: (i) if the state and action spaces are countable, then there are no measurability issues, and (ii) if the utility function is the discounted sum or the limsup of daily payoffs, then existence was proved in general set-ups (see, e.g., Mertens, Sorin and Zamir (1994) for the discounted sum, and Maitra and Sudderth (1993) for the limsup).

Note that in general the punishment strategy and the strategy of player i that defends the punishment level depend on the past play, rather than only on the current state. This is the case whenever the future payoff depends on past play. When using the discounted or the limsup evaluation, future payoff does not depend on past play, and indeed in these cases those strategies depend only on the current state.

3 Expected Irrationality

In this section we assign for each correlated profile σ and every player $i \in I$ a non-negative number $U^i(\sigma)$, that measures how much player i can profit by deviating from σ^i, provided his deviation is followed by an indefinite punishment. In other words, $U^i(\sigma)$ measures the expected irrationality of following σ for player i.

For every correlated profile σ, every finite history $h_n \in H$ and every action $a^i \in A^i$, define $\sigma(h_n) \mid a^i$ to be the conditional probability over A^{-i} given a^i.[1] If player i received the signal a^i, $\sigma(h_n) \mid a^i$ is his conditional probability on the joint action played by his opponents.

[1]Formally, this is the disintegration of $\sigma(h_n)$ w.r.t. the function $f : A \to A^i$ that

Let $h_n \in H$ be a finite history, $a^i \in A^i$ an action, and σ be a correlated profile. Define

$$U^i(h_n, \sigma, a^i) = \max\{\sup_{b^i \neq a^i} \mathbf{E}_{\sigma(h_n)|a^i}(v^i_{h_n, b^i, a^{-i}} - \gamma^i((h_n, a^i, a^{-i}), \sigma)), 0\}.$$

The first term in the maximization is the maximal amount that player i can profit by deviating after the history h_n, given the action he should have played was a^i, and his deviation is followed by an indefinite punishment. Therefore, $U^i(h_n, \sigma, a^i)$ is equal to 0 if player i cannot profit, while it his equal to his maximal profit, if such a profit is available.

In Aumann's (1992) terms, $U^i(h_n, \sigma, a^i)$ is the amount that player i can profit in the information state (h_n, a^i).

Any measurable stopping time $t : H_\infty \to \mathbf{N}$ and every correlated profile σ induce, by Ionescu-Tuclea Theorem, a probability measure over H. Denote expectation w.r.t. this measure by $\mathbf{E}_{t,\sigma}$. Define the *expected irrationality of* σ w.r.t. player i by

$$U^i(\sigma) = \sup_t \mathbf{E}_{t,\sigma} U^i(h_t, \sigma, a_t)$$

where the supremum is over all measurable stopping times. In other words, given that the players should follow σ, player i may stop following σ whenever he chooses. However, one stage afterwards, he is being punished with his punishment level. $U^i(\sigma)$ measures the maximal amount that player i can profit by such a process, where the profit is measured relative to following σ indefinitely.

DEFINITION 3.1 *Let $\epsilon > 0$. A payoff vector $g \in \mathbf{R}^I$ is* ϵ-rational *if there exists a correlated strategy profile σ such that (i) $\gamma(s_1, \sigma) = g$, and (ii) $U^i(\sigma) < \epsilon$ for every $i \in I$.*

A payoff vector $g \in \mathbf{R}^I$ is rational *if it is the limit of ϵ-rational payoffs as ϵ goes to 0; that is, for every $\epsilon > 0$ there exists an ϵ-rational payoff vector g_ϵ such that $\|g - g_\epsilon\| < \epsilon$.*

We denote the set of rational payoffs by E_0.

Note that $E_0 \in \mathbf{R}^I$, and it depends on the initial state s_1.

A payoff vector g is rational if there exists a sequence of correlated strategy profiles such that the corresponding payoffs converge to g (feasibility) and their expected irrationality converge to 0 (individual rationality).

---

is defined by $f(a) = a^i$, projected on A^{-i} (see Dellacherie and Meyer, 1978). Since we require that $(\sigma(h_n) \mid a^i)[A^i] = 1$, regular conditional probabilities do not suffice.

Thus, we rule out as irrational payoff vectors only those vectors that either (i) cannot be supported by correlated strategy profiles, or (ii) can be supported by correlated strategy profiles, but those profiles are irrational for at least one player: if these profiles are played, at least one player can substantially profit by deviating, whatever threats his opponents make. In particular, if a payoff vector is individually irrational, it cannot be the expected outcome of the game. As we see in the next section, any payoff that is individually rational can be an equilibrium payoff in some extended game that includes an autonomous correlation device.

3.1 Properties of E_0

In finite stage games with finite state and action sets, the set of rational payoffs is a compact and closed polyhedron. This fact can be proved directly, or deduced from Corollary 2 in Forges (1986) and our characterization theorem.

It is easy to verify that in a general setup, the set E_0 is closed and convex by definition.

However, in general it needs not be a polyhedron. We provide two examples where E_0 is not a polyhedron. The first is of a two-player one shot game, where the action spaces of the two players are the unit intervals, and the second is of an infinite stage game where the action spaces of the players are finite. Moreover, in the second example the punishment level is independent of the history. It follows from Theorem 4.7 below that in both examples, the set of correlated equilibrium payoffs, that coincides with the set E_0, is not a polyhedron.

As the example in section 4.3 shows, the set of correlated equilibrium payoffs may be a strict subset of E_0.

Example 1: Consider a two-player one shot game, where the action space of each player is the closed unit interval, and the payoff function is

$$u(a^1, a^2) = \begin{cases} (0,0) & a^1 \neq a^2 \\ (\cos(a^1), \sin(a^1)) & a^1 = a^2 \end{cases}$$

For every $x \in [0,1]$, (x,x) is an equilibrium, hence $(\cos(x), \sin(x))$ is an equilibrium payoff, and in particular in E_0. However, the set $\{(\cos(x), \sin(x)) \mid x \in [0,1]\}$ is the pareto frontier of E_0, hence E_0 is not a polyhedron.

Example 2: Consider a two-player game, where $A^1 = \{0, 1, \text{Stop}\}$ and $A^2 = \{\text{Continue}, \text{Stop}\}$. The game terminates once at least one player stops.

If a player stops at any stage, he receives -2. If a player does not stop while his opponent stops, he receives -1. If the play continues forever, then player 2 continued at all stages. In this case, the moves of player 1 are a sequence of zeroes and ones, and define naturally a number x in the unit interval. The payoff is, then, $(\cos(x), \sin(x))$. As in Example 1, the set $\{(\cos(x), \sin(x)) \mid x \in [0,1]\}$ is the pareto frontier of E_0, hence E_0 is not a polyhedron.

4 Extensive Form Correlated Equilibria

In this section we define autonomous correlation devices. We then extend the original stochastic game by introducing such a device. Equilibrium payoffs in the extended game are extensive form correlated equilibrium payoffs. We prove that the set of extensive form correlated equilibrium payoffs coincides with the set E_0 of rational payoffs. If for every player the punishment level is independent of the history, a stronger result is obtained: E_0 coincides with the set of correlated equilibrium payoffs. Finally, by studying an example, we see that there exist games, and extensive form correlated equilibrium payoffs in these games, that pareto dominate all correlated equilibrium payoffs. We use this example to illustrate the autonomous correlation devices we use in the proofs.

4.1 The Extended Game

DEFINITION 4.1 *An* autonomous correlation device $\mathcal{D}$ *is given by*

- *For every player* $i \in I$ *and every* $n \in \mathbf{N}$*, a measurable space* M_n^i *of* signals *or* messages*. Denote* $M_n = \times_{i \in I} M_n^i$*.*
- *For every* $n \in \mathbf{N}$*, a measurable function* $\mu_n : M_1 \times \cdots \times M_{n-1} \to \mathcal{P}(M_n)$*,*

Given a stochastic game G and an autonomous correlation device $\mathcal{D}$, we define an extended game $G(\mathcal{D})$ that is played as follows. At each stage n, a signal $m_n = (m_n^i) \in M_n$ is chosen according to $\mu_n(m_1, \ldots, m_{n-1})$, and each player i is informed of m_n^i. Each player then chooses an action $a_n^i \in A^i(s_n)$, and a new state s_{n+1} is chosen according to $q(s_n, a_n)$, where $a_n = (a_n^i)_{i \in I}$. Both the action combination a_n that was played and the new state s_{n+1} are publicly announced.

Note that the signals are payoff irrelevant, and they are independent of past play.

We assume that players have infinite recall, so each player i can base his choice of an action at stage n on past play $(s_1, a_1, \ldots, s_n)$ and on past signals $(m_1^i, \ldots, m_n^i)$ *he* has received.

A *correlation device* is an autonomous correlation device $\mathcal{D} = ((M_n^i), \mu_n)$ such that M_n^i contains one element for every $n \geq 2$. That is, the players do not get informative signals after the first stage.

Let $H^i(M)$ be the space of all finite histories that player i can observe in $G(\mathcal{D})$; that is, the space of all sequences $(s_1, m_1^i, a_1, \ldots, s_{n-1}, m_{n-1}^i, a_{n-1}, s_n, m_n^i)$ such that $a_k \in A(s_k)$ and $m_k^i \in M_k^i$. Note that, since the signals are private, each player observes a (possibly) different history. Let $H(M)$ be the space of all finite histories that an outside observer, who observes *both* the actions of the players and the signals sent to *all* the players, can observe. Let $H_\infty(M)$ be the space of all infinite histories that this outside observer can observe. We endow $H_\infty(M)$ with the σ-algebra generated by all the finite cylinders. Note that the spaces $(H^i(M))_{i\in I}$, $H(M)$ and $H_\infty(M)$ are independent of $(\mu_n)_{n\in\mathbf{N}}$.

A *strategy* for player i in $G(\mathcal{D})$ is a measurable function $\tau^i : H^i(M) \to \mathcal{P}(A_0^i)$ such that $\tau^i(h_n)[A^i(s_n)] = 1$ for every $h_n \in H(M)$. A *strategy profile* $\tau = (\tau^i)_{i\in I}$ (or simply a *profile*) is a vector of strategies, one for each player.

In the sequel, σ always refers to correlated profiles in the game G, and τ refers to (non-correlated) profiles in the extended game $G(\mathcal{D})$.

For every history $(s_1, m_1, a_1, \ldots, s_n, m_n) \in H(M)$ we denote

$$\tau(s_1, m_1, a_1, \ldots, s_n, m_n) = (\tau^i(s_1, m_1^i, a_1, \ldots, s_n, m_n^i))_{i\in I}.$$

By Ionescu-Tuclea Theorem, every autonomous correlation device $\mathcal{D}$, every profile τ in $G(\mathcal{D})$ and every finite history $h_n \in H(M)$ induce a probability measure $\mathbf{P}_{h_n,\mathcal{D}\tau}$ over $H_\infty(M)$. We denote expectation w.r.t. this measure by $\mathbf{E}_{h_n,\mathcal{D},\tau}$. Define for every finite history $h_n \in H(M)$, the expected payoff w.r.t. τ by

$$\gamma_{\mathcal{D}}^i(h_n, \tau) = \mathbf{E}_{h_n,\mathcal{D},\tau} u^i(s_1, a_1, \ldots).$$

DEFINITION 4.2 *A payoff vector $g \in \mathbf{R}^I$ is an* extensive form correlated ϵ-equilibrium payoff *(resp. correlated ϵ-equilibrium payoff) if there exists an autonomous correlation device $\mathcal{D}$ (resp. a correlation device $\mathcal{D}$) and a strategy profile τ in $G(\mathcal{D})$ such that for every player $i \in I$ and every strategy τ'^i of*

player i in $G(\mathcal{D})$,

$$\gamma^i_{\mathcal{D}}(s_1, \tau) \geq g^i - \epsilon \geq \gamma^i_{\mathcal{D}}(s_1, \tau^{-i}, \tau'^i) - 2\epsilon.$$

DEFINITION 4.3 *A payoff vector $g \in \mathbf{R}^I$ is an* extensive form correlated equilibrium payoff *(resp. correlated equilibrium payoff) if it is the limit of extensive form correlated ϵ-equilibrium payoffs (resp. correlated ϵ-equilibrium payoffs) as ϵ goes to 0.*

4.2 Equivalence Results

The main result of this section is:

THEOREM 4.4 *The set E_0 of rational payoffs coincides with the set of extensive form correlated equilibrium payoffs.*

The theorem follows from the following two propositions, that are proved in the next section. Proposition 4.5 implies that every extensive form correlated equilibrium payoff is a rational payoff, while Proposition 4.6 implies the converse.

PROPOSITION 4.5 *Let $\epsilon > 0$. For every autonomous correlation device $\mathcal{D}$ and every ϵ-equilibrium profile τ in $G(\mathcal{D})$ there exists a correlated profile σ such that (i) $\gamma_{\mathcal{D}}(s_1, \tau) = \gamma(s_1, \sigma)$, and (ii) $U^i(\sigma) \leq \epsilon$ for every $i \in I$.*

The intuition of Proposition 4.5 is as follows. The autonomous correlation device $\mathcal{D}$ induces a probability distribution over plays, and therefore a correlated profile σ. Since, by definition, σ induces the same probability distribution over plays, the proposition follows.

PROPOSITION 4.6 *For every correlated profile σ and every $\epsilon > 0$ there exists an autonomous correlation device $\mathcal{D}$ and a profile τ in $G(\mathcal{D})$ such that (i) $\gamma_{\mathcal{D}}(s_1, \tau) = \gamma(s_1, \sigma)$, and (ii) $\gamma^i_{\mathcal{D}}(s_1, \tau^{-i}, \tau'^i) \leq \gamma^i_{\mathcal{D}}(s_1, \tau) + U^i(\sigma) + \epsilon$ for every player $i \in I$ and every strategy τ'^i of player i in $G(\mathcal{D})$.*

The intuition here is to construct an autonomous correlation device that *mimics* the profile σ: at every stage it chooses an action combination according to the probability distribution given by σ, and it sends each player the action that he should play. To deter deviations, the device reveals, at each stage, the actions it recommended to *all* players in the previous stage. This

way any deviation is detected immediately, and can be punished by the other players.

The only difficulty here is a measure theoretic one: how can one mimic a profile σ when the state and action spaces are general.

When the punishment level of each player is constant over the space of finite histories, a stronger result holds. In such a case, no correlation is needed *along* the play in order to sustain any rational payoff as a correlated equilibrium payoff.

THEOREM 4.7 *If for every player $i \in I$, $v^i_{h_n}$ is independent of $h_n \in H$, then the set of rational payoffs coincides with the set of correlated equilibrium payoffs.*

A result with similar flavor was proved by Forges (1988b) for one-shot games with incomplete information.

The intuition of Theorem 4.7 is as follows. If the expected irrationality of some correlated profile is small, than it must be small after "most" of the possible realized histories. Since the individually rational level is constant, that means that after most histories, the expected payoff of the players is, up to some ϵ, independent of the history. In particular, one can choose a pure profile before start of play, and transmit it to everyone. With high probability, no player can profit too much by deviating at any stage.

Note that Theorem 4.4 implies that the set of equilibrium payoffs cannot increase if we allow players to send private messages to the correlation device. The reason is that whatever be the extension under discussion, as long as (i) the outcome of the coin used by any player is private, (ii) all actions are played simultaneously, and (iii) each player is informed of past play, any equilibrium payoff is a rational payoff. Indeed, assume that g is an equilibrium payoff that is not rational. Then, for every $\epsilon > 0$ there is a profile τ_ϵ that is an ϵ-equilibrium, and its payoff is ϵ-close to g. Fix $\epsilon > 0$. The profile τ_ϵ induces a correlated probability distribution σ_ϵ over the space of infinite histories. Since g is not rational, the expected irrationality of σ_ϵ for some player i is strictly more than ϵ. It follows that player i could profit more than ϵ by deviating from σ^i_ϵ: there is a stopping time t such that if player i deviates whenever t stops, and defends his punishment level thereafter, he profits on average more than ϵ. Since the players observe past play, player i can deviate from τ_ϵ whenever t stops. Since the outcome of the coin he uses is private,

and since actions are played simultaneously, he can defend his punishment level in the extended game as well.

Remark: Though uniform equilibrium payoffs (see, e.g., Mertens, Sorin and Zamir (1994)) are not in the scope of our model (since the uniform equilibrium payoff cannot be defined as a limit of ϵ-equilibrium payoffs using some utility function) similar results can be obtained, with analogous proofs.

4.3 An Example

In this subsection we present an example of a two-player two-stage game. We find that this game has a unique correlated equilibrium payoff, and that it has an extensive form correlated equilibrium payoff that pareto dominates the unique correlated equilibrium payoff. The autonomous correlation device that we use illustrates the structure of the devices that are used in the proof of Proposition 4.6.

Consider the following two-player two-stage game:

stage 1

	L	C	R
T	$1,1$	$1,0$	$0,2$
B	**2**	$0,0$	$1,-4$

stage 2

L	R
$3,-1$	$0,-2$

Figure 1

One can verify that the unique Nash equilibrium of the game is:

- At stage 1, player 1 plays $(1/2, 1/2)$ and player 2 plays $(1/3, 2/3, 0)$.
- If the game reaches stage 2, player 2 plays L.

The corresponding equilibrium payoff is $(1, 0)$. Moreover, the unique correlated equilibrium coincides with the probability distribution over the entries of the matrices induced by this Nash equilibrium.

Consider now an extended game that includes an autonomous correlation device. The extended game is played as follows:
Stage 1A: the device chooses two signals, and sends one signal to each player.
Stage 1B: the players choose simultaneously actions for stage 1 of the original

game.
If the players chose (B, L), then:
Stage 2A: the device chooses a signal, which may depend on the previous signals that it chose, and sends it to player 2.
Stage 2B: player 2 chooses an action for stage 2 of the original game.

We claim that any point on the interval $(3/2, 1/2)$-$(2, 0)$ is an equilibrium payoff in the extended game, for a suitably defined autonomous correlation device. In particular, both players can profit by using such a device.

Indeed, let $x \in [0, 1]$, and consider the following device:

1. At stage 1A, the device chooses (T, L) with probability x and (B, L) with probability $1 - x$, and sends to each player his element in the chosen pair.

2. At stage 2A, the device sends its choice of stage 1A to player 2 (that is, it reveals its previous recommendation to player 2).

It is easy to verify that if $1/2 \leq x \leq 3/4$ then the following pair of strategies form a Nash equilibrium in the extended game, that yields the players an expected payoff $(3 - 2x, 2x - 1)$:

- At stage 1B, the players follow the signal they received at stage 1A.

- At stage 2B, player 2 plays L if player 1 followed the recommendation of the device at stage 1A, and plays R otherwise.

This device has the features that we will see in the proof of Proposition 4.6.

1. The device chooses at every stage a recommended action to each player, according to some known joint distribution, and sends to each player the action he is supposed to play.

2. In addition, the device reveals his recommendations for all the players at the previous stage.

3. The players are required to follow the recommendation of the device.

4. Since the recommendation becomes public after one stage, a deviation is detected immediately and is punished by his punishment level.

5 Proofs of the Equivalence Theorems

5.1 Representing Correlated Profiles as Autonomous Devices

In this subsection we develop some measure theoretic results that are needed to prove Proposition 4.6.

Given a correlated profile σ, we have to define an autonomous correlation device that mimics it. That is, a device that will recommend, at every stage, an action combination according to the probability distribution given by σ. Since the device is autonomous, it cannot base its choice on the actual play. However, for every realized play, σ may indicate a different probability distribution over action combinations. Thus, one needs to choose at stage n a recommended action combination for *every* possible history of length n. The players, who observe the realized history, can choose the recommended action that corresponds to that history, and disregard all other recommendations.

Since the setup is general, the space H_n of histories of length n may be uncountable, hence one cannot choose each recommendation independently. However, there is no need to choose the recommendations independently. As long as the recommendations at stage n are independent from the recommendations of previous stages, the distribution on plays will be equal to the one induced by σ.

The goal of this subsection is to prove the following result.

PROPOSITION 5.1 *Let $\sigma : H \to \mathcal{P}(A_0)$ be a correlated profile. Then there exists a sequence $(Y_n)_{n\in\mathbf{N}}$ of i.i.d r.v. uniformly distributed over $[0,1]$, and a measurable function $\delta_n : H_n \times [0,1]$, such that for every $h_n \in H$ and every measurable subset $C \subseteq A_0$,*

$$\sigma(h_n)[C] = \mathbf{P}(\delta_n(h_n, Y_n) \in C).$$

In words, the Proposition asserts that for every correlated profile σ and for every $n \in \mathbf{N}$ there exists a *countable* collection of i.i.d. r.v.s and a measurable function $\delta_n : H_n \times [0,1] \to A$ that represent $\sigma(h_n)$. That is, the probability that $\delta_n(h_n, \cdot)$ is in some set $C \subseteq A_0$ is equal to $\sigma(h_n)[C]$.

Proposition 5.1 readily follows from the following lemma.

LEMMA 5.2 *Let H be a measurable space, let X be a complete separable metric space, and let $\mathcal{X}$ be the σ-algebra of Borel subsets of X. Let $\mu : H \to$*

$\mathcal{P}(X)$ be measurable. Let Y be a r.v. uniformly distributed over $[0,1]$. Then there exists a measurable function $\delta : H \times [0,1] \to X$, such that

$$\mathbf{P}(\delta(h,Y) \in C) = \mu(h)[C] \qquad \forall h \in H, C \in \mathcal{X}. \tag{1}$$

Proof: We first deal with the case that X is at most countable. Denote $X = (x_n)_{n=1}^N$, where $N = |X|$ can be equal to $+\infty$. Let Y be a r.v. uniformly distributed over $[0,1]$. Define

$$\delta(h,Y) = \min\left\{k \mid \sum_{n=1}^{k} \mu(h)[x_n] \geq Y\right\}.$$

Note that δ is measurable. Eq. (1) holds, since for every $n = 1, \ldots, N$, $\mathbf{P}(\delta(h,Y) = x_n) = \mu(h)[x_n]$.

Assume now that X is uncountable. Since X is complete, separable and metric, it is isomorphic to $([0,1], \mathcal{B})$, where $\mathcal{B}$ is the collection of Borel subsets of $[0,1]$ (see, e.g., Parthasarathy 1967, Theorems 2.8 and 2.12). Hence, it is sufficient to prove the Lemma for the case $(X, \mathcal{X}) = ([0,1], \mathcal{B})$.

We shall now define the function $\delta : H \times [0,1] \to [0,1]$:

$$\delta(h,y) = \sup\{x \in [0,1] \mid \mu(h)[0,x] \leq y\}.$$

Note that δ is measurable. Indeed, for every fixed $x \in [0,1]$,

$$\begin{aligned} \{(h,y) \mid \delta(h,y) > x\} &= \{(h,y) \mid \mu(h)[0,x] < y\} \\ &= \cup_{q \in \mathbf{Q} \cap [0,1]} \{h \mid \mu(h)[0,x] < q\} \times [q,1]. \end{aligned}$$

Since a countable union of measurable sets is measurable, and since μ is measurable, δ is measurable.

Let Y be a r.v. uniformly distributed over $[0,1]$. Then for every $h \in H$ and every $x \in [0,1]$,

$$\mathbf{P}(\delta(h,Y) \leq x) = \mu(h)[0,x].$$

Since the intervals $\{[0,x], x \in [0,1]\}$ generate the Borel σ-algebra, it follows that for every $C \in \mathcal{B}$,

$$\mathbf{P}(\delta(h,Y) \in C) = \mu(h)[C],$$

as desired. ■

5.2 Standard Revealing Devices

We will be interested in a class of autonomous correlation devices, which we call *standard revealing devices*. Those devices have three special features: (i) they choose an element in $[0,1]$ according to the uniform distribution, (ii) the private signal space at stage n of each player $i \in I$ is the space of universally measurable functions from H_n to A^i, and (iii) at stage $n+1$ they publicly announce the signals that were sent at stage n.

DEFINITION 5.3 *A* standard revealing autonomous correlation device $\mathcal{D}$ *is given by a sequence* $(\delta_n)_{n\in\mathbf{N}}$ *of measurable functions, where* $\delta_n : H_n \times [0,1] \to A$, *such that for every* $y \in [0,1]$, *and every* $h_n \in H_n$, $\delta_n(h_n, y) \in A(s_n)$.

A standard revealing device chooses, at every stage $n \in \mathbf{N}$, an element $Y_n \in [0,1]$ according to the uniform distribution, and then sends to each player $i \in I$ a pair $m_n^i = (m_{n-1}, \delta_n^i(\cdot, Y_n))$, where $m_{n-1} = (m_{n-1}^i)_{i\in I}$ is the vector of signals sent at the previous stage, and $\delta_n^i(\cdot, Y_n) : H_n \to A^i$. $\delta_n^i(h_n, Y_n)$ can be interpreted as a recommended action for player i if the realized history up to stage n is h_n.

Since δ_n is measurable, it follows by Theorem III.23 in Castaing and Valadier (1977) that δ_n^i is universally measurable, for every player $i \in I$.

Note that a standard revealing device is in particular an autonomous correlation device. Indeed, fix $n \in \mathbf{N}$. Every $x \in [0,1]$ defines a function $\delta_n(\cdot, x) : H_n \to A^i$. Let $M_n'^i$ be the space of all these functions. The Borel measurable structure of $[0,1]$ induces a measurable structure on $M_n'^i$, and the uniform distribution over $[0,1]$ induces a probability distribution ν_n over $M_n'^i$. Finally, the signal space of player i at stage n is $M_n^i = M_{n-1} \times M_n'^i$, where $M_{n-1} = \times_{i\in I} M_{n-1}^i$, and the distribution over M_n^i is $\{m_{n-1}\} \otimes \nu_n$.

5.3 The Proofs

Proof of Proposition 4.5:
Let $\epsilon > 0$, let $\mathcal{D}$ be an autonomous correlation device, and let τ be an ϵ-equilibrium profile in $G(\mathcal{D})$.

Recall that $\mathbf{P}_{s_1,\mathcal{D},\tau}$ is the probability distribution over the space H_∞ of infinite histories induced by $\mathcal{D}$ and τ. Let σ be a correlated profile that induces the same distribution over H_∞; that is, $\sigma(h_n)[C] = \mathbf{P}_{h_n\mathcal{D},\tau}(a_n \in C)$ for every measurable subset $C \subseteq A_0$. By definition, $\gamma_{\mathcal{D}}(s_1, \tau) = \gamma(s_1, \sigma)$.

We shall now prove that $U^i(\sigma) \leq \epsilon$ for every $i \in I$. Otherwise, there exists a player $i \in I$, and a stopping time t such that $\mathbf{E}_{t,\sigma} U^i(h_t, \sigma, a_t) > \epsilon + \rho$, for some $\rho > 0$. Define a strategy τ'^i for player i in $G(\mathcal{D})$ as follows. Follow τ^i until t. Afterwards, play a strategy that maximizes (up to ρ) your payoff against τ^{-i} given h_t.

It is easy to verify that

$$\gamma^i_{\mathcal{D}}(s_1, \tau^{-i}, \tau'^i) \geq \gamma^i(s_1, \sigma) + \mathbf{E}_{t,\sigma} U^i(h_t, \sigma, a_t) - \rho \geq \gamma^i_{\mathcal{D}}(s_1, \tau) + \epsilon,$$

a contradiction. ■

Proof of Proposition 4.6:
Let σ be a correlated profile and let $\epsilon > 0$. By Proposition 5.1, there exists a countable sequence $(Y_n)_{n \in \mathbf{N}}$ of i.i.d. r.v.s, uniformly distributed over $[0,1]$, and a measurable function $\delta_n : H_n \times [0,1] \to A$ such that for every $h_n \in H_n$ and every measurable subset C of A_0,

$$\sigma(h_n)[C] = \mathbf{P}(\delta_n(h_n, Y_n) \in C). \tag{2}$$

Consider the autonomous correlation device defined by $(\delta_n)_{n \in \mathbf{N}}$. Thus, at each stage n, an element $Y_n \in [0,1]$ is chosen according to the uniform distribution, and each player i receives the function $\delta^i_n(h_n, Y_n) : H_n \to A^i$.

Define a profile τ in $G(\mathcal{D})$ as follows. At every stage n, each player is checked whether his realized action at stage $n-1$ coincides with the recommendation of the device $\delta^i_{n-1}(h_{n-1}, Y_{n-1})$ (which is revealed at stage n). If at least one player deviated, then the deviator who has a minimal index is punished, from that stage on, with an ϵ-punishment correlated strategy profile forever. Otherwise, each player i plays at stage n the action $\delta^i_n(h_n, Y_n)$, where h_n is the realized history until stage n.

Note that we have not specified how, once a deviator is detected, his opponents correlate their actions. This can be done by the following procedure. Before the start of play, the device chooses, for every player $i \in I$, a sequence of i.i.d. r.v.s $(W^i_n)_{n \in \mathbf{N}}$ with $\mathbf{P}(W^i_n = 1) = 1/2$. The device then sends the sequence (W^i_n) to all players *except* player i. If the necessity arises, players $N \setminus \{i\}$ use the sequence (W^i_n) to correlate their moves and follow an ϵ-punishment correlated strategy $\tilde{\sigma}^{-i}_\epsilon$ against player i.

It is easy to verify that $\mathbf{P}_{s_1,\mathcal{D},\tau} = \mathbf{P}_{s_1,\sigma}$, and therefore $\gamma_{\mathcal{D}}(s_1, \tau) = \gamma(s_1, \sigma)$.

We shall now show that $\gamma^i_{\mathcal{D}}(s_1, \tau^{-i}, \tau'^i) \leq \gamma^i_{\mathcal{D}}(s_1, \tau) + U^i(\sigma) + \epsilon$. Indeed, let τ'^i be a strategy of player i in $G(\mathcal{D})$. Let t be the stopping time defined

by

$$t = \min\{n \in \mathbf{N} \mid a_n \neq \delta_n^i(h_n, Y_n)\} + 1.$$

Then, under t^{-i}, at stage t players $N \setminus \{i\}$ switch to a punishment profile against player i. In particular,

$$\gamma_{\mathcal{D}}^i(s_1, \tau^{-i}, \tau'^i) \leq \gamma^i(\sigma) + U^i(\sigma) + \epsilon,$$

as desired. ■

Proof of Theorem 4.7:
Assume now that for every fixed player $i \in N$, $v_{h_n}^i$ is independent of $h_n \in H$, and denote this common value by v^i.

In view of Theorem 4.4, it suffices to prove that every rational payoff is a correlated equilibrium payoff.

Fix $\epsilon > 0$. We denote by P^i the space of *pure* strategies of player i, and $P = \times_{i \in N} P^i$. Every correlated profile σ induces a probability measure over P. This probability measure is also denoted by σ.

Let σ be a correlated profile such that $U^i(\sigma) < \epsilon$ for each player $i \in N$. For every $\delta > 0$, denote by H_∞^δ the set of all histories $h_\infty \in H_\infty$ such that $\gamma^i(h_n, \sigma) < v^i - \delta$ for some beginning h_n of h_∞.

Since $U^i(\sigma) < \epsilon$, and since $v_{h_n}^i$ is independent of h_n for every $i \in I$, it follows that $\mathbf{P}_{s_1,\sigma}(H_\infty^{\sqrt{\epsilon}}) \leq \sqrt{\epsilon}$.

Define a correlation device $\mathcal{D}$ with a signal space $M^i = P$ for each player i. The device chooses a *pure* profile according to σ, and reveals to all the players the profile that was chosen. The players are then requested to follow the pure profile that was chosen by the device. A deviator, who will be noticed immediately, will be punished with his punishment level, which is independent of the history.

There is one technical difficulty we have ignored so far: how to choose a pure profile in P? To do this one needs to impose a measurable structure on the space of pure profiles. Note that each realization of the sequence (Y_n) that was defined in the proof of Proposition 4.6 defines a pure strategy profile. Thus, the measurable structure is the one induced by the mapping that maps $[0,1]^{\mathbf{N}}$ to the space of pure profiles. The measure on P is the one induced by the uniform distribution over $[0,1]^{\mathbf{N}}$ (that is, the infinite product of independent copies of the uniform distributions over $[0,1]$). This measure induces the same expected payoff for the players as σ, for every finite history h_n. Formally, denote by π the correlated profile that corresponds to the

uniform distribution over $[0,1]^{\mathbf{N}}$. Then for every $h_n \in H$ and every player $i \in I$,

$$\gamma^i(h_n, \pi) = \gamma^i(h_n, \sigma).$$

With probability greater than $1 - \sqrt{\epsilon}$, $\gamma^i(h_n, \sigma) > v^i_{h_n} - \sqrt{\epsilon}$ for every $n \in \mathbf{N}$, hence no player can profit more than $\sqrt{\epsilon}$. In particular, this profile is a $((1-\sqrt{\epsilon})\sqrt{\epsilon} + \sqrt{\epsilon}R)$-equilibrium in $G(\mathcal{D})$ (recall that R is a bound of u). ■

6 Communication and Correlation

In this section we show that if the action set is finite and there are at least 4 players, then communication can substitute correlation; that is, if we extend the original stochastic game by allowing the players to send countably many private messages at every stage between each other, then the set of equilibrium payoffs in the extended game coincides with the set E_0 of rational payoffs.

Since the proof is a simple application of a result due to Bárány (1992), and a result with a similar flavor was proved by Forges (1988b) for games with incomplete information, we will only sketch the ideas.

A *direct communication protocol* is a finite sequence of rules, known to all the players, specifying what players should do. Each rule has the form: Some player i should choose a message from his message space according to some probability distribution, that may depend on previous messages he sent or received, and send it to some other player i'.

Bárány (1992) has proved the following theorem:

THEOREM 6.1 (BÁRÁNY (1992)) *Consider a one shot game with player set I and action sets $(A^i)_{i\in I}$. If $|I| \geq 4$ and if A is finite, then every correlated equilibrium payoff that is supported by a rational joint probability distribution over A is a Nash equilibrium payoff in some extended game, that has a pre-play direct communication phase, in which players follow a direct communication protocol.*

The protocol devised by Bárány (1992) can be extended so that if a player deviates from the rules of the protocol in a way that affects the payoff, his identity is revealed. This feature is necessary for effective punishment.

A stochastic game with *direct communication* is a stochastic game where every stage contains two sub-stages: a direct communcation sub-stage and a decision sub-stage. At the direct communication sub-stage, players send messages according to a given direct communication protocol, while at the decision sub-stage, each player chooses an action, as a function of past play (where play here includes realized states, realized actions and messages sent and received by that player).

An equilibrium payoff in the extended game is a *direct communication equilibrium payoff*.

The main result of this section is:

THEOREM 6.2 *If $|I| \geq 4$ and A is finite, the set E_0 of rational payoffs coincides with the set of direct communication equilibrium payoffs.*

Proof: The proof that every communication equilibrium payoff is a rational payoff follows the same lines as the proof of Proposition 4.5.

We shall now prove the converse. If $h \mapsto \mu(h) \in \mathcal{P}(A)$ is measurable, and if each $\mu(h)$ is rational, then the function that assigns to each $h \in H$ the protocol devised by Bárány is also measurable (where the space of protocols is equipped with the discrete topology).

Let now σ be a correlated profile such that $U^i(\sigma) < \epsilon$ for every $i \in I$. Choose a measurable function $\sigma' : H \to \mathcal{P}(A)$ such that for every $h_n \in H$ (i) $\sigma'(h_n) \in \mathcal{P}(A(s_n))$, (ii) $\sigma'(h_n)$ is rational, and (iii) $\|\sigma'(h_n) - \sigma(h_n)\| < \epsilon/2^n$. It then follows that $U^i(\sigma') < 2\epsilon$ for every $i \in I$.

Define now a strategy in the extended stochastic game as follows. At the direct communication sub-stage, follow the protocol devised by Bárány for $\sigma'(h_n)$, and at the decision sub-stage, play the action suggested by the protocol given the messages you have sent and received during the last direct communication sub-stage. If a deviation from the direct communication protocol is detected, the deviator is identifies and punished by his punishment level.

It is easy to verify that this strategy profile is a 2ϵ-equilibrium. ■

Comment: If the transition rule q is norm continuous, the action space $A(s)$ is compact metric for every $s \in S$, and the payoff function $u^i(s, a)$ is continuous over A for every fixed $s \in S$, then one can find, for every $\delta > 0$, a finite approximation of $A(s)$; that is, for every player i a finite set $B^i_\delta(s) \subset A^i(s)$ such that for every $a^i \in A^i(s)$ there exists $b^i \in B^i_\delta(s)$ with $d(a^i(s), b^i(s)) < \delta$.

Moreover, one can choose $B^i_\delta(s)$ so that the correspondence $(s,\delta) \mapsto B^i_\delta(s)$ is measurable. Denote $B_\delta(s) = \times_{i\in I} B^i_\delta(s)$.

For every $\epsilon > 0$, every $s \in S$ and every distribution $\mu \in \mathcal{P}(A(s))$, there exists $\delta > 0$ and a distribution $\nu \in \mathcal{P}(B(s))$ such that $\|\mathbf{E}_\mu u(s,a) - \mathbf{E}_\nu u(s,a)\| < \epsilon$ (continuity of u) and $\|\mathbf{E}_\mu q(s,a) - \mathbf{E}_\nu q(s,a)\| < \epsilon$ (norm-continuity of q).

Thus, under these continuity conditions, Theorem 6.2 holds.

References

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

[2] Aumann R. J. (1992) "Irrationality in Game Theory," *Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn*, P. Dasgupta, D. Gale, O. Hart and E. Maskin (eds), 214-227, The MIT Press, Cambridge and London.

[3] Bárány I. (1992) "Fair Distribution Protocols or How the Players Replace Fortune," *Mathematics of Operations Research*, 17, 327-340.

[4] Blackwell D. and Ferguson T. S. (1968) "The Big Match," *The Annals of Math. Stat.*, 39, 159-163.

[5] Castaing C. and Valadier M. (1977) "Convex Analysis and Measurable Multifunctions," Springer-Verlag, Berlin Heidelberg New-York.

[6] Dellacherie C. and Meyer P.-A. (1978) "Probabilities and Potential," North-Holland, amsterdam.

[7] Forges F. (1986) "An Approach to Communication Equilibria," *Econometrica*, 54, 1375-1385.

[8] Forges F. (1988a) "Communication Equilibria in Repeated Games with Incomplete Information," *Mathematics of Operation Research*, 13, V. 2, 77-117.

[9] Forges F. (1988b) "Universal Mechanisms," *Econometrica*, 58, 1341-1364.

[10] Maitra A., and W. Sudderth (1993) "Borel Stochastic Games with Limsup Payoff," *Annals of Probability*, 21, 861-885.

[11] Maitra A., and W. Sudderth (1998) "Finitely Additive Stochastic Games with Borel Measurable Payoffs," *International Journal of Game Theory*, 27, 257-267.

[12] Mertens J.F., and Sorin S., and S. Zamir (1994) "Repeated Games," CORE Discussion Paper 9421.

[13] Myerson R.B. (1986) "Multistage Games with Communication," *Econometrica*, 54, 323-358.

[14] Neveu J. (1965): *Mathematical Foundations of the Calculus of Probability*, Holden-Day.

[15] Nowak A.S. (1991) "Existence of Correlated Weak Equilibria in Discounted Stochastic Games with General State Space," in *Stochastic Games and Related Topics*, ed. by T.E.S. Raghavan et al., Kluwer Academic Press.

[16] Parthasarathy K.R. (1967) "Probability Measures on Metric Spaces", Academic Press, New York

[17] Solan E., and N. Vieille (1998) "Correlated Equilibrium in Stochastic Games," *Games and Economic Behavior*, to appear