Available online at www.sciencedirect.com
SCIENCE DIRECT®
Games and Economic Behavior 49 (2004) 313–344
GAMES and Economic Behavior
www.elsevier.com/locate/geb

Communication equilibrium payoffs in repeated games with imperfect monitoring

Jérôme Renault, Tristan Tomala *

CEREMADE, Université Paris Dauphine, Place du Maréchal de Lattre, 75776 Paris cedex 16, France

Received 26 September 2000
Available online 20 March 2004

Abstract

We characterize the set of communication equilibrium payoffs of any undiscounted repeated matrix-game with imperfect monitoring and complete information. For two-player games, a characterization is provided by Mertens, Sorin, and Zamir (Repeated games, Part A (1994) CORE DP 9420), mainly using Lehrer's (Math. Operations Res. (1992) 175) result for correlated equilibria. The main result of this paper is to extend this characterization to the n-player case. The proof of the characterization relies on an analogy with an auxiliary 2-player repeated game with incomplete information and imperfect monitoring. We use Kohlberg's (Int. J. Game Theory (1975) 7) result to construct explicitly a canonical communication device for each communication equilibrium payoff.

JEL classification: C72; C73

Keywords: Repeated games; Imperfect monitoring; Communication equilibria

1. Introduction

We study repeated games with imperfect monitoring. In such interactions, a one-shot matrix-game known by all the players is repeated over and over, and after each stage the players get some signal depending on the actions just played. A general goal is to extend the Folk Theorem to such games, i.e. to characterize the set of equilibrium payoffs according to the original data (one-shot game and signalling functions). The pioneering work in this area is due to Lehrer, who obtained characterizations of equilibrium payoffs

* Corresponding author.
E-mail addresses: renault@ceremade.dauphine.fr (J. Renault), tomala@ceremade.dauphine.fr (T. Tomala).

0899-8256/$ – see front matter
doi:10.1016/j.geb.2003.12.005

for 2-player games in the case of Nash equilibria (Lehrer, 1989, 1992b) and in the case of correlated equilibria (Lehrer, 1992a). In the n-player case, characterizations were provided for particular classes of signalling functions only (Lehrer, 1990 for semi-standard signalling, Renault and Tomala, 1998 for a partial result in case of graph monitoring, Tomala, 1999 for observable payoff vectors).

We deal here with general repeated games with imperfect monitoring, without any assumption on the number of players, the one-shot matrix-game or the signalling functions. We provide a characterization of the set of equilibrium payoffs depending only on the original data. An important point is that we deal with communication equilibrium payoffs and not with Nash equilibrium payoffs. Communication equilibria have been introduced by Forges (1985) and Myerson (1982). These are Nash equilibria of some extension of the repeated game, where a mediator who can communicate with the players has been added. The mediator has no payoff, no commitment power, and he can just communicate in a private way with each player between the stages. He may be seen as an extra player with payoff 0 who can help the players to coordinate themselves. Although communication equilibria are more complicated to define, they are here easier to study. For example, as soon as player i knows player j has deviated, i can tell it to the mediator who will warn the other players.

The set of communication equilibrium payoffs (denoted by C throughout the paper) always contains the set of Nash equilibrium payoffs and the set of correlated equilibrium payoffs of the one-shot game and is thus non empty. For two-player games, a characterization is provided in Mertens et al. (1994, part A), mainly using Lehrer's (1992a) result for correlated equilibria. The main result of the present paper is to extend this characterization to the n-player case. Mertens et al. use the following idea of Lehrer to characterize C. A deviation of a player is undetectable if it can not be directly observed (i.e. has no influence on others' signals) and provides to the deviator at least as much information (consequently the deviation cannot be detected by "asking questions" such as: what did you observe at such stage?). For 2-player games, the characterization can be described as follows: C is the set of individually rational payoffs which are feasible in a way such that there is no profitable and undetectable deviation. In the n-player case studied here, a new phenomenon appears: during the game, it is possible that all players know that a deviation occurred, without knowing who has deviated. Consequently, at equilibria all players that may have deviated must be simultaneously punished, and strategies leading to collective punishments have to be introduced (see Tomala, 1999 for a construction of such punishments in a simple case). This type of punishing strategies inducing simultaneously low payoffs for several criteria is typically a "Blackwell" approachability strategy (see Blackwell, 1956). The notion of individual rationality is then replaced by the one of joint rationality, and the introduction and the characterization of the set of jointly rational payoffs is the main new aspect of our result.

Our proof is based on the consideration of an auxiliary repeated game with two players and incomplete information, and the use for such a game of a result by Kohlberg (1975) in the spirit of Blackwell's approachability. In this auxiliary game, one of the players represents the mediator and the other player is called the cheater, representing any potential deviator in the original game. The state of nature represents the identity of the player who may be deviating in the original game. It is known by the cheater but not by the mediator. Formally, the auxiliary game obtained is a 2-player game with lack of information on one side, where the uninformed player always has payoff 0 and with state dependent signalling. The study of some Nash equilibrium payoffs of such repeated games yields our result. In particular, it provides for each communication equilibrium payoff of the original game, an associated canonical communication device. The analogy between communication equilibria of the original game and specific Nash equilibria in the game is very robust. It is merely an identification of strategy spaces in the two games. In particular it is valid under any kind of evaluation of payoffs e.g. discounted games or finitely repeated games. However, to get a full characterization of the set of equilibrium payoffs, we focus on undiscounted games for which some aspects of the analysis are easier. For example, players can use any finite number of stages for information transmission, without influencing payoffs. As a consequence, our characterization requires no hypothesis on payoffs, as for the Folk Theorem. For convenience, we first present our result in the case of deterministic signals. This simplifies the definitions and the proof, but the characterization will also hold in the general case of random signals (see the last section). This may be of interest since random signals appear in many economic models such as principal-agent problems where the outcome observed by the principal depends stochastically on the action taken by the agent. In such a case, the use of a mediator can also have some economical relevance: think of situations where an outside actor (a regulation authority, e.g. representatives of the state) is used to smooth conflicts inside a firm.

In Section 2, we introduce the model and the definitions. For simplicity, we already start with the notion of canonical communication equilibria, which is payoff-equivalent to that of communication equilibria. The statement of the characterization is given in Section 3. We then derive previously known results from it, and provide illustrative examples showing how to compute C. An example is given where C is not a polytope (i.e. the convex hull of a finite number of points). Note that such an example necessarily involves at least 3 players since for 2-player games, C is a polytope (Lehrer, 1992a). In this case, simultaneous punishments are much more difficult to construct. Section 4 is devoted to the proof and to the analogy with 2-player repeated games with lack of information on one side. We finally conclude with the consideration of random signals (Section 5).

2. The model

We start with a set of players N. Each player i in N has a set of actions A^i and a payoff function g^i from $A = \prod_{j \in N} A^j$ to $\mathbb{R}$. The observation of player i is given by a set of signals U^i and an observation function f^i from A to U^i. The repeated game, denoted by Γ, is played as follows. At each stage $t = 1, 2, \ldots,$ the players independently and simultaneously choose an action in their own set of actions. If a in A is the joint action selected, the stage payoff for player i is $g^i(a)$, and before starting stage $t+1$, player i learns the signal $f^i(a)$ only (hence player i may not know his own payoff $g^i(a)$). Players are assumed to have perfect recall. The infinitely repeated game Γ is thus characterized by the data N, $(A^i)_{i \in N}$, $(g^i)_{i \in N}$, $(U^i)_{i \in N}$, $(f^i)_{i \in N}$, fixed once and for all. We assume that the set of players, the sets of actions and the sets of signals are all non empty and finite. The goal of this paper is to provide a characterization of the set of uniform communication equilibrium payoffs according to these data. Throughout the paper, we will use the following notations. If $(E^i)_{i\in N}$ is a collection of sets indexed by N, E will denote $\prod_{i\in N} E^i$. An element $(e^i)_{i\in N}$ in E will simply be denoted by e, and we will denote by e^{-i} the current element of $E^{-i} = \prod_{j\neq i} E^j$. We will write $e = (e^i, e^{-i})$ when the ith component is stressed. If E is a finite set, $|E|$ will denote its cardinality and $\Delta(E)$ the set of probability distributions over E. An element e in E will be identified with the Dirac mass on e. For $p = (p(e))_{e\in E}$ in $\Delta(E)$, *Supp* p will denote the support of p. $\Delta(E)$ will be viewed as a subset of the Euclidean space $\mathbb{R}^E$, and $p \cdot q$ will denote the canonical inner product of p and q. A communication equilibrium of Γ is a Nash equilibrium of an extended game where a mediator has been added. The mediator communicates with the players through a fixed public knowledge procedure called a communication device: before each stage, he sends a private recommendation to each player, and after each stage, each player sends a message back to the mediator. Notice that there is no need here for an extra initial stage where the players would send messages to the mediator, because the players have no initial private information. Forges in 1986 (see also Mertens et al., 1994) showed, using a revelation principle, that communication equilibria admit a canonical form. Namely, any communication equilibrium outcome can be sustained by a canonical communication equilibrium in which at each stage, the mediator suggests each player which action to play, players actually play these actions and report their observed signals to the mediator. Since the set of canonical communication equilibrium payoffs equals the set of communication equilibrium payoffs, we only formally define the former. We first consider communication devices such that the recommendation sent by the mediator to each player i is an action in A^i and where after each stage each player sends back a message in U^i. To distinguish between recommendations and actions, and between signals and messages, it is actually convenient to define, for each player i: $R^i = A^i$ (R^i will be interpreted as the set of recommendations for player i whereas A^i is the set of actions that player i can take), and similarly $M^i = U^i$ (M^i for messages sent back to the mediator, U^i for signals observed by player i).

Definition 2.1. A canonical communication device is an element $c = (c_t)_{t\geqslant 1}$, where $c_1 \in \Delta(R)$ and for each $t \geqslant 2$, c_t is a mapping from $(R \times M)^{t-1}$ to $\Delta(R)$.

Given a fixed canonical communication device c, we define an infinitely repeated game Γ_c played as follows:

Stage 1. The mediator selects a joint recommendation $(r_1^i)_{i\in N}$ in R according to c_1, and privately sends the recommendation r_1^i to each player i. Then the players simultaneously choose actions and receive signals as in the original game, and to conclude stage 1 every player i chooses a message m_1^i in M^i that he sends in a private way to the mediator.

Stage t. The mediator selects a joint recommendation $(r_t^i)_{i\in N}$ in R according to $c_t((r_1^i, m_1^i)_{i\in N}, \ldots, (r_{t-1}^i, m_{t-1}^i)_{i\in N})$ and privately sends r_t^i to each player i. Then the players simultaneously choose actions $a_t = (a_t^i)_{i\in N}$ and receive signals ($f^i(a_t)$ for player i), and to conclude stage t every player i sends back a private message m_t^i to the mediator.

Players are assumed to have perfect recall, and the whole description of the game, including c, is public knowledge. In Γ_c, a behavior strategy for player i will be denoted by $\sigma^i = (\sigma^i_t, \xi^i_t)_{t\geqslant 1}$ where for each t:

σ^i_t gives the lottery on actions played by player i at stage t depending on his past recommendations from the mediator, the actions he played, the signals he observed and the messages he sent to the mediator. σ^i_t is a mapping $(R^i \times A^i \times U^i \times M^i)^{t-1} \times R^i \longrightarrow \Delta(A^i)$.

ξ^i_t gives the lottery on messages sent by player i at stage t, depending on the past. It is a mapping $(R^i \times A^i \times U^i \times M^i)^{t-1} \times R^i \times A^i \times U^i \longrightarrow \Delta(M^i)$.

Denote by Σ^i the set of behavior strategies for player i in Γ_c. A play in this game is a sequence $((r^i_1)_{i\in N}, (a^i_1)_{i\in N}, (u^i_1)_{i\in N}, (m^i_1)_{i\in N}, \ldots, (r^i_t)_{i\in N}, (a^i_t)_{i\in N}, (u^i_t)_{i\in N}, (m^i_t)_{i\in N}, \ldots)$, hence the set of plays is $\Omega = (R \times A \times U \times M)^\infty$, endowed with the product σ-algebra. A profile of behavior strategies $\sigma \in \Sigma$ naturally induces a probability $\mathbb{P}_{\sigma,c}$ over Ω. We define the expected average payoffs as:

$$\forall i \in N, \forall T \geqslant 1, \quad \gamma^i_{T,c}(\sigma) = \mathbb{E}_{\mathbb{P}_{\sigma,c}}\left(\frac{1}{T}\sum_{t=1}^{T} g^i(a_t)\right).$$

We think of players as maximizing the expectation of their average payoffs and use the classical notion of uniform (Nash) equilibria (see for example Sorin, 1992).

Definition 2.2. $\sigma \in \Sigma$ is an equilibrium of Γ_c if: (i) for each player i, $(\gamma^i_{T,c}(\sigma))_{T\geqslant 1}$ converges as T goes to infinity to some $\gamma^i_c(\sigma)$,

(ii) for all $\varepsilon > 0$, there exists T_0 s.t. σ is an ε-Nash equilibria in finitely repeated games with at least T_0 stages, that is:

$$\forall T \geqslant T_0, \forall i \in N, \forall \tau^i \in \Sigma^i, \quad \gamma^i_{T,c}\big(\tau^i, \sigma^{-i}\big) \leqslant \gamma^i_{T,c}(\sigma) + \varepsilon.$$

$(\gamma^i_c(\sigma))_{i\in N} \in \mathbb{R}^N$ is then called an equilibrium payoff of Γ_c.

In Γ_c, each player i has a special strategy σ^{i*}: at each stage, σ^{i*} plays the recommendation just received, and sends back to the mediator the signal just observed by player i. We will refer to σ^{i*} as the faithful strategy of player i. We now define the set we are interested in.

Definition 2.3. If c is a canonical communication device and if the faithful strategy σ^* is an equilibrium of Γ_c, the payoff $(\gamma^i_c(\sigma^*))_{i\in N} \in \mathbb{R}^N$ is called a canonical communication equilibrium payoff of the original repeated game Γ. Let C be the set of such payoffs as c varies.

As already said, C is indeed the set of all communication equilibrium payoffs of Γ. We will consequently often omit the word canonical while dealing with communication equilibrium payoffs (CEP for short).

3. The characterization

We first present the main strategic aspects of communication equilibria of repeated games with imperfect monitoring, then state our result and finally provide comments and examples.

3.1. Feasibility

We denote by $g : A \longrightarrow \mathbb{R}^N$ the vector payoff function $(g(a) = (g^i(a))_{i \in N}$ for each $a \in A)$ and extend g to $\Delta(A)$ by setting $g(p) = \mathbb{E}_p(g)$ for all p in $\Delta(A)$. If P is a subset of $\Delta(A)$, we denote $\{g(p), p \in P\}$ by $g(P)$. Since a CEP is defined as a limit of expectations of stage-average payoffs, any CEP must belong to the convex compact set of feasible payoffs $g(\Delta(A))$. Hence $C \subset g(\Delta(A))$.

3.2. Individual rationality

The correlated minmax of player i is defined by

$$w^i = \min_{p^{-i} \in \Delta(A^{-i})} \max_{p^i \in \Delta(A^i)} g^i\left(p^i \otimes p^{-i}\right) = \max_{p^i \in \Delta(A^i)} \min_{p^{-i} \in \Delta(A^{-i})} g^i\left(p^i \otimes p^{-i}\right),$$

and the set of individually rational payoffs is

$$IR = \left\{x = (x^i)_{i \in N} \in \mathbb{R}^N, x^i \geqslant w^i \ \forall i \in N\right\},$$

where w^i is the lowest quantity that player i can be punished to. Note that the minimum is taken over the set $\Delta(A^{-i})$, because the players can correlate their actions, with the help of the mediator, in order to punish player i. It is plain that in any extended game player i can always obtain at least a payoff of w^i by playing the maximizing p^i at each stage. Hence at equilibrium his payoff should always be at least w^i. So $C \subset IR$. Recall that when the players perfectly observe all actions played ($\forall i$, $U^i = A$ and f^i is the identity map), the Folk Theorem (see Sorin, 1992) for CEP states that $C = g(\Delta(A)) \cap IR$. The last two aspects (see Sections 3.3 and 3.4) are due to imperfect monitoring.

3.3. Undetectable deviations

This phenomenon already appears in two-player games, and the following definitions are due to Lehrer (1989). Assume player i deviates in a way such that:

(i) it induces the same signals for every player j in $N \backslash \{i\}$.
(ii) it gives to player i at least as much information.

It is then impossible to detect this deviation, because no player in $N \backslash \{i\}$ will be aware of it, and player i is able to continue the play as if he did not deviate. Such deviation should not give a better payoff to him. Formally, let a^i and b^i be two actions in A^i. Think of a^i as the action recommended to player i, and of b^i as a possible deviation. We write $b^i \geqslant a^i$ if:

(i) $\forall a^{-i} \in A^{-i}, \forall j \in N\backslash\{i\},\ f^j(b^i, a^{-i}) = f^j(a^i, a^{-i})$ (b^i and a^i are said to be *equivalent*),
(ii) $\forall a^{-i} \in A^{-i},\ \forall b^{-i} \in A^{-i},\ f^i(a^i, a^{-i}) \neq f^i(a^i, b^{-i})$ implies $f^i(b^i, a^{-i}) \neq f^i(b^i, b^{-i})$ (b^i is said to be *more informative* than a^i).

Remark 1. Note that (ii) is equivalent to: $\exists \tilde{\mu}^i : U^i \longrightarrow M^i$ s.t. $\forall a^{-i}, \tilde{\mu}^i(f^i(b^i, a^{-i})) = f^i(a^i, a^{-i})$. Knowing his recommendation a^i, the action he played b^i and the signal he observed, player i is able to send back the signal $f^i(a^i, a^{-i})$ to the mediator, i.e. to play as if no deviation had occurred (see Lehrer, 1992a).

Assume now that the mediator selects a joint recommendation $a = (a^j)_{j\in N}$ in A according to some probability p, and privately recommends each player j to play a^j. If player i deviates and plays some action b^i, his expected payoff will be $\sum_{a^{-i}\in A^{-i}} p(a^{-i}|a^i) g^i(b^i, a^{-i}) = (\sum_{a^{-i}\in A^{-i}} p(a^i, a^{-i}) g^i(b^i, a^{-i}))/p(a^i)$. If moreover $b^i \geqslant a^i$, the deviation will not be detected, hence at equilibrium, it should not give player i a greater payoff. We thus put:

$$\mathcal{P} = \Bigg\{ p \in \Delta(A), \forall i \in N, \forall b^i, a^i \in A^i \text{ s.t. } b^i \geqslant a^i, \sum_{a^{-i}\in A^{-i}} p\big(a^i, a^{-i}\big) g^i\big(a^i, a^{-i}\big) \geqslant \sum_{a^{-i}\in A^{-i}} p\big(a^i, a^{-i}\big) g^i\big(b^i, a^{-i}\big) \Bigg\}.$$

Note that $g(\mathcal{P})$ is obviously included into the set of feasible payoffs $g(\Delta(A))$. Following the work of Lehrer (1992a), Mertens et al. (1994) showed that for 2-player repeated games, $C = g(\mathcal{P}) \cap IR$.

3.4. Joint rationality

With more than two players, a new phenomenon appears: it may be the case that everyone knows a deviation occurred, without knowing who did deviate. For example, consider a 3-player game where:

(i) player 1 and player 2 have trivial monitoring (they always observe the same signal after each stage i.e. U^1 and U^2 are singletons).
(ii) the signal observed by player 3 does not depend on his own move and the values of the signalling function f^3 are given by the following matrix (player 1 is the row player and $A^1 = \{T, B\}$, player 2 is the column player and $A^2 = \{L, R\}$):

	L	R
T	u	v
B	v	v

Consider a canonical communication equilibrium where at some stage, the device recommends player 1 to play T and player 2 to play L. Suppose that at this stage, player 3 observes the signal v and faithfully sends it back to the mediator. Considering unilateral deviations only, the mediator knows that one and only one of the three following statements is correct: (player 1 has deviated by playing B) or (player 2 has deviated by playing R) or (player 3 has deviated by reporting v). But he can not decide which one is true. Consequently, we show that it is necessary to punish simultaneously all players suspected of deviation to prevent profitable deviations. Concerning CEP, this implies that for any subset of players that can be simultaneously suspected, the equilibrium payoffs of these players may be linked and will have to satisfy several inequalities.

Example 3.1. Consider the following game: $N = \{1, 2, 3\}$, $A^1 = \{T, B\}$, $A^2 = \{L, R\}$, $A^3 = \{W, M, E\}$, $U^1 = U^2 = \{*\}$, $U^3 = \{u, v\}$. Payoffs are given by the following matrices:

W

	L	R
T	(0,0,0)	(0,2,0)
B	(2,0,0)	(1,1,0)

M

	L	R
T	(0,1,0)	(0,1,0)
B	(0,1,0)	(0,1,0)

E

	L	R
T	(1,0,0)	(1,0,0)
B	(1,0,0)	(1,0,0)

and signals for player 3 are as follows:

W

	L	R
T	u	v
B	v	u

M

	L	R
T	v	v
B	v	v

E

	L	R
T	v	v
B	v	v

First note that player 3 always has payoff 0. Hence deviations of this player will never be profitable, so we can simply forget about these deviations and always think player 3 as playing his faithful strategy. The set of feasible payoffs $g(\Delta(A))$ is here the convex hull of $(0, 2, 0)$, $(2, 0, 0)$, and $(0, 0, 0)$. If player 3 plays M, player 1's payoff is 0 whatever he does: hence $w^1 = 0$. Similarly $w^2 = 0$. Note that the two actions of player 1 (respectively player 2) are not equivalent, hence $g(\mathcal{P}) = g(\Delta(A))$. So we get $g(\mathcal{P}) \cap IR = \mathrm{conv}\{(0, 2, 0), (2, 0, 0), (0, 0, 0)\}$ where conv stands for convex hull. We claim that $(0, 0, 0)$ is not a CEP. The intuition is the following. Assume that c is a canonical communication device such that the faithful strategy $(\sigma^{1*}, \sigma^{2*}, \sigma^{3*})$ is an equilibrium of Γ_c with payoff $(0, 0, 0)$. Since $(0, 0, 0)$ is an extreme point of $g(\Delta(A))$, (T, L, W) must be recommended to the players in most stages with high probability. Consider now the deviation σ^1 of player 1 consisting of playing at each stage T and B with equal probability, independently of what happened before, and in particular independently of the recommendation of the mediator. Assume that $(\sigma^1, \sigma^{2*}, \sigma^{3*})$ is played. Then at each stage where player 3 is recommended to play M or E, the message reported to the mediator is v, and at each stage where player 3 is recommended to play W, the law of the message reported to the mediator, is uniform on $\{u, v\}$. In particular, at each stage where (T, L, W) is recommended, it occurs with probability $1/2$ that player 1 plays B, has a payoff of 2 and the reported message is v. Since σ^1 should not be a profitable deviation, player 1 has to be punished, and the device should recommend at (almost) every stage player 3 to play M. But consider now the deviation σ^2 of player 2 consisting of playing at each stage L and R with equal probability, independently of what happened before. If $(\sigma^{1*}, \sigma^2, \sigma^{3*})$ is played, the message reported to the mediator is again uniformly distributed on $\{u, v\}$ at each stage where W is recommended to player 3, and the message reported to the mediator is v at each other stage. We have seen that player 3 will then play M very often, giving a payoff of 1 for player 2. So σ^2 is a profitable deviation, and $(0,0,0)$ can not be a CEP. The point here is that it is not possible to punish simultaneously players 1 and 2 at the level $(0,0)$. We now quantify the levels of simultaneous punishments. By playing at each stage $\lambda M + (1-\lambda)E$, with λ in $[0,1]$, player 3 can make sure that the expected payoff for player 1 is $(1-\lambda)$ whereas player 2 has payoff λ. This will imply that any payoff $x = (x^1, x^2, x^3)$ in $\operatorname{conv}\{(0,2,0),(2,0,0),(0,0,0)\}$ which verifies $x^1 + x^2 \geqslant 1$ can be obtained as a CEP. This can be proved as follows: take an infinite sequence of pure joint actions $a_1, a_2, \ldots, a_t, \ldots$ giving on the average the payoff x, and construct a canonical communication device c recommending to play a_t at each stage t, as long as the messages reported by player 3 do not imply that a deviation has occurred. To avoid profitable deviations, if at some stage t the message reported by player 3 shows that a deviation has occurred, then at any subsequent stage c will recommend player 3 to play M with probability λ and E with probability $1-\lambda$. If λ is chosen such that $1-\lambda \leqslant x^1$ and $\lambda \leqslant x^2$, no deviation from the faithful strategy will be profitable. This proves that

$$\bigl\{(x^1, x^2, x^3) \in g\bigl(\Delta(A)\bigr), x^1 + x^2 \geqslant 1\bigr\} \subset C.$$

The reverse inclusion is also true: let c be a canonical communication device inducing a CEP (x^1, x^2, x^3). Consider as before the strategies σ^1 (play T and B with equal probability, independently of what happened) for player 1 and σ^2 (play L and R with equal probability, independently of what happened) for player 2. The point is that $(\sigma^1, \sigma^{2*}, \sigma^{3*})$ and $(\sigma^{1*}, \sigma^2, \sigma^{3*})$ induce the same probability distribution on the sequences of recommendations of the mediator, so they induce the same probability distribution on the sequences of actions played by player 3. For any T denote by λ_1^T, λ_2^T, λ_3^T the induced expected frequencies at which player 3 respectively plays W, M and E up to stage T. At each stage, the payoff of player 1 is 1 if player 3 plays E, and the expected payoff of player 1 given that player 3 plays W is at least $1/2$. Consequently,

$$\gamma_{T,c}^1\bigl(\sigma^1, \sigma^{2*}, \sigma^{3*}\bigr) \geqslant \frac{1}{2}\lambda_1^T + \lambda_3^T.$$

Similarly we obtain

$$\gamma_{T,c}^2\bigl(\sigma^{1*}, \sigma^2, \sigma^{3*}\bigr) \geqslant \frac{1}{2}\lambda_1^T + \lambda_2^T.$$

Hence

$$\gamma_{T,c}^1\bigl(\sigma^1, \sigma^{2*}, \sigma^{3*}\bigr) + \gamma_{T,c}^2\bigl(\sigma^{1*}, \sigma^2, \sigma^{3*}\bigr) \geqslant \lambda_1^T + \lambda_2^T + \lambda_3^T = 1,$$

and the equilibrium condition (ii) of Definition 2.2 gives $\gamma_c^1(\sigma^{1*}, \sigma^{2*}, \sigma^{3*}) + \gamma_c^2(\sigma^{1*}, \sigma^{2*}, \sigma^{3*}) \geqslant 1$. Consequently,

$$C = \bigl\{(x^1, x^2, x^3) \in g\bigl(\Delta(A)\bigr), x^1 + x^2 \geqslant 1\bigr\}.$$

The level of simultaneous punishment is given here by the inequality $x^1 + x^2 \geqslant 1$.

Back to the general case, we give the following definition.

Definition 3.2.

- For each player i, a stage-decision-rule (in short a decision) for player i is a pair of mappings $d^i = (\alpha^i, \mu^i)$ with $\alpha^i : R^i \longrightarrow A^i$ and $\mu^i : R^i \times U^i \longrightarrow M^i$.
- We let D^i be the set of stage-decision-rules for player i.
- Let d^{i*} be the stage-decision-rule for player i such that:

$$\forall \big(a^i, u^i\big) \in R^i \times U^i, \quad \alpha^i\big(a^i\big) = a^i, \quad \mu^i\big(a^i, u^i\big) = u^i.$$

  This decision will be henceforth called the faithful decision of player i.
- Any decision rule $d^i \neq d^{i*}$ will be called a deviation of player i.

The interpretation of a decision-rule d^i is as follows. When player i is told the recommendation r^i, he plays $\alpha^i(r^i)$, and if u^i is the signal he observes, he sends back the message $\mu^i(r^i, u^i)$. D^i is the set of pure stage-strategies of player i in any extended game Γ_c: instead of first observing his recommendation, then choosing an action, observing a signal and sending back a message, player i can equivalently choose an element of D^i and play according to it. What we are doing here is to reduce each stage of the extended game to its normal form.

If $a = (a^k)_{k \in N} \in A$ is recommended to all players and player i plays according to $d^i = (\alpha^i, \mu^i) \in D^i$, whereas the other players play according to a, the joint message received by the mediator is

$$\Big(\big(f^k\big(\alpha^i\big(a^i\big), a^{-i}\big)\big)_{k \neq i}, \mu^i\big(a^i, f^i\big(\alpha^i\big(a^i\big), a^{-i}\big)\big)\Big) \in U.$$

It will be denoted by $\psi^i(d^i, a)$. We will consider mixed decisions, i.e. probabilities over the finite set D^i. Assume now that $a \in A$ is recommended and player i chooses a decision according to some lottery $\delta^i = (\delta^i(d^i))_{d^i \in D^i} \in \Delta(D^i)$. We extend the previous definition and denote by $\psi^i(\delta^i, a)$ the law of the joint message received by the mediator. We have

$$\psi^i\big(\delta^i, a\big) = \sum_{d^i \in D^i} \delta^i\big(d^i\big)\psi^i\big(d^i, a\big) \in \Delta(U),$$

and for any $u = (u^k)_{k \in N}$ in U, the probability of u under $\psi^i(\delta^i, a)$ is the probability under δ^i of the set

$$\Big\{\big(\alpha^i, \mu^i\big) \in D^i,\ \forall k \neq i,\ f^k\big(\alpha^i\big(a^i\big), a^{-i}\big) = u^k \text{ and } \mu^i\big(a^i, f^i\big(\alpha^i\big(a^i\big), a^{-i}\big)\big) = u^i\Big\}.$$

For any player k, we will also denote by $\psi^{i,k}(\delta^i, a)$ the marginal of $\psi^i(\delta^i, a)$ on U^k, i.e. the law of the message reported to the mediator by player k. We will later use $g^k_{\delta^i}(a) = \mathbb{E}_{\delta^i}(g^k(\alpha^i(a^i), a^{-i}))$ for the expected payoff of player k if player i uses δ^i whereas the other players play according to a.

Two mixed decisions $\delta^i \in \Delta(D^i)$ and $\delta^j \in \Delta(D^j)$ will lead to the suspicion of both players i and j if the mediator does not see the difference between {player i deviating with δ^i} and {player j deviating with δ^j}, i.e. if the law of the joint message sent back by all players is the same in both cases: $\psi^i(\delta^i, a) = \psi^j(\delta^j, a)$, $\forall a$. A consequence of this is that the marginals of $\psi^i(\delta^i, a)$ and $\psi^j(\delta^j, a)$ coincide for each a:

– no player $k \neq i, j$ can see the difference between {player i deviating by playing δ^i} and {player j deviating by playing δ^j}: $\psi^{i,k}(\delta^i, a) = \psi^{j,k}(\delta^j, a)$,
– while deviating to δ^i, player i reports the same signals as if j was deviating: $\psi^{i,i}(\delta^i, a) = \psi^{j,i}(\delta^j, a)$,
– while deviating to δ^j, player j reports the same signals as if i was deviating: $\psi^{j,j}(\delta^j, a) = \psi^{i,j}(\delta^i, a)$.

Definition 3.3. Let $J \subset N$ be a subset of players. The set of similar decisions of players in J is

$$SD(J) = \left\{\delta = (\delta^i)_{i\in J} \in \prod_{i\in J} \Delta(D^i),\ \forall i \in J,\ \forall j \in J,\ \psi^i(\delta^i, a) = \psi^j(\delta^j, a)\ \forall a \in A\right\}.$$

If δ is in $SD(J)$ and some player i in J plays according to δ^i, the reported signals give no information about to the mediator about the identity of i. At a canonical communication equilibrium, the mediator will suspect each player of J to have deviated. Thus, there must exist some punishing strategy giving a low payoff simultaneously to every player i in J *if* i *is deviating*. This is where Blackwell's approachability strategy naturally appears. On the other hand, if two decisions are not similar there is an action profile that induce different reported signals for some pair of players. By choosing full support distributions for recommended actions, the mediator can differentiate any two decisions which are not similar.

We always have $(d^{i*})_{i\in J} \in SD(J)$. Since $SD(J)$ is defined via finitely many linear equalities, it is a polytope. Hence $SD(J)$ is a non empty compact convex subset of $\prod_{i\in J} \Delta(D^i)$.

Definition 3.4. The set of jointly rational payoffs is

$$JR = \left\{x \in \mathbb{R}^N,\ \forall q \in \Delta(N),\ x \cdot q \geqslant l(q)\right\},$$

where for any $q \in \Delta(N)$,

$$l(q) = \max_{\delta\in SD(Supp\, q)} \min_{a\in A} \sum_{i\in N} q^i g^i_{\delta^i}(a).$$

For each $q \in \Delta(N)$, the minmax theorem (see, e.g., Raghavan, 1994) gives:

$$l(q) = \min_{p\in\Delta(A)} \max_{\delta\in SD(Supp\, q)} \sum_{a\in A} p(a) \sum_{i\in N} q^i g^i_{\delta^i}(a).$$

Our main result is the following.

Theorem 3.5. *For any repeated game with imperfect monitoring, the set of communication equilibrium payoffs is the set of feasible payoffs that are robust to undetectable deviations and jointly rational*:

$$C = g(\mathcal{P}) \cap JR.$$

We now illustrate the definition of JR with several comments and examples.

3.4.1. Joint rationality implies individual rationality

Let q be the Dirac measure on player i. It is plain that $SD(\{i\}) = \Delta(D^i)$: any deviation of player i makes each player in $\{i\}$ a suspect! Then:

$$\begin{aligned} l(q) &= \min_{p\in\Delta(A)} \max_{\delta^i\in\Delta(D^i)} \sum_{a\in A} p(a) g^i_{\delta^i}(a), \\ &= \min_{p\in\Delta(A)} \max_{\alpha^i : A^i\mapsto A^i} \sum_{a\in A} p(a) g^i\big(\alpha^i\big(a^i\big), a^{-i}\big), \\ &= \min_{p\in\Delta(A)} \sum_{a^i\in A^i} \max_{b^i\in A^i} \sum_{a^{-i}\in A^{-i}} p\big(a^i, a^{-i}\big) g^i\big(b^i, a^{-i}\big), \\ &= w^i. \end{aligned}$$

Hence $(x\cdot q \geqslant l(q)) \Longleftrightarrow (x^i \geqslant w^i)$. Jointly rational payoffs are individually rational.

3.4.2. The two-player case

A specificity of the 2-player case is that both players can be simultaneously punished to their minmax level. Assume that $N = \{1,2\}$. Fix $\bar p^1$ in $\Delta(A^1)$ and $\bar p^2$ in $\Delta(A^2)$ such that $\bar p^1$ (respectively $\bar p^2$) realizes the minimum in $w^1 = \min_{p^1\in\Delta(A^1)} \max_{p^2\in\Delta(A^2)} g^2(p^1\otimes p^2)$ (respectively $w^2 = \min_{p^2\in\Delta(A^2)} \max_{p^1\in\Delta(A^1)} g^1(p^1\otimes p^2)$). Consider the product distribution $\bar p^1\otimes \bar p^2$ in $\Delta(A)$. Then for any $q = (q^1, q^2) \in \Delta(N)$,

$$\begin{aligned} l(q) &\leqslant \min_{p\in\Delta(A)} \max_{(\delta^1,\delta^2)\in\Delta(D^1)\times\Delta(D^2)} \sum_{a\in A} p(a)\big(q^1 g^1_{\delta^1}(a) + q^2 g^2_{\delta^2}(a)\big), \\ &\leqslant \max_{(\delta^1,\delta^2)\in\Delta(D^1)\times\Delta(D^2)} \sum_{(a^1,a^2)\in A} \bar p^1\big(a^1\big)\bar p^2\big(a^2\big)\big(q^1 g^1_{\delta^1}\big(a^1,a^2\big) + q^2 g^2_{\delta^2}\big(a^1,a^2\big)\big). \end{aligned}$$

But for all (δ^1, δ^2),

$$\begin{aligned} &\sum_{(a^1,a^2)\in A} \bar p^1\big(a^1\big)\bar p^2\big(a^2\big)\big(q^1 g^1_{\delta^1}\big(a^1,a^2\big) + q^2 g^2_{\delta^2}\big(a^1,a^2\big)\big) \\ &\quad = q^1 \sum_{a^1} \bar p^1\big(a^1\big) \sum_{a^2} \bar p^2\big(a^2\big) g^1_{\delta^1}\big(a^1,a^2\big) + q^2 \sum_{a^2} \bar p^2\big(a^2\big) \sum_{a^1} \bar p^1\big(a^1\big) g^2_{\delta^2}\big(a^1,a^2\big) \\ &\quad \leqslant q^1 w^1 + q^2 w^2. \end{aligned}$$

Hence we obtain $l(q) \leqslant q^1 w^1 + q^2 w^2$. Since in addition, we proved that individual rationality should always be respected, we deduce that $JR = \{(x^1,x^2)\in\mathbb{R}^2,\ x^1 \geqslant w^1, x^2 \geqslant w^2\} = IR$. For two-player games, we recognize the result from Mertens et al. (1994): $C = g(\mathcal{P}) \cap IR$.

3.4.3. The perfect observation case

If there are at least two players then all deviations can be detected ($\mathcal{P} = \Delta(A)$). For at least three players, as soon as a deviation is detected, the identity of the deviator is clear since a strict majority of players will report the actions actually played to the mediator. Formally, for each $J \subset N$, such that $|J| \geqslant 2$, $SD(J) = \{(d^{i*})_{i\in J}\}$. Consequently, $JR = IR$.

Since we have $JR = IR$ also for two players, for any number of players we get the Folk Theorem: $C = g(\Delta(A)) \cap IR$.

3.4.4. The trivial observation case

No deviation is then detectable: $\forall i \in N\ \forall a^i, b^i \in A^i$, we have $b^i \geqslant a^i$, and for all $J \subset N$, all decisions from players in J are similar: $SD(J) = \prod_{i \in J} \Delta(D^i)$. Consequently, $\mathcal{P}$ is the set of correlated equilibrium distributions of the one-shot game:

$$\mathcal{P} = \Bigg\{ p \in \Delta(A),\ \forall i \in N,\ \forall a^i, b^i \in A^i, \\ \sum_{a^{-i} \in A^{-i}} p\big(a^i, a^{-i}\big) g^i\big(a^i, a^{-i}\big) \geqslant \sum_{a^{-i} \in A^{-i}} p\big(a^i, a^{-i}\big) g^i\big(b^i, a^{-i}\big) \Bigg\}.$$

Moreover, for every $q \in \Delta(N)$ with $Supp\, q = J$,

$$\begin{aligned} l(q) &= \min_{p \in \Delta(A)} \max_{\delta \in \prod_{i \in J} \Delta(D^i)} \sum_{a \in A} p(a) \sum_{i \in J} q^i g^i_{\delta^i}(a), \\ &= \min_{p \in \Delta(A)} \sum_{i \in J} q^i \sum_{a^i \in A^i} \max_{b^i \in A^i} \sum_{a^{-i} \in A^{-i}} p\big(a^i, a^{-i}\big) g^i\big(b^i, a^{-i}\big). \end{aligned}$$

So for p in $\mathcal{P}$,

$$\begin{aligned} g(p) \cdot q &= \sum_{i \in J} q^i g^i(p) = \sum_{i \in J} q^i \sum_{a^i \in A^i} \sum_{a^{-i} \in A^{-i}} p\big(a^i, a^{-i}\big) g^i\big(a^i, a^{-i}\big), \\ &= \sum_{i \in J} q^i \sum_{a^i \in A^i} \max_{b^i \in A^i} \sum_{a^{-i} \in A^{-i}} p\big(a^i, a^{-i}\big) g^i\big(b^i, a^{-i}\big) \geqslant l(q), \end{aligned}$$

and we obtain $C = g(\mathcal{P})$. In case of trivial observation, the set of communication equilibrium payoffs is the set of correlated equilibrium payoffs of the one-shot game.

3.4.5. Back to Example 3.1

Let us see how the characterization solves this example. The interesting case is when 1 and 2 are the suspected players. By definition,

$$SD(\{1, 2\}) = \big\{ \delta = \big(\delta^1, \delta^2\big) \in \Delta\big(D^1\big) \times \Delta\big(D^2\big),\ \forall a \in A,\ \psi^1\big(\delta^1, a\big) = \psi^2\big(\delta^2, a\big) \big\}.$$

Since player 1 has trivial observation, he has nothing to report to the deviator and therefore a decision-rule is well defined by specifying the action function α^1 only. Since player 1 has two actions, D^1 contains four elements which we denote as follows: d^{1*} (faithful decision), d^{1T} (always play T, whatever the recommendation), d^{1B} (always play B, whatever the recommendation), and $d^{1\leftrightarrow}$ (play T if B is recommended and vice-versa). A mixed decision for player 1 is an element of $\Delta(D^1)$, i.e. a probability distribution on $\{d^{1*}, d^{1T}, d^{1B}, d^{1\leftrightarrow}\}$. Notice that several mixed decisions of player 1 may be "equivalent", in the sense that for each recommendation, they induce the same probability distributions on plays. For example, the mixed decision $d^{1T}/2 + d^{1B}/2$ is equivalent to $d^{1*}/2 + d^{1\leftrightarrow}/2$: whatever the recommendation, player 1 plays in both cases T and B with equal probability. In fact, given a mixed decision of player 1, only two numbers are relevant: the probability of playing T when recommended T and the probability of playing T when recommended B. This simply amounts to say that in an extensive game, several mixed strategies may be equivalent with the same behavior strategy. With similar notations we put $D^2 = \{d^{2*}, d^{2L}, d^{2R}, d^{2\leftrightarrow}\}$.

Take $\delta = (\delta^1, \delta^2) \in SD(\{1,2\})$. The probability under δ^1 that player 1 plays T when he is recommended T is $\delta^1(d^{1*}) + \delta^1(d^{1T})$, and we denote this quantity by $\lambda \in [0,1]$. Since players 1 and 2 have trivial observation, we just have to consider the marginals of $\psi^1(\delta^1, a)$ and $\psi^2(\delta^2, a)$ on the set of signals of player 3. For any a in A, $\psi^{1,3}(\delta^1, a) = \psi^{2,3}(\delta^2, a)$.

- Consider the case $a = (T, L, W)$. Under $\psi^{1,3}(\delta^1, (T, L, W))$, u occurs with probability λ and v occurs with probability $1-\lambda$. Since $\psi^{2,3}(\delta^2, (T, L, W)) = \psi^{1,3}(\delta^1, (T, L, W))$, we obtain that the probability under δ^2 of playing L when L is recommended is also λ. So $\delta^2(d^{2*}) + \delta^2(d^{2L}) = \lambda$.
- Consider now $a = (T, R, W)$. $\psi^{2,3}(\delta^2, (T, R, W)) = \psi^{1,3}(\delta^1, (T, R, W))$, which is the probability distribution $\lambda v + (1-\lambda)u$. When R is recommended, δ^2 plays R with probability λ, i.e. $\delta^2(d^{2*}) + \delta^2(d^{2R}) = \lambda$.
- For $a = (B, L, W)$, we obtain that δ^1 plays B with probability λ when B is recommended. So, $\delta^1(d^{1*}) + \delta^1(d^{1B}) = \lambda$.
- $\psi^{1,3}(\delta^1, (B, R, W)) = \lambda u + (1-\lambda)v = \psi^{2,3}(\delta^2, (T, R, W))$, so $a = (B, R, W)$ gives no other condition.
- The other cases for a give no condition, since player 3's signal will always be v.

We have obtained: $\delta^1(d^{1*}) + \delta^1(d^{1T}) = \delta^1(d^{1*}) + \delta^1(d^{1B}) = \delta^2(d^{2*}) + \delta^2(d^{2L}) = \delta^2(d^{2*}) + \delta^2(d^{2R})$, and these conditions are equivalent to: $(\delta^1, \delta^2) \in SD(\{1,2\})$. For example, $(d^{1T}/2 + d^{1B}/2, d^{2L}/2 + d^{2R}/2)$ belongs to $SD(\{1,2\})$. Notice that there are two pure elements in $SD(\{1,2\})$: (d^{1*}, d^{2*}) and $(d^{1\leftrightarrow}, d^{2\leftrightarrow})$.

Let now q be in $\Delta(N)$ with $Supp\, q = \{1,2\}$. For each a in A, $\sum_{i\in N} q^i g^i_{\delta^i}(a) = q^1 g^1_{\delta^1}(a) + q^2 g^2_{\delta^2}(a)$. Using again the notation $\lambda = \delta^1(d^{1*}) + \delta^1(d^{1T})$, we have for example

$$\sum_{i\in N} q^i g^i_{\delta^i}(T, L, W) = q^1\big(0\lambda + 2(1-\lambda)\big) + q^2\big(0\lambda + 2(1-\lambda)\big) = 2(1-\lambda).$$

We represent the mapping $(a \mapsto \sum_{i\in N} q^i g^i_{\delta^i}(a))$ by the following matrices.

W	L	R
T	$2(1-\lambda)$	$q^1(1-\lambda) + 2\lambda q^2$
B	$q^2(1-\lambda) + 2\lambda q^1$	λ

M	L	R
T	q^2	q^2
B	q^2	q^2

E	L	R
T	q^1	q^1
B	q^1	q^1

Since $\sum_{i\in N} q^i g^i_{\delta^i}(a)$ only depends on q, a and λ, we denote this quantity by $l(q, \lambda, a)$, and we have $l(q) = \max_{\lambda\in[0,1]} \min_{a\in A} l(q, \lambda, a)$. $l(q)$ is the value of the following matrix-game, with the row player as the maximizer:

$$\begin{pmatrix} q^1 & q^2 & 0 & 1 & 2q^2 & 2q^1 \\ q^1 & q^2 & 2 & 0 & q^1 & q^2 \end{pmatrix}.$$

The top row of this matrix corresponds to the case $\lambda = 1$, i.e. to the faithful decision (d^{1*}, d^{2*}), whereas choosing the bottom row of the above matrix corresponds to the case $\lambda = 0$, i.e. to the decision $(d^{1\leftrightarrow}, d^{2\leftrightarrow})$ (play the action not recommended). For each λ in $[0, 1]$, $\min_{a\in A} l(q, \lambda, a) = \min\{2(1-\lambda), \lambda, q^1, q^2\} \leqslant \min\{q^1, q^2\}$. Considering $\lambda = 1/2$ gives that $l(q) = \min\{q^1, q^2\}$. Thus the condition $x \cdot q \geqslant l(q)$ for each q in $\Delta(N)$ with $\mathit{Supp}\, q = \{1, 2\}$ becomes: $\forall t \in [0, 1]$, $x^1 t + x^2(1-t) \geqslant \min\{t, 1-t\}$ which is equivalent to $x^1 \geqslant 0$, $x^2 \geqslant 0$ and $x^1 + x^2 \geqslant 1$.

If q in $\Delta(N)$ is such that $\mathit{Supp}\, q = \{1, 3\}$, we have:

$$0 \leqslant l(q) \leqslant \max_{\delta\in\prod_{i\in N}\Delta(D^i)} \min_{a\in A} q^1 g^1_{\delta^1}(a).$$

Taking for example $a = (T, L, M)$ gives $l(q) = 0$. Similarly, $l(q) = 0$ if $\mathit{Supp}\, q = \{2, 3\}$. Finally, since $(\delta^1, \delta^2, \delta^3) \in SD(N)$ implies $(\delta^1, \delta^2) \in SD(\{1, 2\})$, we have if $\mathit{Supp}\, q = \{1, 2, 3\}$:

$$0 \leqslant l(q) \leqslant \max_{(\delta^1,\delta^2)\in SD(\{1,2\})} \min_{a\in A} q^1 g^1_{\delta^1}(a) + q^2 g^2_{\delta^2}(a),$$

so in this case, $l(q) \leqslant \min\{q^1, q^2\}$ and if x is such that $x^1 \geqslant 0$, $x^2 \geqslant 0$ and $x^1 + x^2 \geqslant 1$, we have $x \cdot q \geqslant l(q)$. We thus obtain

$$JR = \{(x^1, x^2, x^3),\ x^1 \geqslant 0,\ x^2 \geqslant 0,\ x^3 \geqslant 0,\ x^1 + x^2 \geqslant 1\}$$

and

$$\begin{aligned} C &= g(\Delta(A)) \cap \{(x^1, x^2, x^3),\ x^1 + x^2 \geqslant 1\} \\ &= \mathrm{conv}\{(0, 2, 0), (0, 1, 0), (1, 0, 0), (2, 0, 0)\}. \end{aligned}$$

3.4.6. Pure and mixed decisions

For $J \subset N$, $SD(J)$ is a polytope, but it might not be the convex hull of pure elements of $SD(J)$. Moreover, in the expression:

$$l(q) = \min_{p\in\Delta(A)} \max_{\delta\in SD(\mathit{Supp}\, q)} \sum_{a\in A} p(a) \sum_{i\in N} q^i g^i_{\delta^i}(a),$$

one can not in general replace $SD(\mathit{Supp}\, q)$ by the set of pure elements of $SD(\mathit{Supp}\, q)$. The following example illustrates these facts.

Example 3.6. Consider, as in Example 3.1, three players with trivial observation for players 1 and 2 and the following observation for player 3:

W

	L	C	R
T	u	v	u
M	v	u	v
B	u	v	u

MW

	L	C	R
T	u	u	v
M	v	u	v
B	u	v	u

ME

	L	C	R
T	w	w	w
M	w	w	w
B	w	w	w

	L	C	R
T	z	z	z
M	z	z	z
B	z	z	z

E

We assume that payoff vectors can be deduced from signals of player 3: signal u gives $(0,0,0)$, signal v gives $(1,1,0)$, signal w gives $(0,1,0)$ and signal z gives $(1,0,0)$.

The point here is that if player 1 plays $\frac{1}{2}M+\frac{1}{2}B$, or if player 2 plays $\frac{1}{2}C+\frac{1}{2}R$, it induces the law of signals $\frac{1}{2}u+\frac{1}{2}v$ as soon as player 3 plays W or MW. We compute $SD(\{1,2\})$. Again we observe that, since players 1 and 2 have trivial observation, we do not have to specify the signals they report. Take $\delta=(\delta^1,\delta^2)\in\Delta(D^1)\times\Delta(D^2)$ such that $\forall a\in A$, $\psi^{1,3}(\delta^1,a)=\psi^{2,3}(\delta^2,a)$.

- Taking $a=(T,L,W)$ gives the equality between the probabilities under δ^1 that $\alpha^1(T)=M$ and under δ^2 that $\alpha^2(L)=C$. We denote this probability by $\lambda/2$.
- $a=(T,L,MW)$ gives $\mathbb{P}_{\delta^1}(\alpha^1(T)=M)=\mathbb{P}_{\delta^2}(\alpha^2(L)=R)$.
- Trying all other values of a, we obtain that $\lambda\in[0,1]$.

We get $\delta=(1-\lambda)(d^{1*},d^{2*})+\lambda(\hat{d}^1,\hat{d}^2)$, where $\hat{d}^1$ plays with probability $1/2$ always M and with probability $1/2$ always B, and $\hat{d}^2$ plays with probability $1/2$ always C and with probability $1/2$ always R. Such a δ cannot be obtained as a convex combination of pure elements. The only pure decision pair in $SD(\{1,2\})$ is (d^{1*},d^{2*}).

Player 3 can punish player 1 by playing ME and player 2 by playing E, hence $w^1=w^2=0$. If $Supp\, q=\{1,2\}$, $l(q)=\max_{\lambda\in[0,1]}\min_{a\in A} l(q,\lambda,a)$, with

$$l(q,\lambda,a)=q^1 g^1_{(1-\lambda)d^{1*}+\lambda\hat{d}^1}(a)+q^2 g^2_{(1-\lambda)d^{2*}+\lambda\hat{d}^2}(a).$$

Fixing λ, $(a\mapsto l(q,\lambda,a))$ is given by the following matrices, where $s=\lambda/2$ and $t=1-\lambda/2$.

W:

	L	C	R
T	s	t	s
M	t	s	t
B	s	t	s

MW:

	L	C	R
T	s	s	t
M	t	s	t
B	s	t	s

ME:

	L	C	R
T	q^2	q^2	q^2
M	q^2	q^2	q^2
B	w	w	w

E:

	L	C	R
T	q^1	q^1	q^1
M	q^1	q^1	q^1
B	q^1	q^1	q^1

So $\forall\lambda$, $\min_{a\in A} l(q,\lambda,a)=\min\{\frac{\lambda}{2},q^1,q^2\}$ and $l(q)=\min\{q^1,q^2\}$. If $(x^1,x^2,0)\in C$, $x^1q^1+x^2q^2\geqslant\min\{q^1,q^2\}$ for all $q\in\Delta(\{1,2\})$, and this is equivalent to $x^1\geqslant 0$, $x^2\geqslant 0$ and $x^1+x^2\geqslant 1$. We thus have here:

$$C=\left\{\left(x^1,x^2,x^3\right)\in g\left(\Delta(A)\right), x^1+x^2\geqslant 1\right\}=\operatorname{conv}\{(0,1,0),(1,0,0),(1,1,0)\}.$$

3.4.7. Structure of the set of communication equilibrium payoffs

It is plain that C is a convex compact subset of $\mathbb{R}^N$. $\mathcal{P}$ is indeed a polytope and so is $g(\mathcal{P})$. For two-player games, C is thus also a polytope. But in the general case (at least three players), the definition of JR involves infinitely many affine inequalities. The following example shows that C need not be a polytope.

Example 3.7. Consider a three-player game, $N = \{1, 2, 3\}$ where each player has two actions. Player 3 has payoff 0 and players 1 and 2 have trivial observation. The signal for player 3 is given by the following matrices.

W:

	L	R
T	u	v
B	v	v

E:

	L	R
T	u	v
B	v	v

The payoffs are the following.

W:

	L	R
T	$(0, 0, 0)$	$(0, 1, 0)$
B	$(0, 1, 0)$	$(1, 1, 0)$

E:

	L	R
T	$(1, 0, 0)$	$(1, 0, 0)$
B	$(0, 1, 0)$	$(1, 1, 0)$

First note that players 2 and 3 can punish player 1 by playing L and W. Hence $w^1 = 0$. Similarly, $w^2 = 0$. Since $\mathcal{P} = \Delta(A)$, $g(\mathcal{P}) \cap IR = g(\Delta(A))$. We now compute $SD(\{1, 2\})$. Let $\delta = (\delta^1, \delta^2)$ be such that for each a in A, we have $\psi^{1,3}(\delta^1, a) = \psi^{2,3}(\delta^2, a)$.

- Consider the case $a = (B, L, W)$. $\psi^{2,3}(\delta^2, (B, L, W))$ is the Dirac mass on v and thus so is $\psi^{1,3}(\delta^1, (B, L, W))$. Hence δ^1 assigns probability one to the elements such that $\alpha^1(B) = B$.
- Considering now $a = (T, R, W)$ gives that δ^2 assigns probability one to the elements such that $\alpha^2(R) = R$.
- For $a = (T, L, W)$, we obtain the equality between the probability under δ^1 that $\alpha^1(T) = B$ and the probability under δ^2 that $\alpha^2(L) = R$.
- The other cases for a give no condition, since player 3's signal will always be v.

Thus we obtain $SD(\{1, 2\}) = \{\lambda d^* + (1 - \lambda)\hat{d}, \lambda \in [0, 1]\}$, where $d^* = (d^{1*}, d^{2*})$ is the faithful decision and $\hat{d} = (d^{1B}, d^{2R})$ where d^{1B} always plays B and d^{2R} always plays R. Let $q = (q^1, q^2)$ be in $\Delta(\{1, 2\})$. Then $l(q) = \max_{\lambda \in [0,1]} \min_{a \in A} l(q, \lambda, a)$, with

$$\begin{aligned} l(q, \lambda, a) &= q^1 g^1_{\lambda d^{1*} + (1-\lambda)\hat{d}^1}(a) + q^2 g^1_{\lambda d^{2*} + (1-\lambda)\hat{d}^2}(a), \\ &= q^1\bigl(\lambda g^1(a) + (1 - \lambda) g^1\bigl(B, a^{-1}\bigr)\bigr) + q^2\bigl(\lambda g^2(a) + (1 - \lambda) g^2\bigl(R, a^{-2}\bigr)\bigr). \end{aligned}$$

$(a \mapsto l(q, \lambda, a))$ is given by following the matrices.

W:

	L	R
T	$(1 - \lambda)q^2$	$(1 - \lambda)q^1 + q^2$
B	q^2	1

E:

	L	R
T	$q^1\lambda$	q^1
B	q^2	1

Hence $l(q) = \max_{\lambda \in [0,1]} \min\{q^1\lambda, (1-\lambda)q^2\} = q^1 q^2$. We then obtain:

$$JR = \left\{(x^1, x^2, x^3) \in \mathbb{R}^3_+,\ x^1 t + x^2(1-t) \geqslant t(1-t)\ \forall t \in [0,1]\right\}.$$

A simple computation shows that $JR = \{(x^1, x^2, x^3) \in \mathbb{R}^3_+,\ \sqrt{x^1} + \sqrt{x^2} \geqslant 1\}$. Thus,

$$C = \left\{(x^1, x^2, x^3) \in g(\Delta(A)),\ \sqrt{x^1} + \sqrt{x^2} \geqslant 1\right\},$$

which is not a polytope. It is here much more complicated to construct simultaneous punishments of players 1 and 2. We now give an intuition for the construction of a canonical communication device c yielding the CEP $(1/4, 1/4, 0)$.

First consider an infinite play $(a_1, a_2, \ldots, a_t, \ldots)$ composed of (T, L, W) with a frequency of $3/4$ and of (B, R, W) with a frequency of $1/4$. As long as the signal reported by player 3 does not prove that a deviation has occurred, c recommends to play a_t at every stage t. This gives a payoff of $(1/4, 1/4, 0)$ if no player deviates. As soon as player 3 reports the signal v at some stage where (T, L, W) was recommended, it is clear that a deviation has occurred. Since player 3 has payoff 0, players 1 and 2 only have to be suspected. To avoid profitable deviations, c will then play a punishing strategy giving simultaneously players 1 and 2 no more than $1/4$. The mediator approaches the set of payoffs $\{(x_1, x_2, x_3), x_1 \leqslant 1/4, x_2 \leqslant 1/4\}$ in the game with vector payoffs given by player 1 and player 2's payoffs. This strategy roughly unfolds as follows. At each stage, c recommends to play the action (T, L, W) with some probability p and the action (T, L, E) with probability $1 - p$. So at every stage player 1 is asked to play T and player 2 is asked to play L and the deviating player can not condition his play on the action recommended to player 3, having no information about it. The communication device (or the mediator) computes then after each stage the frequency λ in $[0, 1]$ of stages where the signal u was reported by player 3. If player 1 is deviating, his average payoff is approximately $(1 - p)\lambda$. If player 2 is deviating, his payoff is approximately $p(1 - \lambda)$. As the number of stages goes to infinity, the mediator adapts this strategy so as to have p close to λ. Is is thus possible to control the vector payoff (x^1, x^2), x^1 being the payoff of player 1 in case he is deviating and x^2 similarly being the payoff of player 2 in case he is deviating, in order to get $(x^1, x^2) \in \{((1-\lambda)\lambda, \lambda(1-\lambda)), \lambda \in [0, 1]\}$. So $x^1 \leqslant 1/4$ and $x^2 \leqslant 1/4$, and no deviation is profitable.

4. An auxiliary 2-player game with incomplete information and the proof

We first present the well-known model of 2-player repeated games with incomplete information, where one of the players is fully informed about the state of nature. Such games are called 2-player repeated games with lack of information on one side, and were introduced by Aumann and Maschler in the sixties (their work, also introducing games with lack of information on both sides, can be found in Aumann and Maschler, 1995 book). In Section 4.2, we show how C can be seen as the set of some equilibrium payoffs of such a game. Finally, we describe this set and derive our characterization of C from it.

4.1. Two-player repeated games with lack of information on one side

In these games, we have two players: player I is called the informed player whereas player II is called the uninformed player. There is a set of states K, and an initial probability q_0 on K. Initially, nature chooses some state k according to q_0. k is then fixed and told to player I, not to player II. Then, at each stage $t = 1, 2, \ldots,$ both players simultaneously select an action in their own set of actions and observe some signal before starting stage $t + 1$. Stage payoffs are not necessarily observed.

The sets of stage actions and of signals for player I may depend on the selected state k, and will respectively be denoted by D^k and V^k. E (respectively W) will stand for player II's set of actions (respectively signals). Payoffs functions in state k are $G^k : D^k \times E \longrightarrow \mathbb{R}$ for player I, and $H^k : D^k \times E \longrightarrow \mathbb{R}$ for player II. Similarly, signalling functions in state k will be denoted by $\varphi^k : D^k \times E \longrightarrow V^k$ for player I and by $\psi^k : D^k \times E \longrightarrow W$ for player II. All sets of states, actions and signals are assumed to be non empty and finite. We also assume that q_0 has full support. Strategies and equilibria are defined as in Section 2. A behavior strategy for player I will be denoted by $\sigma = (\sigma^k)_{k\in K}$, where for each k, σ^k is the strategy used if the state is k. We denote by $\mathcal{S}$ (respectively $\mathcal{T}$) the set of behavior strategies for player I (respectively II). A joint behavior strategy (σ, τ) induces, for any state k, a probability $\mathbb{P}^k_{\sigma,\tau}$ over the set $(D^k \times E \times V^k \times W)^\infty$ of infinite sequences of actions and signals played if the state is k and (σ^k, τ) is played. q_0 and (σ, τ) also induce a probability $\mathbb{P}_{\sigma,\tau}$ over the set of plays $\{(k, h), k \in K, h \in (D^k \times E \times V^k \times W)^\infty\}$. The expected payoffs are then, for all $T \geqslant 1$:

$$\gamma_T^{I,k}(\sigma,\tau) = \mathbb{E}_{\mathbb{P}^k_{\sigma,\tau}}\left(\frac{1}{T}\sum_{t=1}^{T} G^k(d_t, e_t)\right), \qquad \gamma_T^{II,k}(\sigma,\tau) = \mathbb{E}_{\mathbb{P}^k_{\sigma,\tau}}\left(\frac{1}{T}\sum_{t=1}^{T} H^k(d_t, e_t)\right),$$

$$\gamma_T^{I}(\sigma,\tau) = \mathbb{E}_{\mathbb{P}_{\sigma,\tau}}\left(\frac{1}{T}\sum_{t=1}^{T} G^k(d_t, e_t)\right) = \sum_{k\in K} q_0^k \gamma_T^{I,k}(\sigma,\tau),$$

$$\gamma_T^{II}(\sigma,\tau) = \mathbb{E}_{\mathbb{P}_{\sigma,\tau}}\left(\frac{1}{T}\sum_{t=1}^{T} H^k(d_t, e_t)\right) = \sum_{k\in K} q_0^k \gamma_T^{II,k}(\sigma,\tau).$$

Definition 4.1. A (uniform) equilibrium of the repeated game with lack of information on one side is a joint strategy (σ, τ) such that:

(i) for each state k, $(\gamma_T^{I,k}(\sigma,\tau))_{T\geqslant 1}$ converges as T goes to infinity to some $\gamma^{I,k}(\sigma,\tau)$, and $(\gamma_T^{II,k}(\sigma,\tau))_{T\geqslant 1}$ converges as T goes to infinity to some $\gamma^{II,k}(\sigma,\tau)$,

(ii) for all $\varepsilon > 0$, there exists T_0 s.t. (σ, τ) is an ε-Nash equilibria in finitely repeated games with at least T_0 stages, that is $\forall T \geqslant T_0$:

$$\gamma_T^{I}(\sigma',\tau) \leqslant \gamma_T^{I}(\sigma,\tau) + \varepsilon, \quad \forall \sigma' \in \mathcal{S},$$
$$\gamma_T^{II}(\sigma,\tau') \leqslant \gamma_T^{II}(\sigma,\tau) + \varepsilon, \quad \forall \tau' \in \mathcal{T}.$$

$((\gamma^{I,k}(\sigma,\tau))_{k\in K}, \gamma^{II}(\sigma,\tau)) \in \mathbb{R}^K \times \mathbb{R}$ is then called an equilibrium payoff of the repeated game.

As usual in games with incomplete information, a strategy for the informed player is an ε-best response if and only if it is so in each state which has positive probability. It is thus equivalent to replace in (ii) the condition for player I by:

$$\forall \varepsilon > 0,\ \exists T_0,\ \forall T \geqslant T_0,\ \forall k \in K, \quad \gamma_T^{I,k}(\sigma', \tau) \leqslant \gamma^{I,k}(\sigma, \tau) + \varepsilon, \quad \forall \sigma' \in \mathcal{S}.$$

4.2. *An auxiliary game with lack of information on one side*

We associate here to the original repeated game Γ defined in Section 2, an auxiliary game Γ_{inc} with lack of information on one side.

The informed player (player I) is called the cheater, the uninformed player (player II) represents by the mediator. The set of states K is defined as the original set of players N and the initial probability q_0 as the uniform probability on K. The set of actions for the cheater in state k will be the set D^k of decisions of player k defined in Section 3.4, whereas the set E of actions for the mediator will be the original set of joint actions A. The analogy between the original game Γ and the auxiliary game Γ_{inc} is the following. The selected state represents the deviating player in the original game. At each stage, the mediator selects a joint action representing his recommendation in Γ, whereas the cheater selects some decision d^k if the state is k representing the deviation of player k in Γ. Payoffs for the cheater in state k are given by payoffs for player k in Γ: $G^k(d^k, a) = g^k(\alpha^k(a^k), a^{-k})$ for any state k, $d^k = (\alpha^k, \mu^k) \in D^k$ and $a \in A$. The mediator has payoff $H^k(d^k, a) = 0$ for any k, d^k and a. Signals are similarly defined as follows. For the cheater in state k, it consists of the recommendation of the mediator and the signal of player k in Γ: $\varphi^k(d^k, a) = (a^k, f^k(\alpha^k(a^k), a^{-k}))$ for any state k, $d^k \in D^k$ and $a \in A$, and thus the set of signals V^k is $A^k \times U^k$. For the mediator, the signal consists of the messages sent back by all players: $W = U$, and $\psi^k(d^k, a) = ((f^j(\alpha^k(a^k), a^{-k}))_{j \neq k}, \mu^k(a^k, f^k(\alpha^k(a^k), a^{-k})))$. Notice that the notation $\psi^k(d^k, a)$ is the same as in Section 3.4.

With respect to general repeated games with lack of information on one side, Γ_{inc} has specific features that are exploited in the sequel. Notice that player I has a special strategy: we define d^{**} as the strategy that plays in each state k the faithful decision d^{k*}, at every stage regardless of what happened before. Notice also that the set of behavior strategies for player II is $\mathcal{T} = \{\tau = (\tau_t)_{t \geqslant 1}$, with for each t, $\tau_t : (A \times U)^{t-1} \longrightarrow \Delta(A)\}$, and we identify this set with the set of canonical communication devices defined in Section 2. The following result is the main interest of our auxiliary game.

Proposition 4.2. (1) *Let c be a canonical communication device. Then, the faithful strategy σ^* is an equilibrium of Γ_c if and only if (d^{**}, c) is an equilibrium of Γ_{inc}.*

(2) $C = \{x \in \mathbb{R}^N,\ \exists c \in \mathcal{T}$ *with* (d^{**}, c) *an equilibrium of* Γ_{inc} *with payoff* $(x, 0)\}$.

Proof. Let c be in $\mathcal{T}$. We first consider the two following situations:

(a) Γ_c is played and every player i uses his faithful strategy σ^{i*}.

(b) Γ_{inc} is played, and player I uses d^{**} whereas player II uses c.

It is just a matter of notations to see that (a) and (b) induce the same probability distributions over streams of payoffs (for all players in Γ_c and for the vector payoff of player I in Γ_{inc}). Hence, for all T, $(\gamma_{T,c}^i(\sigma^*))_{i \in N} = (\gamma_T^{I,i}(d^{**}, c))_{i \in N}$. So condition (i) of Definition 2.2 is equivalent to that of Definition 4.1. And (2) here will then follow from (1).

Considering unilateral deviations, we can first restrict ourselves to pure strategies. Fix i in N. In Γ_{inc}, a pure strategy for player I in state i is an element $\sigma^i = (\sigma^i_t)_{t\geqslant 1}$ with for each t, $\sigma^i_t : (D^i \times V^i)^{t-1} \longrightarrow D^i$ giving the action played at stage t by player I if the state is i, depending on the first $(t-1)$ actions he played and the first $(t-1)$ signals he received. While using a fixed pure strategy, player I can deduce his past actions from his past signals, so one can equivalently think of pure strategies as elements $\sigma^i = (\sigma^i_t)_{t\geqslant 1}$ with for each t, $\sigma^i_t : (V^i)^{t-1} \longrightarrow D^i$. We now consider a pure strategy for player i in Γ_c. It is defined as an element $\sigma^i = ((\sigma^i_t, \xi^i_t)_{t\geqslant 1})$ where for each t:

- $\sigma^i_t : (R^i \times A^i \times U^i \times M^i)^{t-1} \times R^i \longrightarrow A^i$,
- $\xi^i_t : (R^i \times A^i \times U^i \times M^i)^{t-1} \times R^i \times A^i \times U^i \longrightarrow M^i$.

Similarly, we can restrict ourselves to elements $\sigma^i = (\sigma^i_t, \xi^i_t)_{t\geqslant 1}$ with for each t, $\sigma^i_t : (R^i \times U^i)^{t-1} \times R^i \longrightarrow A^i$ and $\xi^i_t : (R^i \times U^i)^{t-1} \times R^i \times U^i \longrightarrow M^i$. Instead of considering several steps (first choose an action, then a message), we can also switch to the normal form for each stage t. This is equivalent with associating to each element of $(R^i \times U^i)^{t-1}$ a mapping from R^i to A^i and a mapping from $R^i \times U^i$ to M^i, that is an element of D^i. Since by definition $R^i = A^i$, and $V^i = A^i \times U^i$, we obtain a unique set of pure reduced strategies for player I in state i in Γ_{inc} and for player i in Γ_c, which is the set of $\sigma^i = (\sigma^i_t)_{t\geqslant 1}$ with for each t, $\sigma^i_t : (A^i \times U^i)^{t-1} \longrightarrow D^i$. We fix such σ^i, and consider again two situations:

(a) Γ_c is played, all players except i use their faithful strategy whereas player i deviates from σ^i.

(b) Γ_{inc} is played, the state is i, player I plays σ^i whereas player II plays c.

Again, (a) and (b) induce the same probability distributions over streams of payoffs (for player i in Γ_c and for player I in Γ_{inc}), so for all T, $\gamma^i_{T,c}((\sigma^{j*})_{j\neq i}, \sigma^i) = \gamma^{I,i}_T(\sigma, c)$ where σ is any strategy of player I that plays σ^i in state i. So conditions (ii) of Definitions 2.2 and 4.1 are equivalent, and Proposition 4.2 is proved. □

4.3. A theorem for some games with lack of information on one side

We now concentrate on repeated games with lack of information on one side. We start with a repeated game Γ_{inc} as defined in Section 4.1. Γ_{inc} is described by a set of states K, an initial probability q_0 with full support on K, sets of actions $(D^k)_{k\in K}$ and E, sets of signals $(V^k)_{k\in K}$ and W, and for each state k payoffs functions G^k and H^k and signalling functions φ^k and ψ^k.

We make the following important assumptions:

- player II has payoff 0, hence $H^k(d^k, e) = 0$ $\forall d^k \in D^k$ and $e \in E$. So the repeated game is with known own payoffs. Moreover, there is no need to check optimality for player II;
- there exist pure actions of player I, $d^{k*} \in D^k$ for each k, that induce the same signals for player II: $\forall k, k' \in K$, $\psi^k(d^{k*}, e) = \psi^{k'}(d^{k'*}, e)$ for each e in E. We denote by d^{**} the strategy of player I that plays at each stage, whatever happens, d^{k*} if the state is k;

and we want to characterize the set

$$\widetilde{C} = \left\{x \in \mathbb{R}^K,\ \exists \tau \in \mathcal{T} \text{ with } (d^{**}, \tau) \text{ an equilibrium of } \Gamma_{inc} \text{ with payoff } (x, 0)\right\}.$$

We are thus interested in strategies τ such that (d^{**}, τ) is an equilibrium. Notice that since for each state, d^{**} plays the same action at each stage whatever happens, we do not have to care about player I's observations.

As in Aumann and Maschler (1995), we introduce the set of non-revealing strategies for player I. If the state is k and at some stage he plays according to the lottery $\delta^k \in \Delta(D^k)$ whereas player II plays some e in E, the law of the signal received by player II is $\sum_{d^k \in D^k} \delta^k(d^k)\psi^k(d^k, e) \in \Delta(W)$. Denote by $\delta^k\psi^k$ the vector $(\sum_{d^k \in D^k} \delta^k(d^k)\psi^k(d^k, e))_{e \in E}$ and define, $\forall q \in \Delta(K)$:

$$NR(q) = \left\{\delta = \left(\delta^k\right)_{k \in K} \in \prod_{k \in K} \Delta\left(D^k\right),\ \forall k, k' \in Supp\, q,\ \delta^k\psi^k = \delta^{k'}\psi^{k'}\right\}.$$

Suppose that $\delta \in NR(q)$ and that the selected state is in $Supp\, q$. If player I plays according to δ^k in state k, player II will receive no further information about the selected state. δ is then called a non-revealing strategy at q. Because of the existence of $d^* = (d^{k*})_{k \in K}$ which induces state independent signals, $NR(q)$ is non-empty and it is a convex compact subset of $\prod_{k \in K} \Delta(D^k)$. The non-revealing payoff function is now defined as:

$$\begin{aligned}\forall q \in \Delta(K), \quad \tilde{l}(q) &= \max_{\delta \in NR(q)} \min_{p \in \Delta(E)} \sum_{k \in K} q^k G^k\left(\delta^k, p\right) \\ &= \min_{p \in \Delta(E)} \max_{\delta \in NR(q)} \sum_{k \in K} q^k G^k\left(\delta^k, p\right),\end{aligned}$$

where $G^k(\delta^k, p)$ is the expected payoff of player I if the probabilities δ^k and p are played.

Using Blackwell's (1956) approachability, Kohlberg (1975) proved the following result. Let $\Gamma_0(q_0)$ be the zero-sum game with incomplete information that is described as Γ_{inc} except that player II wants to minimize player I's payoff. Let $\operatorname{cav}\tilde{l}$ be the least concave function on $\Delta(K)$ which is pointwise greater than $\tilde{l}$. Let $\alpha \in \mathbb{R}^K$ be such that $\forall q \in \Delta(K)$, $\alpha \cdot q \geqslant \tilde{l}(q)$ and $\alpha \cdot q_0 = \operatorname{cav}\tilde{l}(q_0)$. Such a vector always exists since $\operatorname{cav}\tilde{l}$ is concave and continuous.

Theorem 4.3 (Kohlberg, 1975). *The value of $\Gamma_0(q_0)$ exists and is $\operatorname{cav}\tilde{l}(q_0)$. Player II has an optimal strategy that approaches the set $\{\beta \in \mathbb{R}^K,\ \forall k, \beta^k \leqslant \alpha^k\}$. That is, there exists some strategy $\bar{\tau}$ of player II such that*:

$$\forall \varepsilon > 0, \exists T_0,\ \forall T \geqslant T_0,\ \forall \sigma \in \mathcal{S},\ \forall k \in K, \quad \gamma_T^{I,k}(\sigma, \bar{\tau}) \leqslant x^k + \varepsilon.$$

In general, call approachable a vector x in $\mathbb{R}^K$ for which there exists some strategy $\bar{\tau}$ of player II such that:

$$\forall \varepsilon > 0, \exists T_0,\ \forall T \geqslant T_0,\ \forall \sigma \in \mathcal{S},\ \forall k \in K, \quad \gamma_T^{I,k}(\sigma, \bar{\tau}) \leqslant \alpha^k + \varepsilon.$$

It is plain that any vector x which is coordinate-wise greater than α, is also approachable. We will thus use Kohlberg's theorem in the following form.

Lemma 4.4. *For each* $x = (x^k)_{k \in K}$ *in* $\mathbb{R}^K$*:*

$$x \text{ is approachable} \quad \text{if and only if} \quad x \cdot q \geqslant \tilde{l}(q) \quad \forall q \in \Delta(K).$$

In other words, player II can force player I's long term average expected payoff to be not greater than x^k, simultaneously for each k. This type of strategy will be used as a punishing strategy of player II against player I.

Remark. Note that in such repeated games with lack of information on one side and state dependent signals, player *II* may receive signals which are inconsistent with some states. In such a case, he should only care about payoffs in states which are still possible. This is true in particular when he punishes his opponent via Blackwell's approachability strategy. Back to the analogy with our original game, the set of still possible states exactly represents the set of suspected players i.e. those players who have a deviation consistent with the signals observed by the mediator.

We now characterize $\widetilde{C}$.

Theorem 4.5. $\widetilde{C} = \{x \in \mathbb{R}^K,\ \exists p \in \Delta(E)$ *s.t.*

$$\begin{aligned}
&\text{(i) } \forall q \in \Delta(K),\ x \cdot q \geqslant \tilde{l}(q),\\
&\text{(ii) } \forall k \in K,\ \ x^k = G^k(d^{k*}, p),\\
&\text{(iii) } \forall k \in K\ \forall d^k \in D^k \text{ s.t. } d^k\psi^k = d^{k*}\psi^k,\ G^k(d^{k*}, p) \geqslant G^k(d^k, p)\}.
\end{aligned}$$

Proof. We proceed by double inclusion.

(1) Let x be in $\widetilde{C}$, and consider a strategy τ of player II such that (d^{**}, τ) is an equilibrium of Γ_{inc} with payoff $(x, 0)$. From the Definition 4.1 of equilibrium, player I should not get more than x^k in state k, that is, τ is such that

$$\forall \varepsilon > 0,\ \exists T_0,\ \forall T \geqslant T_0,\ \forall \sigma \in \mathcal{S},\ \forall k \in K, \quad \gamma_T^{I,k}(\sigma, \tau) \leqslant x^k + \varepsilon.$$

From Lemma 4.4, we get $x \cdot q \geqslant \tilde{l}(q)\ \forall q \in \Delta(K)$.

For any state k in K, define $\forall T \geqslant 1$, $p_T(k)$ as the expectation of the average action played by player II up to stage T if player I uses d^{**} and player II uses τ, given that the state is k: $p_T(k) = (p_T(k)(e))_{e \in E} \in \Delta(E)$, with $\forall e \in E$,

$$p_T(k)(e) = \mathbb{E}_{\mathbb{P}^k_{d^{**},\tau}}\left(\frac{|\{t \leqslant T, e_t = e\}|}{T}\right).$$

Since d^{**} always plays the same action and under this action the signals for player *II* are independent on the state, we have that $p_T(k)$ actually does not depend on k. We thus set $p_T = p_T(k) \in \Delta(E)$, and consider a cluster point p of the bounded sequence $(p_T)_{T \geqslant 1}$. By definition, for each state k,

$$\begin{aligned}
x^k &= \lim_{T\to\infty} \mathbb{E}_{\mathbb{P}^k_{d^{**},\tau}}\left(\frac{1}{T}\sum_{t=1}^{T} G^k(d^{k*}, e_t)\right)\\
&= \lim_{T\to\infty} G^k(d^{k*}, p_T)\\
&= G^k(d^{k*}, p).
\end{aligned}$$

Hence (ii) is proved.

Assume now for the sake of contradiction that (iii) is not satisfied. There exists some state k and d^k in D^k with $d^k\psi^k = d^{k*}\psi^k$ and $G^k(d^{k*}, p) < G^k(d^k, p)$. Consider the deviation σ of player I consisting of playing at each stage d^k if the state is k, and $d^{k'*}$ if the state is $k' \neq k$. This deviation can not be detected by player II, hence

$$\mathbb{E}_{\mathbb{P}^k_{\sigma,\tau}}\left(\frac{|\{t \leqslant T, e_t = e\}|}{T}\right) = p_T(e)$$

for each e, and $\gamma_T^{I,k}(\sigma, \tau) = G^k(d^k, p_T)$ for all T. Point (ii) of Definition 4.1 is then contradicted.

(2) Let x in $\mathbb{R}^K$ and p in $\Delta(E)$ be such that (i), (ii), and (iii) are satisfied. We construct τ such that (d^{**}, τ) is an equilibrium of Γ_{inc} with payoff $(x, 0)$. Notice that since player I is supposed to play at each stage d^{k*} if the state is k, knowing his action e player II knows which signal he should receive. Let for each stage T, $p^T = (1 - 1/\sqrt{T})p + (1/\sqrt{T})\hat{p}$, where $\hat{p}$ is the uniform distribution on E. The strategy τ is as follows:

- play p^1 at stage 1. If at each stage $t < T$, the signal received was $\psi^k(d^{k*}, e_t)$ for some k, play according to p^T at stage T. Here, $e_t \in E$ denotes the action played by player II at stage t and recall that $\psi^k(d^{k*}, e_t)$ is independent of k.
- if at some stage t the signal received is different from $\psi^k(d^{k*}, e_t)$, play for the remaining of the game according to a strategy $\bar{\tau}$ given by Lemma 4.4 (this strategy exists because of (i)). In other words, punish player I for ever, using the approachability strategy for zero-sum games with signals.

We finally show that (d^{**}, τ) is an equilibrium with payoff $(x, 0)$. Since $(p^T)_{T \geqslant 1}$ converges to p, it is clear that (d^{**}, τ) yields the payoff $x^k = G^k(d^{k*}, p)$ in each state k. It remains to prove that player I plays a best response.

Fix $\varepsilon > 0$, and let $\overline{M}$ be an upper bound for all absolute values of payoffs. Using Lemma 4.4, let $\overline{T}$ be such that: $\forall T \geqslant \overline{T}$, $\forall \sigma \in \mathcal{S}$, $\forall k \in K$, $\gamma_T^{I,k}(\sigma, \bar{\tau}) \leqslant x^k + \varepsilon$ and $\overline{T} \geqslant (\overline{M}/\varepsilon)^2$. Define now T_0 such that $T_0 \geqslant \overline{T}/\varepsilon$ and for all $T \geqslant T_0$, $T\exp(-\varepsilon\sqrt{T}/|E|) \leqslant \varepsilon$. Fix $T \geqslant T_0$, and σ a strategy for player I. We consider the T-stages game and prove that: $\forall k \in K$, $\gamma_T^{I,k}(\sigma, \tau) \leqslant x^k + \varepsilon(1 + 4\overline{M})$.

Fix $k \in K$, and consider the probability $\mathbb{P}^k_{\sigma,\tau}$ induced by σ and τ when k is the selected state. We define the random variable $Z \in \{0, \ldots, T\}$ as the number of stages t in $\{1, \ldots, T\}$ where player II is not punishing player I, and player I is deviating in a way that could be detected by player II while not being detected (i.e. his action d_t satisfies $d_t\psi^k \neq d^{k*}\psi^k$, and $\psi^k(d_t, e_t) = \psi^k(d^{k*}, e_t)$).

Note that if at some stage t player II plays according to p^t and player I deviates by playing some action d_t such that $d_t\psi^k \neq d^{k*}\psi^k$, the probability that player II will detect player I's deviation (i.e. plays some e_t such that $\psi^k(d_t, e_t) \neq \psi^k(d^{k*}, e_t)$) is at least $1/(|E|\sqrt{t}) \geqslant 1/(|E|\sqrt{T})$. Consequently, $\mathbb{P}^k_{\sigma,\tau}(Z = z) \leqslant (1 - 1/|E|\sqrt{T})^z$.

We condition player I's payoff according to the values of z:

$$\gamma_T^{I,k}(\sigma, \tau) = \sum_{z=0}^{T} \mathbb{P}^k_{\sigma,\tau}(Z = z)\mathbb{E}_{\mathbb{P}^k_{\sigma,\tau}}\left(\frac{1}{T}\sum_{t=1}^{T} G^k(d_t, e_t) \,\Big|\, Z = z\right).$$

We compute an appropriate upper bound for $\gamma_T^{I,k}(\sigma,\tau)$. The point is that low values of z do not give too important payoffs for player I, and high values of z occur with small probability.

Case 1. Let z be such that $z \leqslant \varepsilon T$.

Knowing z, we cannot give a better bound than $\overline{M}$ on player I's expected payoff at the following stages:

- when player I deviates whereas player II is not punishing him (at most $(z+1)$ stages),
- at the last $(\overline{T}-1)$ stages, in case player II is punishing player I, but the punishment is not long enough,
- at the first $\overline{T}$ stages.

So we have at most $2\overline{T}+z$ stages where we only know that player I's payoff is at most $\overline{M}$. These stages will have a small influence on the average payoff:

$$\mathbb{E}_{\mathbb{P}^k_{\sigma,\tau}}\left(\frac{1}{T}\sum_{t=1}^{T} G^k(d_t,e_t) \,\middle|\, Z=z\right),$$

since $z \leqslant \varepsilon T$ and $\overline{T} \leqslant \varepsilon T_0 \leqslant \varepsilon T$.

We have at least $T-(2\overline{T}+z)$ other stages. At each of these stages, either player II is punishing player I, or player II is not. If player II is punishing player I at some stage $t \leqslant T-\overline{T}$, then the punishment is long enough so that the average expected payoff of player I over the stages when he is punished is at most $x^k+\varepsilon$. If player II is not punishing player I, player I plays at stage $t \geqslant \overline{T} \geqslant (\overline{M}/\varepsilon)^2$ some action d_t such that $d_t\psi^k = d^{k*}\psi^k$, whereas player II plays p^t. Hence his expected payoff is:

$$\begin{aligned}
G^k\big(d_t,p^t\big) &= \left(1-\frac{1}{\sqrt{t}}\right)G^k(d_t,p)+\frac{1}{\sqrt{t}}G^k(d_t,\hat{p})\\
&\leqslant \left(1-\frac{1}{\sqrt{t}}\right)x^k+\frac{1}{\sqrt{t}}\overline{M}\\
&= x^k+\frac{1}{\sqrt{t}}\big(\overline{M}-x^k\big)\\
&\leqslant x^k+\frac{\varepsilon}{\overline{M}}\big(\overline{M}-x^k\big)\\
&\leqslant x^k+\varepsilon.
\end{aligned}$$

We thus obtain

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^k_{\sigma,\tau}}\left(\frac{1}{T}\sum_{t=1}^{T} G^k(d_t,e_t) \,\middle|\, Z=z\right) &\leqslant \frac{1}{T}\Big(\big(2\overline{T}+z\big)\overline{M}+\big(T-2\overline{T}-z\big)\big(x^k+\varepsilon\big)\Big),\\
&\leqslant x^k+\varepsilon\big(1+3\overline{M}\big),
\end{aligned}$$

since $\overline{T} \leqslant \varepsilon T$ and $z \leqslant \varepsilon T$.

Case 2. We now consider z such that $z > \varepsilon T$. Then,

$$\mathbb{P}^k_{\sigma,\tau}(Z=z)\mathbb{E}_{\mathbb{P}^k_{\sigma,\tau}}\left(\frac{1}{T}\sum_{t=1}^{T} G^k(d_t,e_t)\,\Big|\, z\right) \leqslant \left(1-\frac{1}{|E|\sqrt{T}}\right)^z \overline{M},$$

and for all such z, we have

$$\left(1-\frac{1}{|E|\sqrt{T}}\right)^z \leqslant \mathrm{e}^{-\frac{z}{|E|\sqrt{T}}} \leqslant \mathrm{e}^{-\frac{\varepsilon\sqrt{T}}{|E|}}.$$

So:

$$\sum_{z>T\varepsilon} \mathbb{P}^k_{\sigma,\tau}(Z=z)\mathbb{E}_{\mathbb{P}^k_{\sigma,\tau}}\left(\frac{1}{T}\sum_{t=1}^{T} G^k(d_t,e_t)\,\Big|\, Z=z\right) \leqslant T\overline{M}\mathrm{e}^{-\frac{\varepsilon\sqrt{T}}{|E|}} \leqslant \varepsilon\overline{M}.$$

Summing up, we obtain:

$$\begin{aligned}\gamma_T^{I,k}(\sigma,\tau) &\leqslant \left(\sum_{z\leqslant T\varepsilon} \mathbb{P}^k_{\sigma,\tau}(Z=z)\big(x^k+\varepsilon\big(1+3\overline{M}\big)\big)\right)+\varepsilon\overline{M}\\ &\leqslant x^k+\varepsilon\big(1+4\overline{M}\big). \qquad \square\end{aligned}$$

4.4. Back to the characterization

Using Proposition 4.2 and Theorem 4.5, we now know that:

$$\begin{aligned}C=\big\{x\in\mathbb{R}^K,\ \exists p\in\Delta(E)\text{ s.t.}\\ &\text{(i) } \forall q\in\Delta(K),\ x\cdot q\geqslant\tilde{l}(q),\\ &\text{(ii) } \forall k\in K,\ x^k=G^k\big(d^{k*},p\big),\\ &\text{(iii) } \forall k\in K\ \forall d^k\in D^k \text{ s.t. } d^k\psi^k=d^{k*}\psi^k,\ G^k\big(d^{k*},p\big)\geqslant G^k\big(d^k,p\big)\big\}.\end{aligned}$$

The following proposition explicits the direct links between conditions (i), (ii) and (iii) above from the incomplete information game and the set $g(\mathcal{P})\cap JR$ of feasible payoffs that are robust to undetectable deviations and jointly rational in the original game.

Proposition 4.6. *Let x be in $\mathbb{R}^K$ and p be in $\Delta(E)$.*

(1) $x=g(p) \iff$ (ii) holds.
(2) $p\in\mathcal{P} \iff$ (iii) holds.
(3) $x\in JR \iff$ (i) holds.

Proof. First recall, from the analogy of Section 4.2, the original notations of Section 2. We notably have: $K=N$ (set of players), $E=A$ (set of joint actions), and $G^k(d^{k*},p)=g^k(p)$ (expected payoff of player k if the distribution p is played).

(1) is clear. We now prove (2).

If $d^k=(\alpha^k,\mu^k)$ is a deviation of player k in D^k, $d^k\psi^k=(\psi^k(d^k,a))_{a\in A}=((f^j(\alpha^k(a^k),a^{-k}))_{j\neq k},\mu^k(a^k,f^k(\alpha^k(a^k),a^{-k})))_{a\in A}$, hence (iii) reduces to:

$$\forall k \in K,\ \forall \alpha^k : R^k \longrightarrow A^k,\ \forall \mu^k : R^k \times U^k \longrightarrow M^k \quad \text{s.t.:}$$
$$\Big(\forall a \in A,\ \forall j \neq k,\ f^j\big(\alpha^k\big(a^k\big), a^{-k}\big) = f^j(a) \quad \text{and}$$
$$\mu^k\big(a^k, f^k\big(\alpha^k\big(a^k\big), a^{-k}\big)\big) = f^k(a)\Big),$$

then $g^k(p) \geqslant \sum_{a\in A} p(a) g^k(\alpha^k(a^k), a^{-k})$.

This is equivalent to: $\forall k \in K, \forall \alpha^k : R^k \longrightarrow A^k$ s.t.:

$$\forall a^k,\ \forall a^{-k},\ \forall j \neq k, \quad f^j\big(\alpha^k\big(a^k\big), a^{-k}\big) = f^j(a), \tag{$*$}$$

and

$$\forall a^k,\ \exists \tilde{\mu}^k : U^k \longrightarrow M^k \quad \text{s.t.} \quad \forall a^{-k}, \quad \tilde{\mu}^k\big(f^k\big(\alpha^k\big(a^k\big), a^{-k}\big)\big) = f^k\big(a^k, a^{-k}\big), \tag{$**$}$$

we have $g^k(p) \geqslant \sum_{a\in A} p(a) g^k(\alpha^k(a^k), a^{-k})$.

But ($(*)$ and $(**)$) is equivalent to $\alpha^k(a^k) \geqslant a^k$ for each a^k (see Remark 1 in Section 3.3), so (iii) is also equivalent to:
$\forall k \in N, \forall \alpha^k : R^k \longrightarrow A^k$ s.t. $\forall a^k, \alpha^k(a^k) \geqslant a^k$, we have:

$$\sum_{a^k\in A^k}\sum_{a^{-k}\in A^{-k}} p\big(a^k, a^{-k}\big) g^k\big(a^k, a^{-k}\big) \geqslant \sum_{a^k\in A^k}\sum_{a^{-k}\in A^{-k}} p\big(a^k, a^{-k}\big) g^k\big(\alpha^k\big(a^k\big), a^{-k}\big).$$

And this is simply equivalent to $p \in \mathcal{P}$.

We finally prove (3).

$$\forall q \in \Delta(N), \quad \tilde{l}(q) = \max_{\delta\in NR(q)} \min_{p\in\Delta(A)} \sum_{i\in N} q^i G^i\big(\delta^i, p\big) = \max_{\delta\in NR(q)} \min_{a\in A} \sum_{i\in N} q^i g^i_{\delta^i}(a),$$

with $NR(q) = \{\delta = (\delta^i)_{i\in N} \in \prod_{i\in N} \Delta(D^i),\ \forall i, j \in Supp\, q,\ \delta^i \psi^i = \delta^j \psi^j\}$.

We can first restrict ourselves to the projection $\widetilde{NR}(q)$ of $NR(q)$ on $\prod_{i\in Supp\, q} \Delta(D^i)$. We have $\tilde{l}(q) = \max_{\delta\in\widetilde{NR}(q)} \min_{a\in A} \sum_{i\in N} q^i g^i_{\delta^i}(a)$. Now, by definition $\delta^i \psi^i = (\sum_{d^i\in D^i} \delta^i(d^i)\psi^i(d^i, a))_{a\in A}$ and for all $a \in A$, $\sum_{d^i\in D^i} \delta^i(d^i)\psi^i(d^i, a)$ is just $\psi^i(\delta^i, a)$ as defined in Section 3.4. It is then clear that $\widetilde{NR}(q) = SD(Supp\, q)$ for all q. The set of non-revealing strategies at q in the repeated game with incomplete information is (up to a projection) the set of similar decisions of the players in the support of q. Consequently, $\tilde{l}(q) = l(q)$ for all q, and condition (i) holds if and only if x belongs to the set of jointly rational payoffs JR. □

We can now conclude the proof of Theorem 3.5. From the previous proposition, one has:

$$C = \big\{x \in \mathbb{R}^N,\ \exists p \in \Delta(A) \text{ s.t. } x \in JR,\ x = g(p),\ \text{and } p \in \mathcal{P}\big\}.$$

Hence $C = g(\mathcal{P}) \cap JR$.

5. Random signals

We finally extend our result to the case of random signals. The model of Section 2 is generalized as follows. Instead of having for each player i a deterministic observation function $f^i : A \longrightarrow U^i$, we have a single stochastic function $f : A \longrightarrow \Delta(U)$. After each stage, if $a \in A$ is the joint action played, an element $u = (u^i)_{i \in N}$ is selected according to $f(a)$, and before starting stage $t+1$ the signal u^i is announced to player i. The definition of the set C of communication equilibrium payoffs is similar as before. We first state the characterization of C, and then explain the ideas of the proof, referring to a previous work by Renault (2000) for computations.

The definition of jointly rational payoffs is almost the same. We still define the set of stage decisions of player i as:

$$D^i = \{d^i = (\alpha^i, \mu^i) \text{ with } \alpha^i : R^i \longrightarrow A^i \text{ and } \mu^i : R^i \times U^i \longrightarrow M^i\}.$$

If δ^i is in $\Delta(D^i)$ and a is in A, we denote as before by $\psi^i(\delta^i, a) \in \Delta(U)$ the probability distribution on joint messages received by the mediator if:

- a is recommended to the players,
- each player $j \neq i$ plays faithfully, whereas player i deviates according to δ^i.

The following procedure selects an element u according to $\psi^i(\delta^i, a)$. First draw $d^i = (\alpha^i, \mu^i) \in D^i$ according to δ^i, then choose an element $\tilde{u} = (\tilde{u}^k)_{k \in N}$ in U according to $f(a^{-i}, \alpha^i(a^i))$, and finally take $u = ((\tilde{u}^k)_{k \neq i}, \mu^i(a^i, \tilde{u}^i)) \in U$. We define as before the expected payoff $g^i_{\delta^i}(a)$ of player i if he uses δ^i whereas the other players play according to a, the set $SD(J)$ of similar decisions of players in J and the set JR of jointly rational payoffs. The definition of $\mathcal{P}$ extends as follows:

$$\mathcal{P} = \Big\{ p \in \Delta(A),\ \forall i \in N,\ \forall \delta^i \in \Delta(D^i) \text{ s.t. } \psi^i(\delta^i, a) = f(a)\ \forall a \in A, \\ \sum_{a \in A} p(a) g^i(a) \geqslant \sum_{a \in A} p(a) g^i_{\delta^i}(a) \Big\}.$$

With these generalizations, Theorem 3.5 perfectly extends.

Theorem 5.1. *In case of random signals, one has* $C = g(\mathcal{P}) \cap JR$.

The proof is a technical generalization of the proof of Theorem 3.5. We now explain how to proceed (all missing points and computations can be found in the proof of Proposition 5.1 in Renault, 2000).

Sketch of the proof. The main part of the proof of Theorem 3.5 can be generalized in a straightforward manner. We now have to deal with 2-player repeated games with lack of information on one side and random signals. The model of Section 4.1 is generalized to the case where in each state k, there is a signalling function $\Phi^k : D^k \times E \longrightarrow \Delta(V^k \times W)$. If the state is k, player I plays $d^k \in D^k$ and player II plays $e \in E$, an element (v^k, w) is selected according to $\Phi^k(d^k, e)$, and player I (respectively II) learns v^k (respectively w). The most relevant function is $\psi^k : D^k \times E \longrightarrow \Delta(W)$ associating to each (d^k, e) the marginal of $\Phi^k(d^k, e)$ on W. We then just have to change the definition of the signalling functions in the definition of the auxiliary game (see Section 4.2). For any $d^k \in D^k$

and $a \in A$, $\Phi^k(d^k, a)$ is defined as follows: select an element $\tilde{u} = (\tilde{u}^j)_{j \in N}$ according to $f(a^{-k}, \alpha^k(a^k))$. The signal for player I (the cheater) is then $(a^k, \tilde{u}^k) \in A^k \times U^k$, and the signal for player *II* is $(\tilde{u}^{-k}, \mu^k(a^k, \tilde{u}^k)) \in U$. Proposition 4.2 and its proof then extend word for word. The only serious problem is to generalize Theorem 4.5 (Section 4.4 then extends easily).

We keep the same definitions for $\delta^k \psi^k$, $NR(q)$ and $\tilde{l}(q)$. We still deal with repeated games with incomplete information where player *II* has payoff 0, and where player I has pure actions $d^{k*} \in D^k$ for each k, that induce the same signals for player *II*: $\forall k, k' \in K$, $\psi^k(d^{k*}, e) = \psi^{k'}(d^{k'*}, e)$ for each e in E. Theorem 4.5 extends as follows:

$$\begin{aligned}\widetilde{C} = \bigl\{x \in \mathbb{R}^K,\ \exists p \in \Delta(E) \text{ s.t.}\\ &\text{(i) } \forall q \in \Delta(K),\ x \cdot q \geqslant \tilde{l}(q),\\ &\text{(ii) } \forall k \in K,\ x^k = G^k\bigl(d^{k*}, p\bigr),\\ &\text{(iii) } \forall k \in K\ \forall \delta^k \in \Delta\bigl(D^k\bigr) \text{ s.t. } \delta^k \psi^k = d^{k*}\psi^k,\ G^k\bigl(d^{k*}, p\bigr) \geqslant G^k\bigl(\delta^k, p\bigr)\bigr\}.\end{aligned}$$

The unique difference is in (iii), where one now has to consider mixed deviations in $\Delta(D^k)$ instead of only pure deviations in D^k. The reason is that with random signals, it may be possible to find a mixed deviation δ^k in $\Delta(D^k)$ inducing the same signals as d^{k*} for player *II*, and which is not a convex combination of pure deviations in D^k inducing the same signals as d^{k*}. The proof of this characterization for $\widetilde{C}$ is more difficult. There is no problem in using Kohlberg's result, which has been generalized to the random signalling case by Mertens et al. (1994, part B, Chapter V). Part (1) of the proof of Theorem 4.5 extends word for word, but the problem concerns part (2). The difficulty is that the signals player *II* is supposed to receive are now random, so knowing his own action player *II* does not exactly know which signal he should receive. Hence statistical tests will be required. Namely, instead of playing a mixed action once, the uninformed player will have to play the same mixed action i.i.d. for a large number of stages and then decide from the statistics of signals, whether player I has deviated or not.

Fix x in $\mathbb{R}^K$ and p in $\Delta(E)$ satisfying (i), (ii) and (iii). We have to construct τ such that (d^{**}, τ) is an equilibrium of Γ_{inc} with payoff $(x, 0)$.

In order to define τ, we can adapt the constructions of Lehrer for repeated games with complete information and deterministic signals (see Lehrer, 1990, 1992b), or more directly very slightly adapt the construction of equilibrium strategies of Renault (2000) for 2-player repeated games with incomplete information and random signalling: the unique difference is that in Renault's paper, signals are assumed to be state independent whereas it is not the case here. However, the generalization to the setup we are dealing with now is straightforward, and almost everything can be used word for word. We now construct τ. Back to communication equilibrium payoffs of repeated games with imperfect monitoring, this construction explicits, for every payoff x in $g(\mathcal{P}) \cap JR$, a canonical communication device yielding x as a CEP (in the sense of Definition 2.3).

Definition of the strategy τ. The set of stages $\{1, 2, \ldots\}$ is divided into consecutive blocks $B^1, B^2, \ldots, B^m, \ldots$ such that for each m, $|B^m| = m^{10}$. The play will consist of a main path and of punishments' phases, starting from the main path.

At some block B^m in the main path, player *II* plays independently at each stage the mixed action $(1-1/m)\,p + (1/m)\hat{p}$, with $\hat{p}$ being the uniform distribution on player *II*'s actions.

At the end of such a block B^m, player *II* will compare the frequency of signals he observed and the frequency of signals he expected to observe (i.e. he should observe if player I is not deviating, in the sense that he is playing according to d^{**}). Put, for any $m > 0$, action e in E and signal w in W:

$$TH(w,e) = \psi^k\big(d^{k*}, e\big)(w) \quad \text{and} \quad OB^m(w,e) = \frac{|\{t \in B^m,\ e_t = e, w_t = w\}|}{|\{t \in B^m,\ e_t = e\}|},$$

e_t and w_t being respectively the actions played and signals received by player *II* at stage t. If $\{t \in B^m, e_t = e\} = \emptyset$, just define $OB^m(w,e) = TH(w,e)$.

$TH(w,e)$ is the probability that player *II* observes the signal w at some stage when he plays e, if player I plays d^{k*} for some k. Hence it is the theoretical frequency of observing w while playing e that he should observe if player I is not deviating. Notice that it does not depend on k in K. $OB(w,e)$ is the observed, or empirical, frequency of signals w among the stages where player *II* has played e. The idea is that if player I uses d^{**}, $OB^m(w,e)$ and $TH(w,e)$ will be, by Tchebychev's inequality, very close with great probability for each w and e. If player I plays something else, for example in order to increase his payoff, condition (iii) will imply that for some e and w, $OB^m(w,e)$ and $TH(w,e)$ will be more different than they should be, and player *II* will punish player I via a punishing strategy given by condition (i). More precisely, after the play of some block B^m in the main path there are two cases:

If for each w in W and e in E, $|TH(w,e) - OB^m(w,e)| \leqslant 1/m$, player *II* will consider that player I did play according to d^{**} at block B^m. The play stays in the main path and proceeds to block B^{m+1}.

Otherwise, player *II* will consider that player I has deviated from d^{**}. The play will immediately go out of the main path and a punishment phase will start. This phase will last from block B^{m+1} to block B^{m^2}. Then, whatever happens during the punishment phase, the play will come back to the main path at block B^{m^2+1}.

To complete the definition of τ, it remains to define what is played during a punishment phase. By assumption, $x \cdot q \geqslant \tilde{l}(q)\ \forall q \in \Delta(K)$, thus the generalization of Kohlberg's result by Mertens et al. (1994, part B, Chapter V, Section 3.d) gives, as in Section 4.3, the existence of some strategy $\bar{\tau}$ of player *II* such that: $\forall \varepsilon > 0, \exists T_0\ \forall T \geqslant T_0\ \forall \sigma \in \mathcal{S}$ $\forall k \in K,\ \gamma_T^{I,k}(\sigma, \bar{\tau}) \leqslant x^k + \varepsilon$. Define now, for each m:

$$\varepsilon_m = \min\Big\{\varepsilon \geqslant 0,\ \exists \bar{\tau}_m \in \mathcal{T} \text{ s.t. } \forall \sigma \in \mathcal{S},\ \forall k \in K,$$
$$\forall T \in \big\{(m+1)^{10}, (m+1)^{10}+1, \ldots, m^{20}\big\},\ \gamma_T^{I,k}(\sigma, \bar{\tau}_m) \leqslant x^k + \varepsilon\Big\}.$$

We have $\varepsilon_m \longrightarrow_{m\to\infty} 0$. Consider the case where the play is in the main path at some block B^m and a punishment phase will start at block B^{m+1}. Let $\bar{\tau}_m$ be a strategy of player *II* that pushes player I's payoff down to $x + \varepsilon_m$ in any game of length between $(m+1)^{10}$ and m^{20}. To punish player I at the blocks $m+1, \ldots, m^2$, player *II* will play at any block $m' \in \{m+1, \ldots, m^2\}$ according to the $|B^{m'}|$ first moves of $\bar{\tau}_m$. This completes the definition of τ.

We finally briefly explain why (d^{**}, τ) is an equilibrium of the incomplete information game.

First assume that (d^{**}, τ) is played. Using Tchebychev's inequality and Borel Cantelli Lemma, one can first show that almost surely the number of punishment phases is finite. It is then not difficult to show that for each state k, $(1/T)\sum_{t=1}^{T} G^k(d_t, e_t) \longrightarrow_{T\to\infty} x^k$, $\mathbb{P}_{d^{**},\tau}$ a.s. and to conclude that $\gamma_T^{I,k}(d^{**}, \tau) \longrightarrow_{T\to\infty} x^k$ by the bounded convergence theorem. See Renault (2000), proof of condition (2) for computations.

Finally assume that player I plays some strategy σ whereas player II plays according to τ. Define for each block number m, $\mathcal{B}_m$ as the event {the play is in the main path at block B^m}. Fix some state k, and define X_m as the random variable: $(1/|B^m|)\sum_{t\in B^m} G^k(d_t, e_t) - x^k$. Assume that the state is k, and fix $\varepsilon > 0$.

The key point is the following, which corresponds to Lemma A of Renault (2000). One can show that at each block B^m, m large enough, while player II is in the main path, player I only has a small probability to obtain a good payoff without being punished afterwards. More precisely, we have for m large enough: $\mathbb{P}^k_{\sigma,\tau}(\mathcal{B}_{m+1} \cap \{X_m > \varepsilon\} \mid \mathcal{B}_m) \leqslant \varepsilon$. The main argument in the proof of this property is a lemma by Lehrer (1992b, Appendix 1), which extends Tchebychev's inequality to non independent random variables such as the sequences of actions played by the players at some block. As a consequence, one can proceed "as if" player I was playing the same mixed action i.i.d. at each stage of some large block in the main path. Computations, using condition (iii), conclude the proof of this property.

Lastly, from the efficiency of the punishing strategies and the cardinalities of the blocks, one can prove that $\forall\varepsilon' > 0\ \exists T_0\ \forall T \geqslant T_0\ \gamma_T^{I,k}(\sigma, \tau) \leqslant x^k + \varepsilon'$. For a precise proof of this, see Lemma B of Renault (2000).

References

Aumann, R.J., Maschler, M., 1995. Repeated Games with Incomplete Information. MIT Press, Cambridge, MA. With the collaboration of R. Stearns.

Blackwell, D., 1956. An analog of the Minmax Theorem for vector payoffs. Pacific J. Math. 65, 1–8.

Forges, F., 1985. An approach to communication equilibria. Econometrica 54, 1375–1385.

Kohlberg, E., 1975. Optimal strategies in repeated games with incomplete information. Int. J. Game Theory 4, 7–24.

Lehrer, E., 1989. Lower equilibrium payoffs in two-player repeated games with non-observable actions. Int. J. Game Theory 18, 57–89.

Lehrer, E., 1990. Nash equilibria of n-player repeated games with semi-standard information. Int. J. Game Theory 13, 191–217.

Lehrer, E., 1992a. Correlated equilibria in two-player repeated games with non-observable actions. Math. Operations Res. 17, 175–199.

Lehrer, E., 1992b. On the equilibrium payoffs set of two-player repeated games with imperfect monitoring. Int. J. Game Theory 20, 211–226.

Mertens, J.F., Sorin, S., Zamir, S., 1994. Repeated games. Parts A and B. CORE DP 9420 and 9421.

Myerson, R.B., 1982. Optimal coordination mechanisms in generalized principal agent problems. J. Math. Econ. 10, 67–81.

Raghavan, T.E.S., 1994. Zero-sum two-person games. Chapter 20 in: Aumann, R.J., Hart, S. (Eds.), Handbook of Game Theory, vol. II. Elsevier.

Renault, J., 2000. On two-player repeated games with lack of information on one side and state independent signalling. Math. Operations Res. 25, 552–572.

Renault, J., Tomala, T., 1998. Repeated proximity games. Int. J. Game Theory 27, 539–559.
Tomala, T., 1999. Nash equilibria of repeated games with observable payoff vector. Games Econ. Behav. 28, 310–324.
Sorin, S., 1992. Repeated games with complete information. In: Aumann, R.J., Hart, S. (Eds.), Handbook of Game Theory, vol. I. Elsevier, pp. 71–107.