Int J Game Theory (2001) 30:259–277

Characterization of correlated equilibria in stochastic games

Eilon Solan†

Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, *and* the School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel. (e-mail: eilons@post.tau.ac.il)

Final version July 2001

Abstract. A general communication device is a device that at every stage of the game receives a private message from each player, and in return sends a private signal to each player; the signals the device sends depend on past play, past signals it sent, and past messages it received.

An autonomous correlation device is a general communication device where signals depend only on past signals the device sent, but not on past play or past messages it received.

We show that the set of all equilibrium payoffs in extended games that include a general communication device coincides with the set of all equilibrium payoffs in extended games that include an autonomous correlation device. A stronger result is obtained when the punishment level is independent of the history.

Key words: Stochastic games, general communication device, extensive form correlated equilibrium, correlation.

1. Introduction

In the present paper we consider stochastic games that are extended by introducing a *general communication device*. Each stage in the extended game is composed of four sub-stages. First, the players send private messages to the device. Second, the device, as a function of past play, past messages it received and past signals it sent, sends a private signal to each player. Third, each player chooses an action, independently of his opponents. Finally, a new state

† I thank Francoise Forges, Rakesh Vohra, the associate editor and two anonymous referees for helpful comments which substantially improved the paper.

is chosen according to a probability distribution that depends on the current state and on the action combination chosen at that stage. The payoff of each player is some function of the infinite play (e.g. the discounted or the undiscounted evaluation of daily payoffs).

We assume that players observe past play, but no player observes messages sent by other players to the device, or signals sent from the device to other players. Moreover, we assume players have perfect recall.

Special cases of general communication devices are communication devices (where new signals depend on past messages and past signals but not on past play), autonomous correlation devices (where new signals depend on past signals, but not on past play, and players do not send messages to the device), correlation devices (where the device sends only one signal before start of play), pre-play communication between players (where players can exchange messages before start of play), direct communication between players along the play (where players can exchange messages along the play), and a mediator (who can send private signals to the players along the play as a function of past play and past signals it sent).

Each of the above mentioned cases allows for a different level of correlation between the players, and therefore the sets of equilibrium payoffs in these extensions may differ.

Denote by E_1 the set of all equilibrium payoffs in extended games that include a general communication device, by E_2 the set of all equilibrium payoffs in extended games that include an autonomous correlation device, and by E_3 the set of all equilibrium payoffs in extended games that include a correlation device. It is clear that $E_1 \supseteq E_2 \supseteq E_3$.

The main result of the present paper is that $E_1 = E_2$ – by introducing communication between the players and the correlation device, or a mediator who can observe past play (or both), one cannot enlarge the set of equilibrium payoffs.

When the punishment level of the players is independent of the history, a stronger result can be obtained: $E_1 = E_3$. Thus, in this class of games introducing correlation after the first stage cannot enlarge the set of equilibrium payoffs.

The two equivalence 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.

To prove the main result we define a set E_0 of all *rational payoffs*, and we prove that $E_1 \subseteq E_0 \subseteq E_2$. Since $E_2 \subseteq E_1$, it follows that $E_1 = E_2 = E_0$. The definition of rational payoffs is related to Aumann's (1992) measure of irrationality. We discuss the relation between the two concepts below. Thus, our results can be viewed as a generalization of Aumann's (1987) characterization of correlated equilibrium payoffs in one-shot games to stochastic games, as well as a generalization of the "Folk Theorem".

The paper does not address the question of whether the set E_2 $(= E_1 = E_0)$ is empty or not. Moreover, non-emptiness of this set is known only in special cases, and in all but one, non-emptiness is proved by showing the existence of an equilibrium payoff or of a correlated equilibrium payoff (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 equilibria in discounted games). The only exception is stochastic games with finitely many states and actions, and the lim sup evaluation, where non-emptines of E_0 is proved directly (see Solan and Vieille, 1998). Therefore, our result implies that in this class of games the set E_2 is not empty as well. The existence of an equilibrium payoff or of a correlated equilibrium payoff in this class of games has not been proved yet.

The main assumption that the results hinges on is that, except of the messages sent to/by the device, players have symmetric information. When information is not symmetric, players may profit by communicating and exchanging their information, and therefore E_2 can be a strict subset of E_1 (see, e.g., Forges (1986, Remark 3)). A related result is Forges (1985, Theorem 2) which implies that in repeated games with incomplete information, where the actions of the informed player do not affect the payoff, and the only role of this player is to transmit information, the set E_3 of (normal form) correlated equilibrium payoffs coincides with the set E_2 of extensive form correlated equilibrium payoffs.

The paper is arranged as follows. The model is presented in section 2. In section 3 we introduce general communication devices, and we state the main results. In section 4 we define the notion of the "measure of irrationality", and we discuss its relation to Aumann's (1992) definition of a similar concept. In section 5 we sketch the main ideas of the proofs, and in section 6 we study an example. The proofs of the main results appear in section 7.

2. The model

For every measurable space Y we denote by $\mathscr{P}(Y)$ the space of probability measures over Y. If $\mu \in \mathscr{P}(Y)$ and $C \subseteq Y$ is a measurable set, then $\mu[C]$ is the measure of C under μ. A function $f : X \to \mathscr{P}(Y)$ 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 X-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 ϕ from a measurable space X to a topological space 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_∞:

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

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 $\mathscr{P}(S)$.
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* play 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 $H_n = \{s_1\} \times (A \times S)^{n-1}$. 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 $H = \bigcup_{n \in \mathbf{N}} H_n$. 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* $\sigma^i : H \to \mathscr{P}(A_0^i)$ *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* $\sigma : H \to \mathscr{P}(A_0)$ *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 $\sigma^{-i} : H \to \mathscr{P}(A_0^{-i})$ 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-Tulcea 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, every player $i \in I$ 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 for player i under σ in the subgame beginning with h_n. The payoff for player i under σ in the subgame beginning with (h_n, a_n), $\gamma^i((h_n, a_n), \sigma)$, is defined analogously. 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:

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

$v^i_{h_n}$ 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. The quantity $v^i_{h_n, a_n}$ is defined analogously.

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

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

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

$$\gamma^i(h_n, (\sigma^{-i}, \sigma^i)) > v^i_{h_n} - \varepsilon \quad \forall h_n \in H_n.$$

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, there are no measurability issues, and (ii) if the utility function is the discounted sum or the lim sup 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 lim sup).

3. General communication devices

In the present section we define general communication devices, and the game extended by a general communication device.

Definition 3.1. *A* general communication device $\mathscr{C}$ *is given by:*

- *For every player $i \in I$ and every $n \in \mathbf{N}$, a measurable space R^i_n of* messages *player i can send the device at stage n. Let $R_n = \times_{i \in I} R^i_n$.*
- *For every player $i \in I$ and every $n \in \mathbf{N}$, a measurable space M^i_n of* signals *the device can send player i at stage n. Let $M_n = \times_{i \in I} M^i_n$.*
- *For every $n \in \mathbf{N}$, a measurable function $\mu_n : (\times^n_{j=1} R_j) \times (\times^{n-1}_{j=1} M_j) \times H_n \to \mathscr{P}(M_n)$.*

Given a stochastic game G and a general communication device $\mathscr{C}$, we define an extended game $G(\mathscr{C})$ as follows. At each stage n, (i) each player $i \in I$ sends the device a private message $r^i_n \in R^i_n$. Denote $r_n = (r^i_n)_{i \in I}$. (ii) The device chooses a message $m_n = (m^i_n)_{i \in I} \in M_n$ according to $\mu_n(r_1, \ldots, r_n, m_1, \ldots, m_{n-1}, h_n)$, where h_n is the realized history up to stage n. Each player i is then informed of m^i_n. (iii) Each player chooses an action $a^i_n \in A^i(s_n)$, and (iv) a new state $s_{n+1} \in S$ is chosen according to $q(s_n, a_n)$, where $a_n = (a^i_n)_{i \in I}$. Both the action combination a_n that was played and the new state s_{n+1} are publicly announced.

We assume that players have infinite recall, so each player i can base his choice of an action at stage n on past play $(s_1, a_1, \dots, s_n)$ and on past signals $(r_1^i, m_1^i, \dots, r_n^i, m_n^i)$ *he* has sent and received.

Two special classes of general communication devices will play special role in the paper: autonomous correlation devices and correlation devices.

Definition 3.2. *A general communication device* $\mathscr{C} = ((R_n^i, M_n^i)_{i \in I}, \mu_n)_{n \in \mathbf{N}}$ *is an* autonomous correlation device *if* μ_n *depends only on previous signals* $(m_1, \dots, m_{n-1})$*, and* not *on past messages* $(r_1, \dots, r_n)$ *sent to the device or on past play* h_n*. An autonomous correlation device is a* correlation device *if* M_n^i *is a singleton for every* $i \in I$ *and every* $n \geq 2$.

In other words, a general communication device is an autonomous correlation device if it is independent of the play: it does not observe past play, and players cannot influence its choices by sending it messages. In particular, the messages have no strategic effect, and there is no loss of generality in assuming that the sequence $(m_1, m_2, \dots)$ of signals is chosen before start of play. The device is a correlation device if it sends a signal to the players only once before play starts, and no correlation is done along the play.

Let $H^i(\mathscr{C})$ be the space of all finite histories that player i can observe in $G(\mathscr{C})$. Formally, $H^i(\mathscr{C}) = H_M^i(\mathscr{C}) \cup H_A^i(\mathscr{C})$, where

$$H_M^i(\mathscr{C}) = \{(s_1, r_1^i, m_1^i, a_1, \dots, s_{n-1}, r_{n-1}^i, m_{n-1}^i, a_{n-1}, s_n) \mid$$
$$a_k \in A(s_k), r_k^i \in R_k^i, m_k^i \in M_k^i\}, \quad \text{and}$$
$$H_A^i(\mathscr{C}) = \{(s_1, r_1^i, m_1^i, a_1, \dots, s_{n-1}, r_{n-1}^i, m_{n-1}^i, a_{n-1}, s_n, r_n^i, m_n^i) \mid$$
$$a_k \in A(s_k), r_k^i \in R_k^i, m_k^i \in M_k^i\}.$$

$H_M^i(\mathscr{C})$ is the collection of all finite histories player i can observe before he chooses a message, and $H_A^i(\mathscr{C})$ is the collection of all finite histories he can observe before he chooses an action. Note that, since the signals are private, each player observes a (possibly) different history. Let $H(\mathscr{C})$ be the space of all finite histories that an outside observer, who observes *both* the actions of the players and the signals received from and sent to *all* the players, can observe. Let $H_\infty(\mathscr{C})$ be the space of all infinite histories that this outside observer can observe. We endow $H_\infty(\mathscr{C})$ with the σ-algebra generated by all the finite cylinders. Note that the spaces $(H^i(\mathscr{C}))_{i \in I}$, $H(\mathscr{C})$ and $H_\infty(\mathscr{C})$ are independent of $(\mu_n)_{n \in \mathbf{N}}$.

A *strategy* of player i in $G(\mathscr{C})$ is measurable function $\tau^i : H^i(\mathscr{C}) \to \mathscr{P}(A_0^i) \cup (\bigcup_{n \in \mathbf{N}} \mathscr{P}(R_n^i))$ such that $\tau^i(h_n) \in \mathscr{P}(R_n^i)$ if $h_n \in H_M^i(\mathscr{C})$, and $\tau^i(h_n) \in \mathscr{P}(A_0^i)$ satisfies $\tau^i(h_n)[A^i(s_n)] = 1$ if $h_n \in H_A^i(\mathscr{C})$.

A *profile* $\tau = (\tau^i)_{i \in I}$ 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 $G(\mathscr{C})$.

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

$$\gamma_{\mathscr{C}}^i(h_n, \tau) = \mathbf{E}_{h_n, \mathscr{C}, \tau} u^i(s_1, a_1, \dots).$$

Definition 3.3. *A payoff vector $g \in \mathbf{R}^I$ is a* general correlated ε-equilibrium payoff *(resp.* extensive form correlated ε-equilibrium payoff, correlated ε-equilibrium payoff*) if there exists a general communication device $\mathscr{C}$ (resp. an autonomous correlation device, a correlation device) and a strategy profile τ in $G(\mathscr{C})$ such that for every player $i \in I$ and every strategy τ'^i of player i in $G(\mathscr{C})$,*

$$\gamma^i_{\mathscr{C}}(s_1, \tau) + \varepsilon \geq g^i \geq \gamma^i_{\mathscr{C}}(s_1, \tau^{-i}, \tau'^i) - \varepsilon.$$

Definition 3.4. *A payoff vector $g \in \mathbf{R}^I$ is a* general correlated equilibrium payoff *(resp.* extensive form correlated equilibrium payoff, correlated equilibrium payoff*) if it is the limit of general correlated ε-equilibrium payoffs (resp. extensive form correlated ε-equilibrium payoffs, correlated ε-equilibrium payoffs) as ε goes to 0.*

We denote by E_1 the set of all general correlated equilibrium payoffs, by E_2 the set of all extensive form correlated equilibrium payoffs, and by E_3 the set of all correlated equilibrium payoff. Note that these sets depend on s_1. It is clear that we have

$$E_1 \supseteq E_2 \supseteq E_3. \tag{1}$$

The main results of the paper are:

Theorem 3.5. $E_1 = E_2$.

Theorem 3.6. *If for every player $i \in I$, $v^i_{h_n}$ is independent of $h_n \in H$, then $E_1 = E_3$.*

As an example for a case where Theorem 3.6 applies, take a game where (i) the utility of each player is some Banach limit of daily payoffs, and (ii) in every state each player has a terminating action that punishes everybody (at a level of punishment which is independent of the state). Alternatively, instead of (ii) one can impose certain ergodicity conditions on the transitions (see, e.g., Nowak (1999a,b) and the references therein).

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. For more details, see Solan and Vieille (1998).

4. Expected irrationality

In this section we define a set E_0 of payoff vectors, which we call the set of *rational payoffs*.

To be more specific, we assign to each correlated profile σ and every player $i \in I$ a non-negative number $U^i(\sigma)$, which we call the *expected irrationality* of σ for player i. $U^i(\sigma)$ measures how much player i can profit by deviating from σ^i, provided his deviation is followed by an indefinite punishment. We call a payoff vector $g \in \mathbf{R}^I$ ε-rational if it is the payoff that corresponds to some profile σ such that $U^i(\sigma) < \varepsilon$ for every player i. A payoff vector is rational if it is the limit of ε-rational payoffs, as ε goes to 0.

We then see how our definition of expected irrationality relates to a similar notion defined by Aumann (1992) for one-shot games.

In section 7 we prove that $E_1 \subseteq E_0$, while $E_0 \subseteq E_2$, thereby, using (1), we prove Theorem 3.5. We also show there that when the punishment level is independent of the history, $E_0 \subseteq E_3$, thereby proving Theorem 3.6.

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 receives the signal a^i, $\sigma(h_n) \mid a^i$ is his conditional probability on the joint action played by his opponents.

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

$$U^i(h_n, \sigma, a^i) = \sup_{b^i \neq a^i} \mathbf{E}_{\sigma(h_n) \mid a^i}(v^i_{h_n, b^i, a^{-i}} - \gamma^i((h_n, a^i, a^{-i}), \sigma)). \tag{2}$$

This 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 non-positive if player i cannot profit, while it is positive and equal to his maximal profit, if such a profit is available.

Define the *expected irrationality of* σ for player i by

$$U^i(\sigma) = \sup_t \mathbf{E}_{s_1, \sigma}(U^i(h_t, \sigma, a_t) 1_{t < +\infty}), \tag{3}$$

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^i whenever he chooses. However, one stage afterwards, he is being punished at 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 (by choosing $t = +\infty$).

Definition 4.1. *Let $\varepsilon > 0$. 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) $U^i(\sigma) < \varepsilon$ 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.*

We denote the set of rational payoffs by E_0.

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

Thus, we rule out as irrational payoff vectors only those vectors that either (i) cannot be supported by correlated profiles, or (ii) can be supported by correlated 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.

---

[1] Formally, this is the disintegration of $\sigma(h_n)$ w.r.t. the function $f : A \to A^i$ that is defined by $f(a) = a^i$, projected on A^{-i} (see Dellacherie and Meyer, 1978). Since we require that $(\sigma(h_n) \mid a^i)$ $[A^{-i}(s_n)] = 1$, regular conditional probabilities do not suffice.

4.1. 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 Theorem 3.5.

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 3.6 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 6 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. 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.

4.2. Comparison with Aumann's notion of irrationality

Aumann (1974) studies one-shot games with incomplete information in a finite setup. Such games are given by (i) a set of players I, (ii) for each player i, a finite set of actions A^i, (iii) for each player i, a utility function $u^i : A \to \mathbf{R}$, where $A = \times_{i \in I} A^i$, (iv) a measure space $(\Omega, \mathscr{F}, P)$ of states of the world, and (v) for each player i, a sub-σ-algebra $\mathscr{F}_i$ of $\mathscr{F}$, which represents the information available to player i.

The game proceeds as follows. A state $\omega \in \Omega$ is chosen according to P. Each player i is informed of the sets in $\mathscr{F}_i$ that contain ω.[2] Then each player chooses an action $a^i \in A^i$, independently of his opponents, and receives the payoff $u^i(a)$, where $a = (a^i)_{i \in I}$.

[2] That is, for every $F_i \in \mathscr{F}_i$ player i is told whether $\omega \in F_i$ or whether $\omega \notin F_i$.

Thus, Aumann extends a game $G = (I, (A^i, u^i))$ with complete information to a game $G' = (G, \Omega, \mathscr{F}, P, (\mathscr{F}_i))$ with incomplete information.

In this setup, a strategy for player i is a $\mathscr{F}_i$-measurable function, and one defines equilibria in the usual way.

Aumann (1974) shows that (i) if each $\mathscr{F}_i$ is rich enough,[3] any equilibrium in the game G' induces a correlated equilibrium in the corresponding game G with complete information, and (ii) for any correlated equilibrium in a game $G = (I, (A^i, u^i))$ with complete information there is an information structure $(\Omega, \mathscr{F}, P, (\mathscr{F}_i))$ such that the distribution over A induced by one of the equilibria of the corresponding game $G' = (G, \Omega, \mathscr{F}, P, (\mathscr{F}_i))$ with incomplete information coincides with the original correlated equilibrium.

Aumann (1992) defines for every strategy profile $\sigma = (\sigma^i)_{i \in I}$ and every player $i \in I$ a number which he calls the *measure of irrationality*. It is given by the highest profit of player i by deviating from σ^i.[4]

In a sequential game, as the one studied in the present paper, players may acquire new information along the play. Hence the information structure should involve a measure space $(\Omega, \mathscr{F}, P)$, and, for each player i, an increasing sequence of sub-σ-algebras $(\mathscr{F}_i^n)_{n \in \mathbf{N}}$. $\mathscr{F}_i^n$ is the information available to player i at stage n. To allow players to randomize, we should require that $\mathscr{F}_i^n$ is sufficiently rich relative to $\bigvee_{j \neq i} \mathscr{F}_j^n$ for every $n \in \mathbf{N}$ and every player i.

Note that we assumed that $(\mathscr{F}_i^n)$ are independent of the play, though the model can be altered to allow this flexibility.

Let now G be a stochastic game with complete information, as defined in section 2, and $G' = (G, \Omega, \mathscr{F}, P, (\mathscr{F}_i^n))$ be the corresponding game with incomplete information.

For every player i and every profile τ in G', let $\hat{U}^i(\tau)$ be the maximal profit of player i by deviating from τ^i.

Every profile τ in G' induces a probability distribution over infinite plays, and therefore a correlated profile σ in G.

Define now $\hat{U}^i(\sigma)$ as the infimum of $\hat{U}^i(\tau)$, over all information structures $(\Omega, \mathscr{F}, P, (\mathscr{F}_i^n))$ and over all profiles τ in $G' = (G, \Omega, \mathscr{F}, P, (\mathscr{F}_i^n))$ that induce the correlated profile σ in G.

Our results show that $\hat{U}^i(\sigma)$ coincides with the measure of irrationality defined by Eq. (3). Thus, if a payoff vector is irrational in the sense of Definition 4.1 (for some game G with complete information), it cannot be an equilibrium payoff in *any* extension G' of G. Theorem 3.5 proves the converse – if a payoff vector is rational for G, then it is an equilibrium payoff in some extension G' of G.

5. The main ideas of the proofs

Since the main ideas that underlie the proofs of the main results are intuitive, while, as we work in a general setup, many technical difficulties appear, this section is devoted to an exposition of the main ideas of the proofs.

---

[3] That is, it allows player i to randomize without giving any information on the outcome to his opponents.

[4] The model studied in Aumann (1992) is slightly different than that of Aumann (1974), but it is more natural here to apply Aumann's definition of the measure of irrationality to the model described above.

Recall that $E_1 \supseteq E_2$. Our first result is:

Theorem 3.5. $E_1 = E_2$.

This theorem follows from the following two propositions, that are proved in section 7. Proposition 5.1 implies that $E_0 \supseteq E_1$, while Proposition 5.2 implies that $E_2 \supseteq E_0$.

Proposition 5.1. *Let* $\varepsilon > 0$. *For every general communication device* $\mathscr{C}$ *and every* ε*-equilibrium profile* τ *in* $G(\mathscr{C})$ *there exists a correlated profile* σ *such that (i)* $\gamma_{\mathscr{C}}(s_1, \tau) = \gamma(s_1, \sigma)$, *and (ii)* $U^i(\sigma) \leq \varepsilon$ *for every* $i \in I$.

The intuition of Proposition 5.1 is as follows. The general communication device $\mathscr{C}$ and the profile τ induce a probability distribution over plays, and therefore a correlated profile σ. If $U^i(\sigma) > \varepsilon$ then there exists a stopping time t such that if player i deviates at stage t, and defends his punishment level afterwards, he gains more than ε. But then, in $G(\mathscr{C})$, player i could have deviated from τ at stage t, and could have gained more than ε, which contradicts the assumption that τ is an ε-equilibrium.

Proposition 5.2. *For every correlated profile* σ *and every* $\varepsilon > 0$ *there exists an autonomous correlation device* $\mathscr{C}$ *and a profile* τ *in* $G(\mathscr{C})$ *such that (i)* $\gamma_{\mathscr{C}}(s_1, \tau) = \gamma(s_1, \sigma)$, *and (ii)* $\gamma^i_{\mathscr{C}}(s_1, \tau^{-i}, \tau'^i) \leq \gamma^i_{\mathscr{C}}(s_1, \tau) + U^i(\sigma) + \varepsilon$ *for every player* $i \in I$ *and every strategy* τ'^i *of player i in* $G(\mathscr{C})$.

The idea 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.

Since we have to construct an autonomous correlation device that does not observe the play, the device chooses at stage n a vector of recommendations, one recommendation for each possible history of length n. The players, who observe past play, know which recommendation to take into account, and which to disregard.

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.

Our second result is:

Theorem 3.6. *If for every player* $i \in I$, $v^i_{h_n}$ *is independent of* $h_n \in H$, *then* $E_1 = E_3$.

The intuition here is as follows. Denote by v^i the punishment level of player i, and let σ be a correlated profile in $G(\mathscr{C})$.

Assume for simplicity that there are finitely many actions, and that $U^i(\sigma) = 0$ for every player i. Then $U^i(h_n, \sigma, a^i) \leq 0$ for every player i, every history h_n that occurs with positive probability under σ, and every action a^i such that $\sigma(h_n)[\{a^i\} \times A^{-i}(s_n)] > 0$. In particular, $\mathbf{E}_{\sigma(h_n) \mid a^i} \gamma^i((h_n, a^i, a^{-i}_n), \sigma) \geq v^i$ for every such action a^i. By integrating over a^i and over s_n, we get that

$\gamma^i((h_{n-1}, a_{n-1}), \sigma) \geq v^i$ for every player i, every history h_{n-1}, and every action combination a_{n-1} such that $\sigma(h_{n-1})[a_{n-1}] > 0$. But this means that even when the players know which action combination is going to be played, their expected payoff is at least v^i. Since the punishment level is v^i, independent of the history, no player can profit by deviating, provided his deviation is followed by punishment. Thus, the device can choose a pure profile before start of play, and send it to everyone. The players are requested to follow this profile, and to punish a deviator. Since no player can profit by deviating at any stage, this is an equilibrium.

6. 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 5.2.

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

Fig. 1.

At stage 1, player 1 chooses a row, and player 2 independently chooses a column. If the players chose (B, L) then the game continues to stage 2, where player 2 chooses an entry. If the players chose another pair of actions at the first stage, or after the choice of player 2 at the second stage, the players receive a payoff as indicated in 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 in the interval $(3/2, 1/2)$-$(2, 0)$ is an equilibrium payoff in the extended game, for a properly 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 5.2.

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 at the previous stage to all the players.
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 at the punishment level.

7. Proofs of the equivalence theorems

7.1. Representing correlated profiles as autonomous devices

In this subsection we develop some measure theoretic results that are needed to prove Proposition 5.2.

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. But 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 7.1. *Let $\sigma : H \to \mathscr{P}(A_0)$ be a correlated profile, and $(Y_n)_{n\in\mathbf{N}}$ a sequence of i.i.d r.v.s uniformly distributed over* $[0,1]$. *There exists a sequence $(\delta_n)_{n\in\mathbf{N}}$ of measurable functions, $\delta_n : H_n \times [0,1] \to A_0$, such that for every $n \in \mathbf{N}$, every $h_n \in H_n$, 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 every $n \in \mathbf{N}$, there exists a measurable function $\delta_n : H_n \times [0,1] \to A$ that represent $\sigma(h_n)$. That is, the probability that $\delta_n(h_n, y)$ is in some set $C \subseteq A_0$, where y is uniformly distributed in $[0,1]$, is equal to $\sigma(h_n)[C]$.

Proposition 7.1 readily follows from the following lemma.

Lemma 7.2. *Let H be a measurable space, let X be a complete separable metric space, and let $\mathscr{X}$ be the σ-algebra of Borel subsets of X. Let $\mu : H \to \mathscr{P}(X)$ 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*

$$\mathbf{P}(\delta(h, Y) \in C) = \mu(h)[C] \quad \forall h \in H, C \in \mathscr{X}. \tag{4}$$

Proof: Let Y be a r.v. uniformly distributed over $[0,1]$. 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 $+\infty$. Define the function $\delta : H \times [0,1] \to [0,1]$ by

$$\delta(h, y) = x_{k_0}, \quad \text{where } k_0 = \min\left\{ k \mid \sum_{n=1}^{k} \mu(h)[x_n] \geq y \right\},$$

where the minimum over an empty set is infinity. Note that δ is measurable. Eq. (4) holds, since for every 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 $([0,1], \mathscr{B})$, where $\mathscr{B}$ 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 $(X, \mathscr{X}) = ([0,1], \mathscr{B})$.

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 $x \in [0,1]$,

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

where $\mathbf{Q}$ is the set of rational numbers. Since a countable union of measurable sets is measurable, and since μ is measurable, δ is measurable.

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 $C \in \mathscr{B}$, $\mathbf{P}(\delta(h, Y) \in C) = \mu(h)[C]$, as desired. ■

7.2. 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) at every stage each player receives a private signal as well as a public signal, (iii) 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 (iv) at stage n the device publicly announces the private signals that were sent at stage $n-1$.

Definition 7.3. *A* standard revealing autonomous correlation device $\mathscr{C}$ *is given by a sequence* $(\delta_n)_{n \in \mathbf{N}}$ *of measurable functions,* $\delta_n : H_n \times [0,1] \to A_0$*, 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 $m_n^i = (\delta_{n-1}(\cdot, Y_{n-1}), \delta_n^i(\cdot, Y_n))$. $\delta_k^i(h_k, Y_k)$ can be interpreted as a recommended action for player i if the realized history up to stage k is h_k.

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 $y \in [0,1]$ defines a function $\delta_n(\cdot, y) : H_n \to A_0^i$. Let $M_n^{\prime i}$ be the space of all these functions. The Borel measurable structure of $[0,1]$ induces a measurable structure on $M_n^{\prime i}$, and the uniform distribution over $[0,1]$ induces a probability distribution ν_n over $M_n^{\prime i}$. Finally, the signal space of player i at stage n is $M_n^i = M_{n-1}^{\prime} \times M_n^{\prime i}$, where $M_{n-1}^{\prime} = \times_{i \in I} M_{n-1}^{\prime i}$, and the distribution over M_n^i is $1_{m_{n-1}^{\prime}} \otimes \nu_n$, where $m_{n-1}^{\prime}$ is the recommendation at stage $n-1$.

7.3. The proofs

Proof of Proposition 5.1:
Let $\varepsilon > 0$, let $\mathscr{C}$ be an autonomous correlation device, and let τ be an ε-equilibrium profile in $G(\mathscr{C})$.

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

We shall now prove that $U^i(\sigma) \leq \varepsilon$ for every $i \in I$. Otherwise, there exists a player $i \in I$, and a stopping time t such that $\mathbf{E}_{s_1, \sigma} U^i(h_t, \sigma, a_t) > \varepsilon + \rho$, for some $\rho > 0$. Define a strategy $\tau^{\prime i}$ for player i in $G(\mathscr{C})$ 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

$$\gamma^i_{\mathscr{C}}(s_1, \tau^{-i}, \tau'^i) \geq \gamma^i(s_1, \sigma) + \mathbf{E}_{s_1,\sigma} U^i(h_t, \sigma, a_t) - \rho \geq \gamma^i_{\mathscr{C}}(s_1, \tau) + \varepsilon,$$

a contradiction, since τ is an ε-equilibrium. ∎

Proof of Proposition 5.2:
Let σ be a correlated profile, let $\varepsilon > 0$, and let $(Y_n)_{n\in\mathbf{N}}$ be a sequence of i.i.d. r.v.s, uniformly distributed over $[0,1]$. By Proposition 7.1, there exists a sequence of measurable function $\delta_n : H_n \times [0,1] \to A$, $n \in \mathbf{N}$, such that for every $n \in \mathbf{N}$, every $h_n \in H_n$, and every measurable subset C of A_0,

$$\sigma(h_n)[C] = \mathbf{P}(\delta_n(h_n, Y_n) \in C). \tag{5}$$

Let $\mathscr{C}$ be the standard revealing autonomous correlation device defined by $(\delta_n)_{n\in\mathbf{N}}$.

Define a profile τ in $G(\mathscr{C})$ as follows. At every stage n, the realized action of each player at stage $n-1$ is compared with $\delta^i_{n-1}(h_{n-1}, Y_{n-1})$, the recommendation of the device, 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 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 start of play, the device chooses, for every player $i \in I$, a realization of a sequence of i.i.d. r.v.s uniformly distributed in $[0,1]$. The device then sends the realization to all players *except* player i. If the necessity arises, players $N\setminus\{i\}$ use the realization to correlate their moves and follow an ε-punishment correlated strategy $\tilde{\sigma}^{-i}_{\varepsilon}$ against player i. The realization can be translated into a punishing strategy by applying Proposition 7.1 to $\sigma = \tilde{\sigma}^{-i}_{\varepsilon}$.

One can verify that $\mathbf{P}_{s_1,\mathscr{C},\tau} = \mathbf{P}_{s_1,\sigma}$, and therefore $\gamma_{\mathscr{C}}(s_1,\tau) = \gamma(s_1,\sigma)$.

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

$$t = \min\{n \in \mathbf{N} \mid a^i_n \neq \delta^i_n(h_n, Y_n)\} + 1.$$

Then, under (τ^{-i}, τ'^i), at stage t players $N\setminus\{i\}$ switch to an ε-punishment profile against player i. By the definition of $U^i(\sigma)$,

$$\begin{aligned}\gamma^i_{\mathscr{C}}(s_1, \tau^{-i}, \tau'^i) &\leq \gamma^i(\sigma) + \mathbf{E}_{s_1,\mathscr{C},\tau^{-i},\tau'^i}(U^i(h_t, \sigma, a^i_t) 1_{t<+\infty}) + \varepsilon \\ &\leq \gamma^i(\sigma) + U^i(\sigma) + \varepsilon,\end{aligned}$$

as desired. ∎

Proof of Theorem 3.6:
Assume now that for every fixed player $i \in N$, $v^i_{h_n}$ is independent of $h_n \in H$, and denote this common value by v^i.

In view of Theorem 3.5, it suffices to prove that $E_0 \subseteq E_3$: every rational payoff is a correlated equilibrium payoff.

Fix $\varepsilon > 0$. 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 $U^i(\sigma) < \varepsilon$ for each player $i \in N$. For every $\delta > 0$, denote by H_i^δ the set of all histories $h_\infty \in H_\infty$ such that $\mathbf{E}_{\sigma(h_n) \mid a_n^i} \gamma^i((h_n, a_n^i, a^{-i}), \sigma) < v^i - \delta$ for some beginning (h_n, a_n^i, a_n^{-i}) of h_∞.

We now show that since $U^i(\sigma) < \varepsilon$, and since $v_{h_n}^i$ is independent of h_n for every $i \in I$, $\mathbf{P}_{s_1,\sigma}(H_i^{\sqrt{\varepsilon}}) < \sqrt{\varepsilon}$. Define a stopping time t by

$$t(h_\infty) = \min\{n \in \mathbf{N} \mid \mathbf{E}_{\sigma(h_n) \mid a_n^i} \gamma^i((h_n, a_n^i, a^{-i}), \sigma) < v^i - \sqrt{\varepsilon}\}.$$

Since $v_{h_n}^i$ is independent of the history,

$$U^i(h_n, \sigma, a_n^i) = v^i - \mathbf{E}_{\sigma(h_n) \mid a_n^i} \gamma^i((h_n, a_n^i, a^{-i}), \sigma).$$

By the definition of the measure of irrationality,

$$\varepsilon > U^i(\sigma) \geq \mathbf{E}_{s_1,\sigma}(U^i(h_t, \sigma, a_t^i) 1_{t < +\infty}) > \mathbf{P}_{s_1,\sigma}(H_i^{\sqrt{\varepsilon}}) \times \sqrt{\varepsilon},$$

and therefore $\mathbf{P}_{s_1,\sigma}(H_i^{\sqrt{\varepsilon}}) < \sqrt{\varepsilon}$.

We now restrict ourselves to histories $h_\infty \notin H_i^{\sqrt{\varepsilon}}$. Since $h_\infty \notin H_i^{\sqrt{\varepsilon}}$, $\mathbf{E}_{\sigma(h_n) \mid a_n^i} \gamma^i((h_n, a_n^i, a^{-i}), \sigma) \geq v^i - \sqrt{\varepsilon}$. By integrating over a_n^i we get that for every $n \in \mathbf{N}$, $\gamma^i(h_n, \sigma) \geq v^i - \sqrt{\varepsilon}$. By integrating over s_n we get that for every $n > 1$,

$$\gamma^i((h_{n-1}, a_{n-1}), \sigma) \geq v^i - \sqrt{\varepsilon}.$$

But this means that for $h_\infty \notin H_i^{\sqrt{\varepsilon}}$, even if player i knows the pure action combination that is going to be played, he cannot profit more than $\sqrt{\varepsilon}$ by deviating.

Define a correlation device $\mathscr{C}$ 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 is noticed upon deviation, will be punished at his punishment level v^i.

To allow players to punish a deviator, we use the same idea as in the proof of Proposition 5.2. The device chooses for each player i, before start of play, a realization of a sequence of i.i.d. r.v.s uniformly distributed in the unit interval, and sends it to all players except player i. If player i ever deviates, players $N \backslash \{i\}$ use this realization to punish him. The realization can be translated into a punishing strategy by applying Proposition 7.1 $\sigma = \tilde{\sigma}_\varepsilon^{-i}$.

Let $H^{\sqrt{\varepsilon}} = \bigcup_{i \in I} H_i^{\sqrt{\varepsilon}}$. Then $\mathbf{P}_{s_1,\sigma}(H^{\sqrt{\varepsilon}}) < I\sqrt{\varepsilon}$. Thus, with probability greater than $1 - I\sqrt{\varepsilon}$, the realized history is not in $H^{\sqrt{\varepsilon}}$. Conditional on $h_\infty \notin H^{\sqrt{\varepsilon}}$, for every $n \in \mathbf{N}$ and every player $i \in I$, $\gamma^i((h_n, a_n), \sigma) \geq v^i - \sqrt{\varepsilon}$. In particular, no player i can profit more than $\varepsilon + \sqrt{\varepsilon}$ by deviating. Thus, $\gamma(s_1, \sigma)$ is a $((1 - I\sqrt{\varepsilon})(\varepsilon + \sqrt{\varepsilon}) + I\sqrt{\varepsilon}R)$-equilibrium payoff (recall that R is a bound of u).

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 5.2 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 $i \in I$,

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

Remark: Forges (1988) defined the notion of canonical devices in repeated games with incomplete information. In our context, a correlation device is *canonical* if the signal it sends to each player before start of play is a pure strategy for the whole game, and an autonomous correlation device is *canonical* if the signal it sends to each player at every stage is a vector of recommended actions, one for each possible history.

The correlation devices that we construct in the proofs of Theorems 3.5 and 3.6 are not canonical. However, as Francoise Forges commented, in both cases there are equivalent canonical devices. Indeed, since in both cases the device knows its recommendations, it can calculate, for every history of length n, whether it is compatible with past recommendations or not. Moreover, if it is not compatible, it can calculate who must have deviated. Thus, for every finite history which is not compatible with the last recommendation (and is compatible with all previous recommendations), future recommendations are derived from an ε-punishment profile against one of the deviators.

References

Aumann RJ (1974) Subjectivity and correlation in randomized strategies, *J Math Econ* 1:67–96

Aumann RJ (1987) Correlated equilibrium as an expression of bayesian rationality, *Econometrica* 55:1–18

Aumann RJ (1992) "Irrationality in game theory," *Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn*, P. Dasgupta, D. Gale, O. Hart, E. Maskin (eds), 214–227, The MIT Press, Cambridge and London

Castaing C, Valadier M (1977) "Convex analysis and measurable multifunctions," Springer-Verlag, Berlin Heidelberg New-York

Dellacherie C, Meyer P-A (1978) "Probabilities and potential," North-Holland, Amsterdam

Forges F (1985) "Correlated equilibria in a class of repeated games with incomplete information," *International Journal of Game Theory* 14:129–150

Forges F (1986) "An approach to communication equilibria," *Econometrica* 54:1375–1385

Forges F (1988) "Communication equilibria in repeated games with incomplete information," *Mathematics of Operation Research* 13:191–231

Maitra A, Sudderth W (1993) "Borel stochastic games with limsup payoff," *Annals of Probability* **21**:861–885

Maitra A, Sudderth W (1998) "Finitely additive stochastic games with borel measurable payoffs," *International Journal of Game Theory* 27:257–267

Mertens JF, Sorin S, Zamir S (1994) "Repeated games," CORE Discussion Papers 9420–9422

Neveu J (1965) *Mathematical Foundations of the Calculus of Probability*, Holden-Day

Nowak AS (1991) "Existence of correlated weak equilibria in discounted stochastic games with general state space," in *Stochastic Games and Related Topics*, ed. by TES Raghavan et al., Kluwer Academic Press

Nowak AS (1999a) "Sensitive equilibria for ergodic stochastic games with countable state spaces", *Mathematical Methods of Operations Research* 50:65–76

Nowak AS (1999b) "Optimal strategies in a class of zero-sum ergodic stochastic games", *Mathematical Methods of Operations Research* 50:399–419

Parthasarathy KR (1967) "Probability measures on metric spaces", Academic Press, New York

Solan E, Vieille N (1998) "Correlated equilibrium in stochastic games," *Games and Economic Behavior*, to appear