The MaxMin Value of Stochastic Games with Imperfect Monitoring

Dinah Rosenberg*, Eilon Solan† and Nicolas Vieille‡§

April 22, 2003

Abstract

We study finite zero-sum stochastic games in which players do not observe the actions of their opponent. Rather, in each stage, each player observes a stochastic signal that may depend on the current state and on the pair of actions chosen by the players. We assume that each player observes the state and his/her own action. We prove that the uniform max-min value always exists. Moreover, the uniform max-min value is independent of the information structure of player 2. Symmetric results hold for the uniform min-max value.

1 Introduction

The classical literature on repeated games and stochastic games considers models with perfect monitoring in which past play is observed by the players. The strategies used by the players at equilibrium in such games are usually history dependent, and use the observation of the past sequence of moves to play at any given stage.

In the last two decades, models with imperfect monitoring were explored, and several applications of these models were studied (see, e.g., Radner (1981), Rubinstein and Yaari (1983)). Lehrer (1989, 1990, 1992a, 1992b) has characterized the set of equilibrium payoffs for various notions of undiscounted equilibria in infinitely repeated games with imperfect monitoring. Plainly, zero-sum repeated games have a value, and an optimal strategy for a player is to repeatedly play his optimal strategy in the one-shot game, whatever be the signaling structure. Unlike the situation in repeated games, the value of zero-sum stochastic games might be modified by the introduction of imperfect monitoring.

In the present paper we are interested in two-player zero-sum stochastic games with imperfect monitoring. These games are played as follows. At every stage, the game is in one of finitely many states. Each player chooses an action, independently of his opponent. The current state, together with the pair of actions, determine a daily payoff that player 2 pays player 1, a probability

[*] Laboratoire d'Analyse Géométrie et Applications, Institut Galilée, Université Paris Nord, avenue Jean-Baptiste Clément, 93430 Villetaneuse, France. e-mail: dinah@zeus.math.univ-paris13.fr

[†] MEDS Department, Kellogg School of Management, Northwestern University, *and* School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel. e-mail: eilons@post.tau.ac.il, e-solan@kellogg.northwestern.edu

[‡] Ecole Polytechnique and Département Finance et Economie, HEC, 1, rue de la Libération, 78 351 Jouy-en-Josas, France. e-mail: vieille@hec.fr

[§] We deeply thank E. Lehrer, J.-F. Mertens, A. Neyman and S. Sorin. The discussions we had, and the suggestions they provided, were extremely useful. We acknowledge the financial support of the Arc-en-Ciel/Keshet program for 2001/2002. The research of the second author was supported by the Israel Science Foundation (grant No. 69/01-1).

distribution according to which a new state is chosen, and a probability distribution over pairs of signals, one for each player. Each player is then informed of his private signal, and of the new state. However, no player is informed of his opponent's signal and of the daily payoff.

For every discount factor, the discounted game with perfect monitoring has a value, and each player has an optimal stationary strategy, namely, an optimal strategy that depends only on the current state (see Shapley, 1953). Similarly, for every positive integer n, the n-stage game has a value, and each player has an optimal strategy that depends only on the current state and on the number of remaining stages. Consequently, in both cases, the value is independent of the signaling structure (provided each player always observes the current state), and the optimal strategies remain optimal in the λ-discounted game or the n-stage game with any signaling structure. However, optimal strategies usually depend on the discount factor or on the length of the game.

Here we study the *uniform* max-min value of stochastic games. The uniform max-min value v exists as soon as (i) for every $\varepsilon > 0$ player 1 has a single strategy that ensures that the expected average payoff in *every* sufficiently long game is at least $v - \varepsilon$, and (ii) for every $\varepsilon > 0$ and every strategy of player 1, player 2 has a reply such that the expected average payoff in *every* sufficiently long game is at most $v + \varepsilon$. The uniform min-max value is defined analogously, by exchanging the roles of the two players.

Mertens and Neyman (1981) proved that zero-sum stochastic games with perfect monitoring always have a uniform value – *i.e.*, both the uniform max-min value and the uniform min-max value exist, and they coincide. The ε-optimal strategies they constructed indeed rely on the observation of the sequence of past moves of the opponent.

Coulomb (1992, 1999, 2001) was the first to study stochastic games with imperfect monitoring. He studied the class of absorbing games, and proved that the uniform max-min value (and similarly the uniform min-max value) exists. In addition, he provided a formula for both values. One of Coulomb's main findings is that the uniform max-min value does not depend on the signaling structure of player 2. Similarly, the uniform min-max value does not depend on the signaling structure of player 1. In general, the uniform max-min and the uniform min-max values do not coincide, hence stochastic games with imperfect monitoring need not have a uniform value.

In the present paper we prove that all finite stochastic games have a uniform max-min value and a uniform min-max value. As in the case of absorbing games, the uniform max-min value is independent of the information structure of player 2, and the uniform min-max value is independent of the information structure of player 1. We also prove that these values are limits of max-min and min-max values of certain auxiliary discounted (non-standard) games.

The approach that we take is quite different from that of Coulomb (1992, 1999, 2001). We first define an equivalence relation over mixed actions of player 2, that has similarities with the one used in Lehrer's and Coulomb's works. Basically two actions of player 2 are said to be equivalent with respect to a mixed move of player 1 if they induce the same distribution of signals to player 1. However the definition takes into account the fact that we use discounted games, hence events that occur rarely (relative to the discount factor) do not affect the payoff. Using this equivalence relation we define a new daily payoff function. We then define an auxiliary discounted max-min value as a fixed point of a functional equation that is based on the auxiliary daily payoff function. Finally, we prove that the uniform max-min value is the limit of these auxiliary discounted max-min values.

To prove the last claim we use the method developed by Mertens and Neyman (1981) for stochastic games with perfect monitoring. The method of studying asymptotic properties of auxiliary discounted games by defining a new payoff function already appears in Solan (1999) and in Solan and Vohra (2002), in the study of equilibria in n-player absorbing games.

Independently of our work, Coulomb (2003) proved the same result, using similar tools.

The paper is organized as follows. Section 2 contains the model and the statement of the main results. In Section 3 we introduce a number of tools, define the auxiliary discounted games, and study some of their basic properties. Section 4 contains a reminder of the analysis of Mertens and Neyman. The last two sections are devoted to the two parts of the proof.

2 The model

For every finite set K, $\Delta(K)$ is the set of probability distributions over K. We identify each element $k \in K$ with the element of $\Delta(K)$ that assigns probability one to k.

A *two-person zero-sum stochastic game with imperfect monitoring* is described by: (i) a set S of states, (ii) action sets A and B for the two players, (iii) a daily reward function $r : S \times A \times B \to \mathbf{R}$, (iv) signal sets M^1 and M^2 and (v) a transition function $\psi : S \times A \times B \to \Delta(M^1 \times M^2 \times S)$. All through the paper, the sets S, A, B, M^1 and M^2 are assumed to be finite.

The game is played in stages. The initial state s_1 is known to both players. At each stage $n \in \mathbf{N}$, (a) the players independently choose actions a_n and b_n; (b) player 2 pays player 1 the amount $r(s_n, a_n, b_n)$; (c) a triple (m^1_n, m^2_n, s_{n+1}) is drawn according to $\psi(s_n, a_n, b_n)$; (d) players 1 and 2 are told respectively m^1_n and m^2_n, but they are *not* informed of a_n, b_n, or $r(s_n, a_n, b_n)$; and (e) the game proceeds to stage $n + 1$.

We denote by ψ^1 (resp. ψ^2) the projection of ψ on M^1 (resp. M^2). These functions represent the signal that each of the players receives. The multi-linear extensions of r and ψ to $S \times \Delta(A) \times \Delta(B)$ are still denoted by r and ψ respectively.

We assume throughout that each player always knows the current state, and the action he is playing. In terms of ψ, this amounts to assuming the following: if both probabilities $\psi^1(s, a, b)[m^1, m^2, t]$ and $\psi^1(s', a', b')[m^1, m'^2, t']$ are positive, then $(s, a, t) = (s', a', t')$. A similar property holds for player 2. We also assume perfect recall, so each player remembers the sequence of signals he has received so far.

We denote by $H_n = S \times (A \times B \times M^1 \times M^2 \times S)^{n-1}$ the set of histories up to stage n,[1] and by $H^1_n = S \times (M^1)^{n-1}$ and $H^2_n = S \times (M^2)^{n-1}$ the set of private histories of the two players respectively. We also let $H_\infty = (S \times A \times B \times M^1 \times M^2)^{\mathbf{N}}$ denote the set of infinite plays. For $i = 1, 2$, $\mathcal{H}^i_n$ denotes the cylinder algebra over H_∞ induced by H^i_n, $\mathcal{H}^i_\infty = \sigma(\mathcal{H}^i_n, n \geq 1)$ the σ-algebra of events that are measurable for player i, and $\mathcal{H}_\infty = \sigma(\mathcal{H}^1_\infty, \mathcal{H}^2_\infty)$ the σ-algebra generated by all the cylinder algebras.

A (behavioral) strategy of player 1 (resp. player 2) is a sequence $\sigma = (\sigma_n)_{n\geq 1}$ (resp. $\tau = (\tau_n)_{n\geq 1}$) of functions $\sigma_n : H^1_n \to \Delta(A)$ (resp. $\tau_n : H^2_n \to \Delta(B)$). A stationary strategy depends only on the current stage : hence, a stationary strategy of player 1 is described by a vector $(x^s)_{s\in S}$ in $(\Delta(A))^S$, with the interpretation that x^s is the mixed move used whenever the current state is $s \in S$. Stationary strategies of player 2 are denoted by $(y^s)_{s\in S} \in (\Delta(B))^S$.

We denote by $\mathbf{P}_{s,\sigma,\tau}$ the probability distribution induced over $(H_\infty, \mathcal{H}_\infty)$ by a pair (σ, τ) of strategies and an initial state $s \in S$, and by $\mathbf{E}_{s,\sigma,\tau}$ the corresponding expectation operator. The

[1] Since the signal of each player contains the current state, the next state, and his action, some information in this representation is redundant.

expected average payoff up to stage n is

$$\gamma_n(s,\sigma,\tau) = \mathbf{E}_{s,\sigma,\tau}\left[\frac{1}{n}\sum_{k=1}^{n} r(s_k,a_k,b_k)\right].$$

Definition 1 *$v \in \mathbf{R}^S$ is the (uniform)* max-min value *of the game if:*

- *Player 1 can* guarantee *v: for every $\varepsilon > 0$, there exists a strategy σ of player 1 and $N \in \mathbf{N}$, such that:*
$$\forall s \in S, \forall \tau, \forall n \geq N,\ \gamma_n(s,\sigma,\tau) \geq v(s) - \varepsilon.$$
- *Player 2 can* defend *v: for every $\varepsilon > 0$ and every strategy σ of player 1 there exists a strategy τ of player 2 and $N \in \mathbf{N}$, such that:*
$$\forall s \in S, \forall n \geq N, \gamma_n(s,\sigma,\tau) \leq v(s) + \varepsilon.$$

The definition of the (uniform) min-max value is obtained by exchanging the roles of the two players.

Our main result is the following.

Theorem 2 *Every stochastic game has a max-min value and a min-max value. The max-min value (resp. the min-max value) depends on ψ only through ψ^1 (resp. only through ψ^2).*

Note that if player 1 cannot guarantee a quantity w it does not follow that player 2 can defend it. Therefore the first part of the theorem is not a tautology. Recall that this result assumes that the game has perfect recall, and that each player always knows the current state. The situation in which players are not fully informed of the current state raises additional difficulties, see Rosenberg et al. (2002) for the analysis of the one-player case.

Coulomb (1999, 2001) proved the corresponding statement for the class of absorbing games.

We assume w.l.o.g. that payoffs are non-negative and bounded by 1. We focus on the existence of the max-min value. The existence of the min-max value follows by the same argument by exchanging the roles of players 1 and 2.

3 The max-min value

3.1 Indistinguishable moves

We start by defining an equivalence relation between mixed actions of player 2. This equivalence relation will be used to provide a semi-explicit formula for the max-min value. In essence, two mixed actions y and z of player 2 are equivalent for a mixed action x of player 1 at state s if the probability that player 1 cannot distinguish y from z is high. Variants of this relation have played a central role in earlier analysis of games with imperfect monitoring, such as in the work of Aumann and Maschler (1995), Lehrer (1989, 1990, 1992a, 1992b) and Coulomb (1999, 2001).

Given $\varepsilon, \lambda > 0$,[2] $s \in S$ and $x \in \Delta(A)$, we define a binary relation $\sim_{\lambda,\varepsilon,s,x}$ over $\Delta(B)$ as follows:

$$y \sim_{\lambda,\varepsilon,s,x} z \text{ if and only if } \psi^1(s,a,y) = \psi^1(s,a,z) \text{ for every } a \text{ such that } x[a] \geq \lambda/\varepsilon.$$

[2] λ always stands for a discount factor. Here and in the sequel we omit the condition $\lambda \leq 1$.

Thus, y and z are equivalent for x at s if every action of player 1 that can be used to distinguish between y and z is played under x with low probability (low being defined with respect to λ). Plainly, the relation $\sim_{\lambda,\varepsilon,s,x}$ is an equivalence relation.

Remark: A simple alternative candidate for the definition of the relation is $y \sim_{\lambda,\varepsilon,s,x} z$ if and only if $\|\psi^1(s,x,y) - \psi^1(s,x,z)\| \leq \lambda/\varepsilon$. However, this would not define a transitive relation.

3.2 An auxiliary daily payoff function

We define a function $\widetilde{r}$ that is to be thought of as the worst payoff consistent with a given distribution of signals to player 1. Given $\varepsilon, \lambda > 0$, $s \in S$ and $(x, y) \in \Delta(A) \times \Delta(B)$, we set

$$\widetilde{r}^{\varepsilon}_{\lambda}(s,x,y) = \min_{z \sim_{\lambda,\varepsilon,s,x} y} r(s,x,z). \tag{1}$$

Since the set $\{z \in \Delta(B) : z \sim_{\lambda,\varepsilon,s,x} y\}$ is compact, the minimum in the right-hand side of (1) is reached. Note that $\widetilde{r}^{\varepsilon}_{\lambda}(s,x,y) = \widetilde{r}^{\varepsilon}_{\lambda}(s,x,z)$ whenever $z \sim_{\lambda,\varepsilon,s,x} y$, and

$$\widetilde{r}^{\varepsilon}_{\lambda}(s,x,y) \leq r(s,x,y). \tag{2}$$

The continuity property of $\widetilde{r}$ that we need in the sequel is summarized by the following lemma.

Lemma 3 *For every $\delta > 0$, there is $\eta > 0$ such that for every $s \in S$, every $x \in \Delta(A)$, every $\lambda, \varepsilon > 0$ and every $y, z \in \Delta(B)$, the following is satisfied: if $\|\psi^1(s,a,y) - \psi^1(s,a,z)\| < \eta$ for every $a \in A$ that satisfies $x[a] \geq \lambda/\varepsilon$, then $|\widetilde{r}^{\varepsilon}_{\lambda}(s,x,y) - \widetilde{r}^{\varepsilon}_{\lambda}(s,x,z)| < \delta$.*

The proof of Lemma 3 relies on the next result.

Lemma 4 *For every $\delta > 0$ there is $\eta > 0$ such that for every $s \in S$, every $x \in \Delta(A)$, every $y, z, z' \in \Delta(B)$, and every $\varepsilon, \lambda > 0$, the following is satisfied. If (i) $\|\psi^1(s,a,y) - \psi^1(s,a,z)\| < \eta$ for every $a \in A$ that satisfies $x[a] \geq \lambda/\varepsilon$, and (ii) $z' \sim_{\lambda,\varepsilon,s,x} z$, then there exists $y' \in \Delta(B)$ such that (a) $y' \sim_{\lambda,\varepsilon,s,x} y$, and (b) $\|y' - z'\| < \delta$.*

Observe that Lemma 4 implies Lemma 3. Indeed, let $\delta > 0$ be given, and let $\eta > 0$ be the one given by Lemma 4 w.r.t. δ. Suppose $y, z \in \Delta(B)$ satisfy $\|\psi^1(s,a,y) - \psi^1(s,a,z)\| < \eta$ for every $a \in A$ such that $x[a] \geq \lambda/\varepsilon$. Let $z' \sim_{\lambda,\varepsilon,s,x} z$ satisfy $\widetilde{r}^{\varepsilon}_{\lambda}(s,x,z) = r(s,x,z')$, and let $y' \in \Delta(B)$ satisfy the conclusion of Lemma 4 w.r.t. z'. Then by Lemma 4(b) and (2)

$$\widetilde{r}^{\varepsilon}_{\lambda}(s,x,z) = r(s,x,z') > r(s,x,y') - \delta \geq \widetilde{r}^{\varepsilon}_{\lambda}(s,x,y') - \delta = \widetilde{r}^{\varepsilon}_{\lambda}(s,x,y) - \delta.$$

Exchanging the roles of y and z, one obtains $|\widetilde{r}^{\varepsilon}_{\lambda}(s,x,z) - \widetilde{r}^{\varepsilon}_{\lambda}(s,x,y)| < \delta$, and Lemma 3 follows.

Proof. Since S is finite, we may assume that s is given.

Assume to the contrary that the lemma does not hold. Then there exists $\delta > 0$ such that for every $n \in \mathbf{N}$ there are $x_n \in \Delta(A)$, $y_n, z_n, z'_n \in \Delta(B)$, and $\lambda_n, \varepsilon_n > 0$ such that (i) $\|\psi^1(s,a,y_n) - \psi^1(s,a,z_n)\| < 1/n$ for every $a \in A$ that satisfies $x_n[a] \geq \lambda_n/\varepsilon_n$, (ii) $z'_n \sim_{\lambda_n,\varepsilon_n,s,x_n} z_n$, and for every $y'_n \in \Delta(B)$ that satisfy (a) $y'_n \sim_{\lambda_n,\varepsilon_n,s,x_n} y_n$, we have (b) $\|y'_n - z'_n\| \geq \delta$.

To derive a contradiction, we define a sequence (y'_n) such that for every n, $y'_n \sim_{\lambda_n,\varepsilon_n,s,x_n} y_n$ and $\lim_{n\to\infty} \|y'_n - z'_n\| = 0$.

By taking a subsequence we assume w.l.o.g. that (A) the limits $y = \lim_{n\to\infty} y_n$, $z = \lim_{n\to\infty} z_n$ and $z' = \lim_{n\to\infty} z'_n$ exist, and (B) the set $\{a \in A \mid x_n[a] \geq \lambda_n/\varepsilon_n\}$ is independent of n.

Claim 1: For every $n \in \mathbf{N}$, $z \sim_{\lambda_n,\varepsilon_n,s,x_n} z'$.
Let $a \in A$ satisfy $x_n[a] \geq \lambda_n/\varepsilon_n$ for every $n \in \mathbf{N}$. Then $\psi^1(s,a,z_n) = \psi^1(s,a,z'_n)$. By the continuity of ψ^1, $\psi^1(s,a,z) = \psi^1(s,a,z')$, as desired.

Claim 2: For every $n \in \mathbf{N}$, $y \sim_{\lambda_n,\varepsilon_n,s,x_n} z$.
Let $a \in A$ satisfy $x_n[a] \geq \lambda_n/\varepsilon_n$ for every $n \in \mathbf{N}$. Since $\|\psi^1(s,a,y_n) - \psi^1(s,a,z_n)\| < 1/n$ for every $n \in \mathbf{N}$, and by the continuity of ψ^1, we derive $\psi^1(s,a,y) = \psi^1(s,a,z)$.

Since $y = \lim_{n\to\infty} y_n$, there exists a sequence $(\alpha_n, e_n)_{n\in\mathbf{N}}$ such that $\alpha_n \in [0,1]$, $\lim_{n\to\infty} \alpha_n = 1$, $e_n \in \Delta(B)$, and $y_n = \alpha_n y + (1-\alpha_n)e_n$. Define $y'_n = \alpha_n z' + (1-\alpha_n)e_n$.

Claim 3: For every $n \in \mathbf{N}$, $y_n \sim_{\lambda_n,\varepsilon_n,s,x_n} y'_n$.
Let $a \in A$ satisfy $x_n[a] \geq \lambda_n/\varepsilon_n$ for every n. By Claims 1 and 2, $\psi^1(s,a,y) = \psi^1(s,a,z')$. By the linearity of ψ, $\psi^1(s,a,y'_n) = \alpha_n\psi^1(s,a,z') + (1-\alpha_n)\psi^1(s,a,e_n) = \alpha_n\psi^1(s,a,y) + (1-\alpha_n)\psi^1(s,a,e_n) = \psi^1(s,a,y_n)$, as desired.

The desired contradiction follows from Claim 3 and since $\lim_{n\to\infty} y'_n = z' = \lim_{n\to\infty} z'_n$. ■

Corollary 5 *For every $\lambda, \varepsilon > 0$, and every $s \in S$, the function $\widetilde{r}^{\varepsilon}_{\lambda}(s,\cdot,\cdot)$ is continuous w.r.t. y, and is (jointly) upper semicontinuous.*

Proof. That $\widetilde{r}^{\varepsilon}_{\lambda}(s,\cdot,\cdot)$ is continuous w.r.t. y is an immediate consequence of Lemma 4. Let $s \in S$ be given, and let $(x_n, y_n)_{n\in\mathbf{N}}$ be a convergent sequence in $\Delta(A) \times \Delta(B)$, with limit (x,y). W.l.o.g., assume that the set $\{a \in A : x_n[a] \geq \lambda/\varepsilon\}$ is independent of $n \in \mathbf{N}$, and that $\lim_{n\to\infty} \widetilde{r}^{\varepsilon}_{\lambda}(s,x_n,y_n)$ exists. Let $z \in \Delta(B)$ be such that $z \sim_{\lambda,\varepsilon,s,x} y$. By Lemma 4, for each $k \in \mathbf{N}$, there exists n_k and $z_{n_k} \in \Delta(B)$ such that $z_{n_k} \sim_{\lambda,\varepsilon,s,x} y_{n_k}$ and $\|z_{n_k} - z\| < 1/k$. Since $z_{n_k} \sim_{\lambda,\varepsilon,s,x} y_{n_k}$, one has $z_{n_k} \sim_{\lambda,\varepsilon,s,x_{n_k}} y_{n_k}$. Indeed, if $x_{n_k}[a] \geq \lambda/\varepsilon$ for every k then $x[a] \geq \lambda/\varepsilon$, so that $\psi^1(s,a,z_{n_k}) = \psi^1(s,a,y_{n_k})$.

In particular, by (2), $r(s,x_{n_k},z_{n_k}) \geq \widetilde{r}^{\varepsilon}_{\lambda}(s,x_{n_k},z_{n_k}) = \widetilde{r}^{\varepsilon}_{\lambda}(s,x_{n_k},y_{n_k})$. Letting k go to infinity, one obtains $\lim_{n\to\infty} \widetilde{r}^{\varepsilon}_{\lambda}(s,x_n,y_n) \leq r(s,x,z)$. Since z is arbitrary, this also implies $\lim_{n\to\infty} \widetilde{r}^{\varepsilon}_{\lambda}(s,x_n,y_n) \leq \widetilde{r}^{\varepsilon}_{\lambda}(s,x,y)$. ■

3.3 A functional equation

It is convenient to denote by q the marginal of ψ over S: $q(s'|s,a,b) = \psi(s,a,b)[M^1 \times M^2 \times \{s'\}]$ is the probability of moving from state s to state s' when the actions are a and b. This is the transition function of the game, when one ignores the signals. The multi-linear extension of q to $S \times \Delta(A) \times \Delta(B)$ is still denoted by q.

Given $\lambda, \varepsilon > 0$, we define the operator $T_{\lambda,\varepsilon} : \mathbf{R}^S \to \mathbf{R}^S$ as follows: for every $w : S \to \mathbf{R}$, we set

$$T_{\lambda,\varepsilon} w(s) := \max_{x\in\Delta(A)} \min_{y\in\Delta(B)} \left\{\lambda \widetilde{r}^{\varepsilon}_{\lambda}(s,x,y) + (1-\lambda)\mathbf{E}_{q(\cdot|s,x,y)}[w(\cdot)]\right\},$$

where $\mathbf{E}_{q(\cdot|s,x,y)}$ is the expectation w.r.t. $q(\cdot|s,x,y)$.

Since $\widetilde{r}^{\varepsilon}_{\lambda}$ is continuous w.r.t. y, the minimum in the definition of $T_{\lambda,\varepsilon}$ is attained for each $x \in \Delta(A)$. Since $\widetilde{r}^{\varepsilon}_{\lambda}$ is jointly upper semi-continuous, the maximum is also attained.

Lemma 6 *For each $\lambda, \varepsilon > 0$, the operator $T_{\lambda,\varepsilon}$ has a unique fixed point.*

Proof. Plainly, $T_{\lambda,\varepsilon}$ is non-decreasing. Moreover, $T_{\lambda,\varepsilon}(w + c\mathbf{1}) = T_{\lambda,\varepsilon}w + (1-\lambda)c\mathbf{1}$, for each $c \in \mathbf{R}$ and $w : S \to \mathbf{R}$, where $\mathbf{1} : S \to \mathbf{R}$ is the function defined by $\mathbf{1}(s) = 1$ for every $s \in S$. By Blackwell's criterion, $T_{\lambda,\varepsilon}$ is strictly contracting, hence has a unique fixed point. ■

Given $\lambda, \varepsilon > 0$, we let v_λ^ε denote the unique fixed point of $T_{\lambda,\varepsilon}$. Our characterization of the max-min value is the following.

Theorem 7 *The limit $v = \lim_{\varepsilon\to 0} \lim_{\lambda\to 0} v_\lambda^\varepsilon$ exists, and is the max-min value of the game.*

Observe that v_λ^ε does not depend on ψ^2, the structure of signals to player 2, hence neither does v.

3.4 Algebraic properties

We collect in Proposition 9 below the semi-algebraic properties that are useful for our purposes.

Lemma 8 *For every state $s \in S$, the function $\phi_s : (\varepsilon, \lambda, x, y) \mapsto \widetilde{r}_\lambda^\varepsilon(s, x, y)$ is semi-algebraic.*

Proof. Fix $s \in S$. The set

$$E = \left\{(\varepsilon, \lambda, x, y, y', r) \in (0,1)^2 \times \Delta(A) \times (\Delta(B))^2 \times \mathbf{R} : y \sim_{\lambda,\varepsilon,s,x} y', r = r(s,x,y')\right\}$$

is defined by finitely many polynomial inequalities. In particular, it is semi-algebraic. Therefore the graph of ϕ_s, which is equal to

$$\left\{(\varepsilon, \lambda, x, y, r) \in (0,1)^2 \times \Delta(A) \times \Delta(B) \times \mathbf{R} \colon r = \min\{r' \in \mathbf{R}, \exists y' \in \Delta(B) \text{ s.t. } (\varepsilon, \lambda, x, y, y', r') \in E\}\right\}$$

is semi-algebraic as well. ■

Using Lemma 6 one can now deduce the following.

Proposition 9 *For every state $s \in S$, (i) the function $(\lambda, \varepsilon) \mapsto v_\lambda^\varepsilon(s)$ is semi-algebraic, and (ii) the set*

$$\left\{(\varepsilon, \lambda, x) \in (0,1)^2 \times \Delta(A) \colon \min_{y\in\Delta(B)} \{\lambda \widetilde{r}_\lambda^\varepsilon(s,x,y) + (1-\lambda)\mathbf{E}\left[v_\lambda^\varepsilon | s, x, y\right]\} = v_\lambda^\varepsilon(s)\right\}$$

is semi-algebraic.

In particular, for every fixed $\varepsilon > 0$, $\lim_{\lambda\to 0} v_\lambda^\varepsilon(s)$ exists. Observe that if $\varepsilon_1 > \varepsilon_2$ then $y \sim_{\lambda,\varepsilon_1,s,x} y'$ implies that $y \sim_{\lambda,\varepsilon_2,s,x} y'$, so that $\widetilde{r}_\lambda^{\varepsilon_1}(s,x,z) \geq \widetilde{r}_\lambda^{\varepsilon_2}(s,x,z)$ for every $s \in S$, every $x \in \Delta(A)$ and every $z \in \Delta(B)$. It follows that $v_\lambda^{\varepsilon_1}(s) \geq v_\lambda^{\varepsilon_2}(s)$ for every $\lambda > 0$ and every $s \in S$. In particular, the function $\varepsilon \mapsto \lim_{\lambda\to 0} v_\lambda^\varepsilon(s)$ is monotonic non-decreasing, so that the limit $v(s) = \lim_{\varepsilon\to 0} \lim_{\lambda\to 0} v_\lambda^\varepsilon(s)$ exists.

Set

$$G = \{(\lambda, \varepsilon, z) \in (0,1)^2 \times \mathbf{R}^S \mid \lambda \leq \varepsilon^2, z = v_\lambda^\varepsilon\}. \tag{3}$$

G is a semi-algebraic set, whose closure contains $(0,0,v)$. Indeed, for every $\eta > 0$ there is $\varepsilon_0 > 0$ sufficiently small such that $\| \lim_{\lambda\to 0} v^\varepsilon_\lambda - v\| < \eta$ for every $\varepsilon \in (0,\varepsilon_0)$. Hence for every $\varepsilon \in (0,\varepsilon_0)$ there is $\lambda_0(\varepsilon) \in (0,1)$ such that $\|v^\varepsilon_\lambda - v\| < 2\eta$ for every $\lambda \in (0,\lambda_0(\varepsilon))$.[3]

By the Curve Selection Theorem (see, e.g. Bochnak et al., 1998, Theorem 2.5.5) there is a continuous semi-algebraic function $f : (0,1) \to G$ such that $\lim_{r\to 0} f(r) = (0,0,v)$.

Write $f(r) = (\lambda(r), \varepsilon(r), v^{\varepsilon(r)}_{\lambda(r)})$. The functions $r \mapsto \lambda(r)$, $r \mapsto \varepsilon(r)$ and $r \mapsto v^{\varepsilon(r)}_{\lambda(r)}(s)$ (for $s \in S$) are semi-algebraic, hence monotone in a neighborhood of zero. Since $\lambda > 0$ for each $(\lambda,\varepsilon,v) \in G$, and since $\lim_{r\to 0} \lambda(r) = 0$, the function $\lambda(r)$ is invertible in a neighborhood of zero. Hence, there is a semi-algebraic function $\lambda \mapsto \varepsilon(\lambda)$ such that, in a neighborhood of 0, $(\lambda, \varepsilon(\lambda), v^{\varepsilon(\lambda)}_\lambda) \in G$ and $\lim_{\lambda\to 0} v^{\varepsilon(\lambda)}_\lambda = v$.

We denote by d the degree in λ of the function $\lambda \mapsto \varepsilon(\lambda)$. That is, $\lim_{\lambda\to 0} \lambda^d/\varepsilon(\lambda) \in (0,\infty)$. By the definition of G, $d \in (0,1/2]$.

4 Reminder on zero-sum games

We here recall a result due to Mertens and Neyman (1981, hereafter MN). We let $\lambda \mapsto w_\lambda$ be a $\mathbf{R}^S$-valued semi-algebraic function defined over $(0,1)$, and we set $w := \lim_{\lambda\to 0} w_\lambda$.

Let $\varepsilon > 0$, $Z > 0$ and two functions $\lambda : (0,+\infty) \to (0,1)$ and $L : (0,+\infty) \to \mathbf{N}$ be given. Assume that the following conditions are satisfied for every $z \geq Z$, every $|\eta| \leq 4$ and every $s \in S$:

C1 $|w_\lambda(s) - w(s)| \leq \varepsilon/12$;

C2 $L(z) \leq \varepsilon z/192$;

C3 $|\lambda(z + \eta L(z)) - \lambda(z)| \leq \varepsilon\lambda(z)/48$;

C4 $\left|w_{\lambda(z+\eta L(z))}(s) - w_{\lambda(z)}(s)\right| \leq \varepsilon L(z)\lambda(z)/12$;

C5 $\int_Z^\infty \lambda(z)dz \leq \varepsilon/12$;

C6 $\lambda(z)L(z) \leq \varepsilon/48$.

MN note that **C1**-**C6** hold for Z large enough, in each of the next two cases:

Case 1: $\lambda(z) = z^{-\beta}$ and $L(z) = \lceil \lambda(z)^{-\alpha} \rceil$,[4] where $\alpha \in (0,1)$ and $\beta > 1$ satisfy $\alpha\beta < 1$;

Case 2: $\lambda(z) = 1/(z \ln^2 z)$ and $L(z) = 1$.

Let $(\widehat{r}_k)_{k\in\mathbf{N}}$ be a $[0,1]$-valued process defined on the set of plays. Define recursively processes $(z_k), (L_k), (\lambda_k)$ and (B_k) by the formulas

$$\begin{aligned} z_1 &= Z, B_1 = 1,\\ \lambda_k &= \lambda(z_k), L_k = L(z_k), B_{k+1} = B_k + L_k,\\ z_{k+1} &= \max\left\{ Z, z_k + L_k \widehat{r}_k - \sum_{B_k \leq n < B_{k+1}} w(s_n) + \frac{\varepsilon}{2} \right\}. \end{aligned}$$

[3] The condition $\lambda \leq \varepsilon^2$ in (3) can be replaced by $\lambda \leq \varepsilon^c$ for any $c > 1$.

[4] For every $c \in \mathbf{R}$, $\lceil c \rceil$ is the minimal integer greater than or equal to c.

Theorem 10 (Mertens and Neyman, 1981) *Suppose that $\widehat{r}_k$ is $\mathcal{H}^1_{B_k}$-measurable for every $k \in \mathbf{N}$, and suppose that the strategy pair (σ, τ) satisfies, for every $k \in \mathbf{N}$,*

$$\mathbf{E}_{s,\sigma,\tau}\left[\lambda_k L_k \widehat{r}_k + (1-\lambda_k L_k) w_{\lambda_k}(s_{B_{k+1}}) | \mathcal{H}^1_{B_k}\right] \geq w_{\lambda_k}(s_{B_k}) - \frac{\varepsilon}{12}\lambda_k L_k. \qquad (4)$$

Then there exists $N_0 \in \mathbf{N}$, such that for every $n \geq N_0$

$$\mathbf{E}_{s,\sigma,\tau}\left[\frac{1}{n}\sum_{i=1}^{n} \widehat{R}_i\right] \geq w(s) - \varepsilon, \qquad (5)$$

where $\widehat{R}_i = \widehat{r}_k$ whenever $B_k \leq i < B_{k+1}$. Moreover,

$$\mathbf{E}_{s,\sigma,\tau}\left[\sum_{k=1}^{+\infty} \lambda_k L_k\right] < +\infty. \qquad (6)$$

The result also holds when replacing in (4) and (5) '$\geq$' by '$\leq$', and the '-' sign on the right-hand side by a '+' sign.

In MN, play is divided into blocks. B_k is the first stage of block k, and L_k is the length of this block. As MN study games with perfect monitoring, where payoffs are observed, in their setup $\widehat{r}_k$ is the average payoff in block k. In our model payoffs are not observed, and $\widehat{r}_k$ will be an estimate for the average payoff in block k.

Remark: Theorem 10 differs from the result proven in MN (their Section 3) in two respects. (i) In the definition of z_{k+1} we use the term $L_k \widehat{r}_k$, whereas MN use the sum of stage payoffs along block k, and (ii) since MN study the case of perfect monitoring, they condition on $\mathcal{H}_{B_k}$ in (4), whereas we condition on $\mathcal{H}^1_{B_k}$. Lemma 3.4 in MN can be easily adapted to deal with the first point. On the other hand, the second point does not affect the proof.

5 Player 2 can defend v

We prove in this section that player 2 can defend v.

Let an initial state $s \in S$, $\varepsilon > 0$ and a strategy σ of player 1 be given. It is enough to prove that there exists a strategy $\overline{\tau}$ such that $\gamma_n(s, \sigma, \overline{\tau}) \leq v(s) + 2\varepsilon$ for every n sufficiently large.

For every $s \in S$, $\lambda > 0$ and $x \in \Delta(A)$ we choose $y^s_\lambda(x) \in \Delta(B)$ such that

$$\lambda \widetilde{r}^\varepsilon_\lambda(s, x, y^s_\lambda(x)) + (1-\lambda)\mathbf{E}_{q(\cdot|s,x,y^s_\lambda(x))}\left[v^\varepsilon_\lambda\right] \leq v^\varepsilon_\lambda(s).$$

We also choose $z^s_\lambda(x) \sim_{\lambda,\varepsilon,s,x} y^s_\lambda(x)$ such that $r(s, x, z^s_\lambda(x)) = \widetilde{r}^\varepsilon_\lambda(s, x, y^s_\lambda(x))$.

We are now going to define a strategy τ, in the spirit of MN (see Section 4, **Case 2** ($L(z) = 1$, $\lambda(z) = 1/(z \ln^2 z)$), with $\widehat{r}_n = \widetilde{r}^\varepsilon_{\lambda_n}(s_n, \xi_n, y^{s_n}_{\lambda_n})$ for ξ_n defined below, and $w_\lambda = v^\varepsilon_\lambda$.)

Suppose the strategy τ is defined for the first $n-1$ stages. At stage n, τ plays the mixed action $y^{s_n}_{\lambda_n}(\xi_n)$, where ξ_n is the expected mixed action of player 1 given the sequence of states visited so far: $\xi_n[a] = \mathbf{P}_{s,\sigma,\tau}(a_n = a \mid s_1, \ldots, s_n)$, $a \in A$. Since the computation of ξ_n involves only the restriction of τ to the first $n-1$ stages, there is no circularity in this definition. The calculation of λ_n is explicitly described in Section 4. As ξ_n is $\mathcal{H}^1_n$-measurable, $y^{s_n}_{\lambda_n}(\xi_n)$, $z^{s_n}_{\lambda_n}(\xi_n)$, $\widehat{r}_n$ and λ_n are $\mathcal{H}^1_n$-measurable as well.[5]

Applying Theorem 10 to (σ, τ), we conclude that both (5) (with the inequality reversed) and (6) are satisfied.

By (6), there is $N \in \mathbf{N}$ such that

$$\mathbf{E}_{s,\sigma,\tau}\left[\sum_{n=N}^{+\infty} \lambda_n\right] < \frac{\varepsilon^2}{|A|}. \tag{7}$$

We now define the strategy $\overline{\tau}$ for player 2 as follows. It coincides with τ up to stage N, *i.e.* it plays $y^{s_n}_{\lambda_n}(\xi_n)$ in each stage $n < N$. In each stage $n \geq N$, $\overline{\tau}$ plays $z^{s_n}_{\lambda_n}(\xi_n)$.

The definition of $\overline{\tau}$ is reminiscent of the type of replies defined by Coulomb (2001). Loosely speaking, under $\overline{\tau}$, player 2 plays for good transitions up to stage N, and for low payoffs afterwards. We now check that $\gamma_n(s, \sigma, \overline{\tau}) \leq v(s) + 2\varepsilon$, for n large enough. Note that the strategy $\overline{\tau}$ uses only the sequence of states, and not any additional signal that player 2 may receive.

For every $n \in \mathbf{N}$ define the (random) set

$$A_n = \{a \in A \colon \xi_n[a] \geq \lambda_n/\varepsilon\}.$$

This is the set of all actions that are relevant for the equivalence relation at stage n. By definition,

$$\mathbf{P}_{s,\sigma,\tau}(a_n \in A_n \mid s_1, \ldots, s_n) = \sum_{a \in A_n} \xi_n[a] \geq 1 - \frac{|A|\lambda_n}{\varepsilon}, \text{ for each } n \geq N.$$

By taking expectations, by summation over n and (7) this implies that

$$\mathbf{P}_{s,\sigma,\tau}(\xi_n[a_n] \geq \lambda_n/\varepsilon, \quad \forall n \geq N) \geq 1 - \varepsilon. \tag{8}$$

Set $\overline{H}_\infty = \{h \in H_\infty \colon \xi_n[a_n] \geq \lambda_n/\varepsilon \quad \forall n \geq N\}$. On the set $\overline{H}_\infty$, one has $\psi^1(s_n, a_n, y^{s_n}_{\lambda_n}(\xi_n)) = \psi^1(s_n, a_n, z^{s_n}_{\lambda_n}(\xi_n))$ for each $n \in \mathbf{N}$. In particular, for every $F \in \mathcal{H}^1_\infty$, $\mathbf{P}_{s,\sigma,\overline{\tau}}(F \cap \overline{H}_\infty) = \mathbf{P}_{s,\sigma,\tau}(F \cap \overline{H}_\infty)$. Therefore, by (8), $\sup_{F \in \mathcal{H}^1_\infty} |\mathbf{P}_{s,\sigma,\overline{\tau}}(F) - \mathbf{P}_{s,\sigma,\tau}(F)| < \varepsilon$. Since ξ_n and $y^{s_n}_{\lambda_n}(\xi_n)$ are $\mathcal{H}^1_n$-measurable, this yields

$$\left|\mathbf{E}_{s,\sigma,\tau}\left[\widetilde{r}^{\varepsilon}_{\lambda_n}(s_n, \xi_n, y^{s_n}_{\lambda_n}(\xi_n))\right] - \mathbf{E}_{s,\sigma,\overline{\tau}}\left[\widetilde{r}^{\varepsilon}_{\lambda_n}(s_n, \xi_n, y^{s_n}_{\lambda_n}(\xi_n))\right]\right| \leq \varepsilon, \text{ for every } n \in \mathbf{N}.$$

By the choice of $z^{s_n}_{\lambda_n}(\xi_n)$, $\mathbf{E}_{s,\sigma,\overline{\tau}}\left[\widetilde{r}^{\varepsilon}_{\lambda_n}(s_n, \xi_n, y^{s_n}_{\lambda_n}(\xi_n))\right] = \mathbf{E}_{s,\sigma,\overline{\tau}}\left[r(s_n, a_n, b_n)\right]$, for every $n \geq N$. By summation, one obtains for every $n \geq N/\varepsilon$,

$$\gamma_n(s, \sigma, \overline{\tau}) = \mathbf{E}_{s,\sigma,\overline{\tau}}\left[\frac{1}{n}\sum_{i=1}^{n} r(s_i, a_i, b_i)\right] \leq \mathbf{E}_{s,\sigma,\tau}\left[\frac{1}{n}\sum_{i=1}^{n} \widetilde{r}^{\varepsilon}_{\lambda_i}(s_i, \xi_i, y^{s_i}_{\lambda_i}(\xi_i))\right] + \varepsilon.$$

By (5) with the inequality reversed, this yields

$$\gamma_n(s, \sigma, \overline{\tau}) \leq \lim_{\lambda \to 0} v^{\varepsilon}_{\lambda}(s) + 2\varepsilon,$$

for every n sufficiently large, as desired.

[5] Actually, those random variables are measurable w.r.t. the coarser algebra generated by $s_1, \ldots, s_n$.

6 Player 1 can guarantee v

We fix $\varepsilon \in (0,1)$ once and for all. Our goal is to construct a strategy σ that guarantees v up to 3ε.

The structure of the proof is as follows. In Section 6.1 we define a strategy σ, that plays in blocks, in the spirit of Mertens and Neyman. In Section 6.2 we define the process $(\widehat{r}_k)$, which is used as a sufficient statistic for the average payoff in block k in the definition of σ. To apply Theorem 10 we have to show that (4) is satisfied. We prove this in Section 6.3. We then have to relate the average estimated payoff to the average payoff. This is done in Section 6.4.

6.1 Definition of a strategy

By Section 3.4, there is a semi-algebraic function $\lambda \mapsto \varepsilon(\lambda)$ such that $\lim_{\lambda \to 0} \varepsilon(\lambda) = 0$, and $\lim_{\lambda \to 0} v_\lambda^{\varepsilon(\lambda)}(s) = v(s)$ for every $s \in S$. Recall that $d \in (0, \frac{1}{2}]$ is the degree of $\lambda \mapsto \varepsilon(\lambda)$. For notational simplicity, we abbreviate $v_\lambda^{\varepsilon(\lambda)}$ and $\widetilde{r}_\lambda^{\varepsilon(\lambda)}$ to v_λ and $\widetilde{r}_\lambda$ respectively. For $\lambda \in (0,1)$, we let $x_\lambda \in (\Delta(A))^S$ achieve the maximum in the definition of v_λ. Specifically, $\lambda \mapsto x_\lambda$ is a semi-algebraic function that satisfies for every $s \in S$ and every $\lambda > 0$,

$$\lambda \widetilde{r}_\lambda(s, x_\lambda^s, y) + (1-\lambda)\mathbf{E}_{q(\cdot|s,x_\lambda^s,y)}[v_\lambda(\cdot)] \geq v_\lambda(s) \quad \forall y \in \Delta(B). \tag{9}$$

Proposition 9 implies that such a function exists.

Define for every $s \in S$

$$\overline{A}(s) = \left\{ a \in A \colon x_\lambda^s[a] \geq \frac{\lambda}{\varepsilon(\lambda)}, \quad \text{for every } \lambda \text{ sufficiently small} \right\}.$$

Since $\lambda \mapsto x_\lambda$ and $\lambda \mapsto \varepsilon(\lambda)$ are semi-algebraic, one has $x_\lambda^s[a] < \frac{\lambda}{\varepsilon(\lambda)}$ for every $a \notin \overline{A}(s)$ and every λ sufficiently small.

We now define a strategy σ in the spirit of Mertens and Neyman, see Section 4, **Case 1**, with $w_\lambda = v_\lambda$. We choose $\alpha \in (1-d, 1)$, and $\beta \in (1, 1/\alpha)$. The strategy σ plays in blocks; block k starts at (random) stage B_k, and lasts for L_k stages. During this block, player 1 plays the stationary strategy x_{λ_k}. The processes λ_k, L_k and B_k are explicitly defined in Section 4; the process $(\widehat{r}_k)$ is defined in the next section.

The parameter Z will be chosen later, to satisfy various conditions.

6.2 The statistic $\widehat{\mathbf{r}}_k$

We here proceed with the definition of $\widehat{r}_k$. The value of $\widehat{r}_k$ depends only on the sequence of signals received during block k. Most of the analysis in the subsequent sections deals with a given block. Therefore, for notational simplicity, we drop the subscript k: we thus write L instead of L_k, λ instead of λ_k, etc. We also relabel the stages of block k from 1 to L, so that $B_{k+1} = L + 1$.

For $s \in S$ and $a \in \overline{A}(s)$, we let $\rho_{s,a} \in \Delta(M^1)$ stand for the empirical distribution of signals received by player 1 in the stages where a was played at state s:

$$\rho_{s,a}[m] = \frac{|\{n \leq L, m_n^1 = m\}|}{|\{n \leq L, (s_n, a_n) = (s,a)\}|} \qquad \forall m \in M^1.$$

For $s \in S$, we let $\widehat{y}^s \in \Delta(B)$ minimize $\max_{a\in\overline{A}(s)} \left\|\rho_{s,a} - \psi^1(s,a,\cdot)\right\|_\infty$. Finally, we set

$$\widehat{r} = \sum_{s\in S} \mathbf{N}_s \widetilde{r}_\lambda(s, x^s_\lambda, \widehat{y}^s),$$

where $\mathbf{N}_s = |\{n \leq L, s_n = s\}$ is the number of visits to s during the current block. In effect, at the end of each block, player 1 computes a stationary strategy that is most consistent with the sequence of signals, and $\widehat{r}$ is the corresponding worst payoff.

6.3 The assumptions of Theorem 10 hold

We will prove that inequality (4) always holds, provided Z is large enough. The proof is the same for the different blocks. Hence, we shall focus on a generic block, and will omit the corresponding subscript k.

We set $\delta = \varepsilon/(24|S| + 24)$. Let $\eta \in (0,1)$ satisfy the conclusion of Lemma 3 w.r.t. δ.

We introduce the mixed move y_n used by player 2 at stage n. Specifically, $y_n[b] = \mathbf{P}_{s,\sigma,\tau}(b_n = b \mid \mathcal{H}^2_n)$. We also let $\overline{y} = (\overline{y}^s)_{s\in S}$ denote the empirical stationary strategy used by player 2 during the block. Formally, for $s \in S$, $\overline{y}^s = \frac{1}{\mathbf{N}_s}\sum_{n\colon s_n = s} y_n$.[6]

In the next proof, we use the following observation. For $\varepsilon \in (0, 1/3)$ and $a, b > 0$ one has

$$\left|\frac{a/A}{b/B} - 1\right| < 3\varepsilon \text{ whenever } \left|\frac{a}{b} - 1\right| < \varepsilon \text{ and } \left|\frac{A}{B} - 1\right| < \varepsilon. \tag{10}$$

Proposition 11 *There is $Z_1 > 0$ such that for every k, if $z_k \geq Z_1$, the following holds. For every strategy τ of player 2, and every $s \in S$,*

$$\left|\mathbf{E}_{s,x_\lambda,\tau}[\widehat{r}] - \frac{1}{L}\mathbf{E}_{s,x_\lambda,\tau}\left[\sum_{t\in S}\mathbf{N}_t \widetilde{r}_\lambda(t, x^t_\lambda, \overline{y}^t)\right]\right| \leq 2|S|\delta.$$

Recall that $\lambda = \lambda_k = \lambda(z_k)$ and $L = L_k = L(z_k)$, so that if z_k is large than λ is small and L is large.

Proof. Let an initial state $s \in S$ and a strategy τ be given. All probabilities and expectations below are taken w.r.t. $\mathbf{P}_{s,x_\lambda,\tau}$. Let $t \in S$ and $a \in \overline{A}(t)$ be given, and let $m^1 \in M^1$ be an arbitrary signal such that $\psi^1(t,a,b)[m^1] > 0$ for some $b \in B$. For $i = 1, \ldots, L$, set $W_i = 1$ if $m^1_i = m^1$, and $W_i = 0$ otherwise. Note that $\mathbf{E}[W_i|\mathcal{H}_i] = \psi^1(s_i, x_\lambda, y_i)[m^1]$. The random variables $W_i - \mathbf{E}[W_i|\mathcal{H}_i]$, $i = 1, \ldots, L$, are centered and uncorrelated. By Chebyshev's inequality, letting $S_L = \frac{1}{L}\sum_{i=1}^L (W_i - \mathbf{E}[W_i|\mathcal{H}_i])$, one has for every $c > 0$

$$\mathbf{P}(|S_L| \geq c) \leq \frac{\mathrm{Var}(S_L)}{c^2} = \frac{1}{L^2c^2}\sum_{i=1}^{L}\mathrm{Var}(W_i) \leq \frac{x^t_\lambda(a)}{Lc^2}. \tag{11}$$

Denote by $\mathbf{N}_{t,a} = |\{n \leq L : (s_n, a_n) = (t,a)\}|$ the number of stages in the block where the play visited state t and player 1 played the action a, and by $\mathbf{N}_{m^1} = \left|\left\{n \leq L\colon m^1_n = m^1\right\}\right|$ the number

[6] Note that $\overline{y}^s$ need not coincide with the empirical distribution of the actions actually chosen at state s.

of stages in the block where player 1 observed the signal m^1. Since the signal contains the current state and the chosen action, $\mathbf{N}_{t,a} \geq \mathbf{N}_{m^1}$ (recall that there is a b such that $\psi^1(t,a,b)[m^1] > 0$).

Note that $\sum_{i=1}^{L} W_i = \mathbf{N}_{m^1}$, while $\sum_{i=1}^{L} \mathbf{E}[W_i|\mathcal{H}_i] = \mathbf{N}_t x_\lambda^t[a]\psi^1(t,a,\overline{y}^t)[m^1]$. Hence, by (11),

$$\mathbf{P}\left(\left|\frac{\mathbf{N}_{m^1}}{L} - \frac{\mathbf{N}_t}{L} x_\lambda^t[a]\psi^1(t,a,\overline{y}^t)[m^1]\right| \geq c\right) \leq \frac{x_\lambda^t[a]}{Lc^2}.$$

Keeping a fixed and applying this upper bound for each $m^1 \in M^1$, one gets that, with probability at least $1 - \frac{|M^1|x_\lambda^t[a]}{Lc^2}$, both

$$\left|\frac{\mathbf{N}_{t,a}}{L} - \frac{\mathbf{N}_t}{L} x_\lambda^t[a]\right| < |M^1|c \tag{12}$$

and

$$\left|\frac{\mathbf{N}_{m^1}}{L} - \frac{\mathbf{N}_t}{L} x_\lambda^t[a]\psi^1(t,a,\overline{y}^t)[m^1]\right| < c \tag{13}$$

hold, for every $m^1 \in M^1$. Denote by $\mathcal{E}_0$ the corresponding event.

By (10), on $\mathcal{E}_0$ one has $|\rho_{t,a}[m^1] - \psi^1(t,a,\overline{y}^t)[m^1]| < 3|M^1|\frac{cL}{\mathbf{N}_t x_\lambda^t[a]}$, provided $\frac{|M^1|cL}{\mathbf{N}_t x_\lambda^t[a]\psi^1(t,a,\overline{y}^t)[m^1]} < 1/3$. Set $\mathcal{E}_1 = \mathcal{E}_0 \cap \{\mathbf{N}_t/L \geq \delta\}$.

Since $a \in \overline{A}(t)$, the degree of $\lambda \mapsto x_\lambda^t[a]$ in λ, $\deg_\lambda(x_\lambda^t(a))$, is at most $1-d$. Recall that $1-d < \alpha < 1$. Choose $\gamma \in \left(\deg_\lambda(x_\lambda^t[a]), \frac{\deg_\lambda(x_\lambda^t[a])+\alpha}{2}\right)$. We will use the estimates we derived in the previous paragraphs with $c = \lambda^\gamma$.

Provided that z_k is sufficiently large, when $c = \lambda^\gamma$, we have, on $\mathcal{E}_1$,

$$\frac{|M^1|x_\lambda^t[a]}{Lc^2} \leq \frac{\delta}{|A|}, \text{ and } \frac{|M^1|cL}{\mathbf{N}_t x_\lambda^t[a]} < \eta/12 < 1/2. \tag{14}$$

If in addition $\psi^1(t,a,\overline{y}^t)[m^1] \geq \eta/4$, then

$$\frac{|M^1|cL}{\mathbf{N}_t x_\lambda^t[a]\psi^1(t,a,\overline{y}^t)[m^1]} < 1/3$$

Therefore, on $\mathcal{E}_1$ one has $|\rho_{t,a}[m^1] - \psi^1(t,a,\overline{y}^t)[m^1]| < \eta$ whenever $\psi^1(t,a,\overline{y}^t)[m^1] \geq \eta/4$.

On the other hand, if $\psi^1(t,a,\overline{y}^t)[m^1] < \eta/4$, one has on $\mathcal{E}_1$, by (12), (13), and (14)

$$\begin{aligned}
|\rho_{t,a}[m^1] - \psi^1(t,a,\overline{y}^t)[m^1]| &\leq \frac{\mathbf{N}_{m^1}}{\mathbf{N}_{t,a}} + \psi^1(t,a,\overline{y}^t)[m^1] \\
&\leq \frac{\mathbf{N}_t x_\lambda^t[a]\psi^1(t,a,\overline{y}^t)[m^1] + cL}{\mathbf{N}_t x_\lambda^t[a] - |M^1|cL} + \eta/4 \\
&\leq 2\frac{\mathbf{N}_t x_\lambda^t[a]\psi^1(t,a,\overline{y}^t)[m^1] + cL}{\mathbf{N}_t x_\lambda^t[a]} + \eta/4 \\
&\leq 2\psi^1(t,a,\overline{y}^t)[m^1] + \eta/6 + \eta/4 \leq \eta
\end{aligned}$$

Thus, for every $t \in S$, with probability at least $1-\delta/|A|$, one has $||\rho_{t,a} - \psi^1(t,a,\overline{y}^t)||_\infty < \eta$ whenever $\mathbf{N}_t/L \geq \delta$.

Letting $a \in \overline{A}(t)$ vary, we deduce that

$$\mathbf{P}\left(\frac{\mathbf{N}_t}{L} \leq \delta, \text{ or } \|\rho_{t,a} - \psi^1(t,a,\overline{y}^t)\|_\infty < \eta \ \ \forall a \in \overline{A}(t)\right) \geq 1 - \delta.$$

Since $\|\rho_{t,a} - \psi^1(t,a,\overline{y}^t)\|_\infty < \eta$ implies that $|\widetilde{r}_\lambda(t, x^t_\lambda, \widehat{y}^t) - \widetilde{r}_\lambda(t, x^t_\lambda, \overline{y}^t)| \leq \delta$, we conclude that

$$\mathbf{E}\left[\frac{\mathbf{N}_t}{L}\left|\widetilde{r}_\lambda(t, x^t_\lambda, \widehat{y}^t) - \widetilde{r}_\lambda(t, x^t_\lambda, \overline{y}^t)\right|\right] \leq 2\delta.$$

The result follows by summation over $t \in S$. ■

Proposition 12 *If z_k is sufficiently large the following holds. For every strategy τ of player 2, and every initial state $s \in S$,*

$$\mathbf{E}_{s,x_\lambda,\tau}[\lambda L \widehat{r} + (1 - \lambda L)v_\lambda(s_{L+1})] \geq v_\lambda(s) - \frac{\varepsilon}{12}\lambda L.$$

Proof. All expectations below are taken w.r.t. $\mathbf{P}_{s,x_\lambda,\tau}$. Suppose z_k is large enough so that (i) $\lambda L \leq \delta$, (ii) $0 \leq (1-\lambda)^L - (1 - \lambda L) \leq \delta\lambda L$, (iii) $(1-\lambda)^L \geq 1 - \delta$, and (iv) the conclusion of Proposition 11 holds. Since $L(z) = \lceil \lambda(z) \rceil^{-\alpha}$, $\lim_{z\to\infty}(1-\lambda(z))^{L(z)} = 1$, so that (iii) holds, provided z_k is sufficiently large. Since $e^{-x} + (1-\delta)x < 1$ in a positive neighborhood of 0, (ii) holds provided z_k is sufficiently large. Under (i)-(iv) one has:

$$\mathbf{E}\left[\lambda L \widehat{r} + (1 - \lambda L)v_\lambda(s_{L+1})\right] - v_\lambda(s) + 2|S|\delta\lambda L \tag{15}$$

$$\geq \mathbf{E}\left[\lambda \sum_{t\in S} \mathbf{N}_t \widetilde{r}_\lambda(t, x^t_\lambda, \overline{y}^t) + (1 - \lambda L)v_\lambda(s_{L+1})\right] - v_\lambda(s) \tag{16}$$

$$\geq \mathbf{E}\left[\lambda \sum_{t\in S} \mathbf{N}_t \widetilde{r}_\lambda(t, x^t_\lambda, \overline{y}^t) + (1 - \lambda)^L v_\lambda(s_{L+1})\right] - v_\lambda(s) - \delta\lambda L \tag{17}$$

$$= \mathbf{E}\left[\lambda \sum_{t\in S} \mathbf{N}_t \widetilde{r}_\lambda(t, x^t_\lambda, \overline{y}^t) + \sum_{i=1}^{L}\left((1-\lambda)^i v_\lambda(s_{i+1}) - (1-\lambda)^{i-1}v_\lambda(s_i)\right)\right] - \delta\lambda L \tag{18}$$

$$= \mathbf{E}\left[\lambda \sum_{t\in S} \mathbf{N}_t \widetilde{r}_\lambda(t, x^t_\lambda, \overline{y}^t) + \sum_{i=1}^{L}(1-\lambda)^{i-1}\left((1-\lambda) v_\lambda(s_{i+1}) - v_\lambda(s_i)\right)\right] - \delta\lambda L \tag{19}$$

$$\geq (1-\lambda)^L\mathbf{E}\left[\lambda \sum_{t\in S} \mathbf{N}_t \widetilde{r}_\lambda(t, x^t_\lambda, \overline{y}^t) + \sum_{i=1}^{L}\left((1-\lambda) v_\lambda(s_{i+1}) - v_\lambda(s_i)\right)\right] - 2\delta\lambda L \tag{20}$$

$$= (1-\lambda)^L\mathbf{E}\left[\lambda \sum_{t\in S} \mathbf{N}_t \widetilde{r}_\lambda(t, x^t_\lambda, \overline{y}^t) + \sum_{i=1}^{L}\Big((1-\lambda) \mathbf{E}_{q(\cdot|s_i, x^{s_i}_\lambda, y^{s_i}_i)}[v_\lambda(\cdot)] - v_\lambda(s_i)\Big)\right] - 2\delta\lambda L \tag{21}$$

$$= (1-\lambda)^L\mathbf{E}\left[\sum_{t\in S}\left(\lambda \mathbf{N}_t \widetilde{r}_\lambda(t, x^t_\lambda, \overline{y}^t) + (1-\lambda)\sum_{i:s_i=t} \mathbf{E}_{q(\cdot|t, x^t_\lambda, y^t_i)}[v_\lambda(\cdot)] - v_\lambda(t)\right)\right] - 2\delta\lambda L \tag{22}$$

$$= (1-\lambda)^L\mathbf{E}\left[\sum_{t\in S}\mathbf{N}_t\Big(\lambda \widetilde{r}_\lambda(t, x^t_\lambda, \overline{y}^t) + (1-\lambda) \mathbf{E}_{q(\cdot|t, x^t_\lambda, \overline{y}^t)}[v_\lambda(\cdot)] - v_\lambda(t)\Big)\right] - 2\delta\lambda L \tag{23}$$

$$\geq -2\delta\lambda L. \tag{24}$$

The transition from (15) to (16) follows from Proposition 11. The transition from (16) to (17) holds by (ii). The transitions from (17) to (18) and from (18) to (19) are immediate. The transition from (19) to (20) holds for the following reasons: For the first term, observe that $(1-\lambda)^L < 1$ and payoffs are non-negative. For the second term, observe that $0 \leq (1-\lambda)^{i-1} - (1-\lambda)^L \leq \delta$, while $\mathbf{E}[(1-\lambda)v_\lambda(s_{i+1}) - v_\lambda(s_i)|\mathcal{H}_i] \geq -\lambda$, so that their product is at least $-\delta\lambda$. The transition from (20) to (21) holds by the law of iterated expectations: $\mathbf{E}[v_\lambda(s_{i+1})] = \mathbf{E}[\mathbf{E}[v_\lambda(s_{i+1})|\mathcal{H}_i]]$ and $\mathbf{E}[v_\lambda(s_{i+1})|\mathcal{H}_i] = \mathbf{E}_{q(\cdot|s_i, x_\lambda^{s_i}, y_i^{s_i})}[v_\lambda(\cdot)]$. The transition from (21) to (22) is a simple rewriting. The transition from (22) to (23) holds by the linearity of q and the definition of $\overline{y}$. The transition from (23) to (24) holds by the optimality of x_λ.

The proposition follows since $\delta = \varepsilon/(24|S|+24)$. ■

6.4 The end of the proof: player 1 can guarantee v

We consider the strategy σ that was defined in Section 6.1. We let Z be sufficiently large so that for every $k \in \mathbf{N}$, Propositions 11 and 12 hold, and for every $s \in S$, the set $\{a \in A \colon x_\lambda^s[a] \geq \lambda/\varepsilon(\lambda)\}$ is independent of λ, provided $\lambda \leq \lambda(Z)$. We prove that there exists $N_0 \in \mathbf{N}$, such that $\gamma_n(s,\sigma,\tau) \geq v(s) - 3\varepsilon$ for every $s \in S$, every strategy τ of player 2, and every $n \geq N_0$.

Let an initial state $s \in S$, and a strategy τ be given. All expectations below are taken w.r.t. $\mathbf{P}_{s,\sigma,\tau}$. We first rewrite the conditional average payoff in block k. Denote by $\overline{y}(k) = (\overline{y}^s(k))_{s\in S}$ the empirical mixed move played by player 2 in block k (previously denoted $\overline{y}$). By the law of iterated expectations, and the linearity of r, one has

$$\begin{aligned}
\mathbf{E}\left[\frac{1}{L_k}\sum_{n=B_k}^{B_{k+1}-1} r(s_n,a_n,b_n)|\mathcal{H}_{B_k}\right] &= \mathbf{E}\left[\frac{1}{L_k}\sum_{n=B_k}^{B_{k+1}-1} \mathbf{E}_{s,x_{\lambda_k},\tau}[r(s_n,a_n,b_n) \mid \mathcal{H}_n]|\mathcal{H}_{B_k}\right] \\
&= \mathbf{E}\left[\frac{1}{L_k}\sum_{n=B_k}^{B_{k+1}-1} r(s_n,x_{\lambda_k}^{s_n},y_n^{s_n})|\mathcal{H}_{B_k}\right] \\
&= \mathbf{E}\left[\frac{1}{L_k}\sum_{n=B_k}^{B_{k+1}-1} r(s_n,x_{\lambda_k}^{s_n},\overline{y}^{s_n}(k))|\mathcal{H}_{B_k}\right]. \qquad (25)
\end{aligned}$$

Set $\kappa_n = \sup\{k : B_{k+1} \leq n\}$. This is the index of the last block that ends before or at stage n. One has the identity

$$\sum_{i=1}^{n}(r(s_i,a_i,b_i) - \widehat{R}_i) = \sum_{k=1}^{+\infty}\mathbf{1}_{\kappa_n \geq k}\sum_{i=B_k}^{B_{k+1}-1}(r(s_i,a_i,b_i) - \widehat{R}_i) + \sum_{i=B_{\kappa_n}}^{n}(r(s_i,a_i,b_i) - \widehat{R}_i).$$

By the law of iterated expectations, the triangle inequality, since the event $\{\kappa_n \geq k\}$ is $\mathcal{H}_{B_k}$-measurable, and since payoffs are bounded by 1, this yields

$$\begin{aligned}
&\mathbf{E}\left[\sum_{i=1}^{n}(r(s_i,a_i,b_i) - \widehat{R}_i)\right] \geq \\
&\quad \mathbf{E}\left[\sum_{k=1}^{+\infty}\mathbf{1}_{\kappa_n \geq k}\mathbf{E}\left[\sum_{i=B_k}^{B_{k+1}-1}(r(s_i,a_i,b_i) - \widehat{R}_i)|\mathcal{H}_{B_k}\right]\right] - \mathbf{E}\left[n - B_{\kappa_n} + 1\right]. \qquad (26)
\end{aligned}$$

By (**C2**) and the definition of (z_k),

$$n - B_{\kappa_n} + 1 \leq L_{\kappa_n} \leq \varepsilon z_{\kappa_n}/192 \leq \frac{\varepsilon(z_0 + n(1+\varepsilon/2))}{192}.$$

Moreover, for each k one has by (25), (2) and Proposition 11,

$$\begin{aligned}\mathbf{E}\left[\sum_{i=B_k}^{B_{k+1}-1} (r(s_i,a_i,b_i) - \widehat{R}_i)|\mathcal{H}_{B_k}\right] &= \mathbf{E}\left[\sum_{i=B_k}^{B_{k+1}-1} (r(s_i,x_\lambda^{s_i},\overline{y}^{s_i}(k)) - \widehat{R}_i)|\mathcal{H}_{B_k}\right] \\ &\geq \mathbf{E}\left[\sum_{i=B_k}^{B_{k+1}-1} (\widetilde{r}_\lambda(s_i,x_\lambda^{s_i},\overline{y}^{s_i}(k)) - \widehat{R}_i)|\mathcal{H}_{B_k}\right] \\ &\geq -\varepsilon L_k.\end{aligned}$$

Hence, Eq. (26) implies

$$\mathbf{E}\left[\sum_{i=1}^{n}(r(s_i,a_i,b_i) - \widehat{R}_i)\right] \geq -\varepsilon\mathbf{E}\left[\sum_{k=1}^{+\infty}\mathbf{1}_{\kappa_n \geq k} L_k\right] - \varepsilon(z_0 + n(1+\varepsilon/2))/192 \geq -2\varepsilon n, \tag{27}$$

where the second inequality holds for n large enough.

By Proposition 12 we can apply Theorem 10, and therefore by (5), one has $\mathbf{E}\left[\sum_{i=1}^{n}\widehat{R}_i\right] \geq n(v(s)-\varepsilon)$. By (27),

$$\mathbf{E}\left[\sum_{i=1}^{n} r(s_i,a_i,b_i)\right] \geq n(v(s) - 3\varepsilon),$$

which concludes the proof.

References

[1] Aumann R.J. and Maschler M., with the collaboration of R.E. Stearns (1995), Repeated games with incomplete information, MIT Press

[2] Bochnak J., Coste M. and Roy M. F. (1998) Real Algebraic Geometry, Springer Verlag, Berlin

[3] Coulomb J.M. (1992) Repeated Games with Absorbing States and No Signals, *Int. J. Game Th.*, **21**, 161-174

[4] Coulomb J.M. (1999) Generalized Big-Match, *Math. Oper. Res.*, **24**, 795-816

[5] Coulomb J.M. (2001) Absorbing Games with a Signaling Structure, *Math. Oper. Res.*, **26**, 286-303

[6] Coulomb J.M. (2003) Stochastic Games without Perfect Monitoring, *preprint*

[7] Lehrer E. (1989) Lower Equilibrium Payoffs in Two-Player Repeated Games with Nonobservable Actions, *Int. J. Game Th.*, **18**, 57-89

[8] Lehrer E. (1990) Nash Equilibria of n-Player Repeated Games with Semi-Standard Information, *Int. J. Game Th.*, **19**, 191- 217

[9] Lehrer E. (1992a) Correlated Equilibria in Two-Player Repeated Games with Nonobservable Actions, *Math. Oper. Res.* **17**, 175-199

[10] Lehrer E. (1992b) On the Equilibrium Payoffs Set of Two Player Repeated Games with Imperfect Monitoring, *Int. J. Game Th.*, **20**, 211-226

[11] Mertens J.F. and Neyman A. (1981) Stochastic Games, *Int. J. Game Th.*, **10**, 53-66

[12] Radner R. (1981) Monitoring Cooperative Agreements in Repeated Principle-Agent Relationship, *Econometrica*, **49**, 1127- 1148

[13] Rosenberg D., Solan E. and Vieille N.(2002) Blackwell Optimality in Markov Decision Processes with Partial Observation, *Ann. Stat.*, **30**, 1178-1193

[14] Rubinstein A. and Yaari M. (1983) Repeated Insurance Contracts and Moral Hazard, *J. Econ. Th.*, **30**, 74-97

[15] Shapley L.S. (1953) Stochastic Games, *Proc. Nat. Acad. Sci. U.S.A.*, **39**, 1095-1100

[16] Solan E. (1999) Three-Person Absorbing Games, *Math. Oper. Res.*, **24**, 669-698

[17] Solan E. and Vohra R. (2002), Correlated Equilibrium Payoffs and Public Signalling in Absorbing Games, *Int. J. Game Th.*, **31**, 91-122