# The value of Repeated Games with an informed controller

Jérôme Renault [*]

September 2011

### Abstract

We consider the general model of zero-sum repeated games (or stochastic games with signals), and assume that one of the players is fully informed and controls the transitions of the state variable. We prove the existence of the *uniform* value, generalizing several results of the literature. A preliminary existence result is obtained for a particular class of stochastic games played with pure strategies.

*Key words. Repeated games, stochastic games, uniform value, incomplete information, single controller, Choquet order, Wasserstein distance.*

## 1   Introduction

The context of this work is the characterization of repeated game models where the value exists. We first consider here general repeated games defined with finite sets of states, actions and signals. This model includes usual stochastic games, standard repeated games with incomplete information and also repeated games with signals. At each stage the players will play a matrix game depending on a parameter called state. This state is partially known and evolves from stage to stage, and after each stage every player receives some private signal on the current situation. We make two important assumptions. We first assume that player 1 is informed, in the sense that he can always deduce the current state and player 2's signal from his own signal. Second, we assume that player 1 controls the transitions, in the sense that the law of the pair (new state, signal received by player 2) does not depend on player 2's actions. We call "repeated games with an informed controller" the games satisfying these two assumptions.

This class of games includes Markov chain repeated games with lack of information on one side introduced in Renault, 2006, where the states follow an

exogenous Markov chain and actions are perfectly observed at the end of each stage. It is a much more general class since, in particular, we allow here for transitions of the state depending also on player 1's actions. It also contains stochastic games with a single controller and incomplete information on the side of his opponent, as studied in Rosenberg *et al.*, 2004 (see subsection 4.3.3 here). *A fortiori* it contains the standard repeated games with incomplete information on one side and perfect monitoring introduced by Aumann and Maschler. Note that repeated games with an informed controller contains (weak) forms of the three main features of general repeated games models : the stochastic game aspect (the state evolves from one stage to another and is controlled here by player 1), the incomplete information aspect (player 2 has an incomplete knowledge of the state), and the signalling aspect (players observe signals rather than actions). We believe that the existence result presented here is the first one to significantly deal with these three aspects simultaneously. Yet, repeated games with an informed controller do not include stochastic games, where the transitions are controlled by both players (see Mertens and Neyman, 1981 for the existence of the uniform value in such games).

We prove the existence of the uniform value via several steps, and several games are considered. These are : our original repeated game where player 1 is informed and controls the transitions (level 1), an auxiliary stochastic game (level 2), and finally a one-player repeated game, i.e. a dynamic programming problem (level 3). A crucial point is that in our original game, the set of states $K$ is finite.

The auxiliary stochastic game has the following features. It is played with pure strategies, and derived from the original game by 'assuming' that the mixed moves of player 1 are observed by player 2 : an action for player 1 will be an element of $\Delta(I)^K$, where $I$ is the original set of pure actions of player 1. The new set of states is $X = \Delta(K)$, the set of probabilities over $K$, and represents in the original game the belief of player 2 on the current state [1]. In the auxiliary stochastic game, the new state is known to both players, and actions played are perfectly observed after each stage. Because we assumed that player 1 is informed at level 1, he can guarantee in the original game whatever he can guarantee in the auxiliary game (note that this property is not true for player 2, because in the original game he needs to know the strategy of player 1 to compute his belief on the current state). Using minmax theorems and recursive formulas, one can prove that the $n$-stage values of the original and the auxiliary game are equal. This implies that if player 1 can guarantee the limit value in the auxiliary game so does he in the original game.

Technically, in the auxiliary game it will be convenient to assume that the initial state is initially chosen according to a probability measure with finite support on $X$. We denote by $\Delta_f(X)$ the set of such measures, and one may think of $\Delta_f(X)$ as a slightly extended state space for the auxiliary game. To express an informational gap, we use the Choquet order of *sweeping* of probability measures :

---

1. The idea of considering an auxiliary stochastic game is certainly not new, see for example Mertens 1986, Coulomb, 2003 or Mertens *et al.*, 1994, Part A, Ch IV, section 3.

given $u$ and $v$ in $\Delta_f(X)$, we say that $u$ is *better* than $v$, or $v$ is a sweeping of $u$, if for every concave continuous mapping $f$ from $X$ to the reals, $u(f) \geq v(f)$. And it is essentially possible to model the original informational advantage of player 1 with a "splitting hypothesis" defined via this order. For the topological part, we use the weak* topology on the set $\Delta(X)$ of Borel probability measures on $X$, and more precisely the Wasserstein distance. This allows to carefully control the Lipschitz constant of the value functions.

In order to prove that in the auxiliary game player 1 can guarantee the limit value, a dynamic programming problem will be derived by letting player 2 play a best reply at any period. Consequently the role of player 2 disappears, and this is indeed possible because we assumed that in the original repeated game the transitions are controlled by player 1 and not by player 2. Technically, the set of states of the dynamic programming problem is $Z = \Delta_f(X) \times [0, 1]$. $\Delta_f(X)$ is dense in $\Delta(X)$ for the weak* topology, so $Z$ can be viewed as a precompact metric space. We define, for every $m$ and $n$, a value $w_{m,n}$ as the supremum payoff player 1 can achieve when his payoff is defined as the *minimum*, for $t$ in $\{1, ..., n\}$, of his average rewards computed between stages $m+1$ and $m+t$. One can prove that the family $(w_{m,n})$ is uniformly equicontinuous, and together with the precompactness of the state space, this implies the existence of the uniform value for the the dynamic programming problem. The proof of this implication can be found in a companion paper (Corollary 3.8, Renault 2011), which only deals with 1-player games and can be read independently.

To complete the discussion, it remains to describe what can be guaranteed by player 2 in the original and auxiliary games. Luckily enough, this is an easier aspect of the proof since we assumed that the transitions are controlled by player 1, hence player 2 can not make any irreversible, or long term, mistake. Precisely, any $m$-stage strategy of player 2 can be extended into an $n + m$-stage strategy which is optimal for player 2 in the game with payoff defined as the expected average payoff between stages $m + 1$ and $m + n$.

The paper is organized as follows. We start in the next section with a simple illustrative example. In section 3, we consider a particular class of stochastic games including our auxiliary games of level 2. We think that this class of games is interesting in itself. It is defined with hypotheses making no reference to the original finite set $K$, and we prefer to start by presenting this class, which can be considered as both more general and simpler to study than the auxiliary stochastic games. We prove that these games have a uniform value using the result on dynamic programming proved in Renault, 2011. In section 4, we consider our original repeated game and show how the existence of the uniform value in such a game derives from the existence of the uniform value for the stochastic games of section 3. Finally, we obtain formulas expressing the uniform value in terms of the values of some finite games. More precisely, let $v_{m,n}$ be the value of the game where the global payoff is defined as the average of the payoffs between stage $m + 1$ and stage $m + n$. We show in particular that $\inf_{n \geq 1} \sup_{m \geq 0} v_{m,n} = \sup_{m \geq 0} \inf_{n \geq 1} v_{m,n}$,

and that this quantity is the uniform value (see subsection 4.3.1). We conclude by discussing several hypotheses and present a few open problems.

# 2  A simple example

We consider the following repeated game with an informed controller. The set of states is $K = \{k_1, k_2\}$, with initial distribution the uniform probability $p^* = (1/2, 1/2)$, the set of pure actions of player 1 is $I = \{T, B\}$ and the set of pure actions of player 2 is $J = \{L, R\}$. The payoffs for player 1 and the transitions are the following :

$$
\begin{array}{cc}
\begin{array}{c} \\ T \\ B \end{array}
\begin{array}{c}
\begin{array}{cc} L & R \end{array} \\
\left( \begin{array}{cc} 1_{\circlearrowleft} & 0_{\circlearrowleft} \\ 0_{p*} & 0_{p*} \end{array} \right) \\
\text{state } k_1
\end{array}
&
\begin{array}{c} \\ T \\ B \end{array}
\begin{array}{c}
\begin{array}{cc} L & R \end{array} \\
\left( \begin{array}{cc} 0_{p*} & 0_{p*} \\ 0_{p*} & 1_{p*} \end{array} \right) \\
\text{state } k_2
\end{array}
\end{array}
$$

For example, if the state is $k_1$ and $(T, L)$ is played then player 1's payoff is 1 and the state remains the same, that is $k_1$. If the state is $k_2$ and $(B, L)$ is played, then the payoff to player 1 is 0 and the next state is chosen according to $p^*$. To conclude the description of the game, it remains to describe what is observed by the players. The set of signals of player 1 is $C = K \times I$, and the set of signals of player 2 is $D = I$. At the beginning of every stage, the action previously played by player 1 is publicly announced, and in addition player 1 learns the new state. Note that the action of player 2 is not announced to player 1, and the new state is not announced to player 2.

Let us be very clear about the progress of the game. Initially, the first state $k_1$ is selected according to $p^*$ and announced to player 1, but not to player 2. Then simultaneously player 1 chooses $i_1$ in $I$ and player 2 chooses $j_1$ in $J$. The payoff for player 1 and the new state $k_2$ are determined by the above matrices, and before playing stage 2 the action $i_1$ is publicly announced, whereas $k_2$ is only told to player 1. Then at any stage $t \geq 2$, the new state $k_t$ is selected according to the state $k_{t-1}$ and the actions $(i_{t-1}, j_{t-1})$ of stage $t - 1$, then player 1 learns the signal $c_t = (k_t, i_{t-1})$ and player 2 learns the signal $d_t = i_{t-1}$. Simultaneously, player 1 chooses $i_t$ in $I$ and player 2 chooses $j_t$ in $J$, payoffs are determined (but not observed) and the play proceeds to stage $t + 1$. All the above description of the game is known by the players.

Notice that player 1 is informed, in the sense that he can always deduce the true state and player 2's signal from his own signal, and that player 1 controls the transition, in the sense that the law of the pair (new state, signal for player 2) does not depend on the action of player 2. We will consider in section 4 all games satisfying these two conditions, see the hypotheses **HA'** and **HB'** in subsection 4.1. Having in mind a large number of stages, we are interested in the following questions. What is the long term value ? How should the players behave ? Is it

important for the players to know precisely the number of stages ?

We now describe the auxiliary stochastic game of level 2 associated to our example. The set of states is $X = \Delta(K) = \{p = (p^1, p^2) \in I\!R_+^2, p^1 + p^2 = 1\}$ and a state here corresponds to the belief of player 2 on $K$ in the original game. The set of actions of player 1 is $A = \Delta(I)^K = \{a = (a^1, a^2) \in [0,1]^2\}$, with $a^1$ and $a^2$ being the respective probabilities to play $T$ in state $k_1$ and $k_2$. In the auxiliary stochastic game, player 2 observes after each stage the action in $A$ played by player 1 : this is a main difference with the original game. The set of actions of player 2 is $B = \Delta(J) = \{b \in [0,1]\}$, with $b$ being the probability to play $L$. The payoff to player 1 in the auxiliary game is given by the following mapping $g$ from $X \times A \times B$ to $[0,1]$ :

$$g(p,a,b) = p^1 a^1 b + p^2 (1-a^2)(1-b).$$

Suppose that in the original game player 2's belief on $K$ is given by $p$, and that player 1 plays according to some $a$ in $A$. Then with probability $(p^1 a^1 + p^2 a^2)$ player 1 plays $T$ and the belief of player 2 on the next state becomes $\left( \frac{p^1 a^1 + 1/2 p^2 a^2}{p^1 a^1 + p^2 a^2}, \frac{1/2 p^2 a^2}{p^1 a^1 + p^2 a^2} \right)$, whereas with the remaining probability player 1 plays $B$ and the belief of player 2 on the next state is $p^*$. Consequently, the transition in the auxiliary game is defined by the mapping $l$ from $X \times A \times B$ to the set $\Delta_f(X)$ of probability measures with finite support on $X$ such that ($\delta_p \in \Delta_f(X)$ denotes the Dirac measure on $p$) :

$$l(p,a,b) = \left( p^1 a^1 + p^2 a^2 \right) \delta_{\left( \frac{p^1 a^1 + 1/2 p^2 a^2}{p^1 a^1 + p^2 a^2}, \frac{1/2 p^2 a^2}{p^1 a^1 + p^2 a^2} \right)} + \left( p^1 (1-a^1) + p^2 (1-a^2) \right) \delta_{p^*}.$$

Notice that $l(p,a,b)$ does not depend on $b$, i.e. the transition is controlled by player 1 (more than this, one can check that all the following hypotheses H1, ..., H7 of section 3 are satisfied here). Note also that the state variable $p$ always satisfies $p^1 \geq 1/2$, so that $p^2 \leq p^1$.

The last element for describing the auxiliary game is the initial distribution on states in $X$, it is here $u = \delta_{p^*}$ in $\Delta_f(X)$. This game is played as follows : at any stage $t$, the state $p_t$ is selected according to $l(p_{t-1}, a_{t-1}, b_{t-1})$ (according to $u$ if $t = 1$) and announced to both players. Simultaneously, player 1 chooses $a_t$ in $A$ and player 2 chooses $b_t$ in $B$. The stage payoffs are $g(p_t, a_t, b_t)$ for player 1 and the opposite for player 2. Then $a_t$ and $b_t$ are publicly announced, and the play proceeds to stage $t + 1$.

From the auxiliary stochastic game, it is then possible to define a dynamic programming problem of level 3 by assuming that player 2 plays a payoff best-reply at every stage, i.e. chooses $L$ if $p^1 a^1 < p^2 (1-a^2)$ and $R$ if $p^1 a^1 > p^2 (1-a^2)$. The study of this dynamic programming problem can be achieved (in this case, through the Average Cost Optimality Equation, for example as in Hörner *et al.* 2010) and yields the following results :

The value of the auxiliary stochastic game (and of the dynamic programming problem) is 1/3, and an optimal stationary strategy for player 1 is given by playing

$a^2 = 0$, i.e always $B$ in state $k_2$, and $a^1 = p^2/p^1$, i.e. play $T$ in state $k_1$ with probability $p^2/p^1$ if the state is $(p^1, p^2)$. This strategy defines a simple Markov chain with two states on $X$, the states being $p^* = (1/2, 1/2)$ and $\hat{p} = (1, 0)$. The payoffs are $1/2$ in $p^*$ and $0$ in $\hat{p}$, and the state moves with probability one from $\hat{p}$ to $p^*$, and from $p^*$ the next state is determined with uniform probability. Consequently, the unique invariant measure is $2/3\,\delta_{p^*} + 1/3\,\delta_{\hat{p}}$ and the strategy guarantees player 1 a payoff of $(2/3)(1/2) + (1/3)(0) = 1/3$.

What is an optimal strategy for player 2 in the auxiliary stochastic game? If player 2 plays always $R$, he guarantees $1/2$ because player 1 can play $B$ at every stage. If player 2 plays $R$ if $p^1 > 1/2$ and $L$ if $p = p^*$, then player 1 can reply by playing $T$ if $p = p^*$ and $B$ if $p^1 > 1/2$, and a simple computation of the induced invariant measure yields an expected payoff of $3/8$ for player 1, which is more than $1/3$ and hence too large. We now prove that a stationary optimal strategy for player 2 is here given by : play $R$ if $p^1 > 1/2$, and play $2/3L + 1/3R$ if $p = p^*$. Fix this stationary strategy for player 2, and consider the Markov decision process (or MDP) faced by player 1. Without loss of generality we assume that after player 1 has played $T$, the belief of player 2 on the next state will satisfy $p^1 > 1/2$ and player 2 plays $R$, whereas after player 1 has played $B$, the belief of player 2 on the next state is $p^*$ and player 2 plays $2/3L + 1/3R$. Hence, player 1 faces an MDP (which is *not* the dynamic programming problem of level 3) with four states : $\alpha^1 = (k_1, T)$, $\alpha^2 = (k_2, T)$, $\alpha^3 = (k_1, B)$, $\alpha^4 = (k_2, B)$, each state $\alpha$ including the current state in $K$ and the last action played by player 1. Player 1 has two actions $T$ and $B$. Probability transitions (p) and rewards (r) are given in the figure below, where only positive rewards are written.
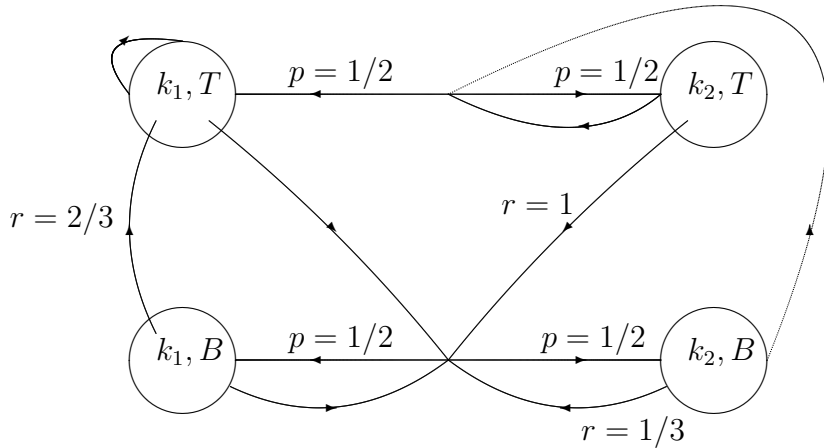


For example, in $\alpha^1 = (k_1, T)$ player 2 will play $R$, hence player 1 can either play $T$, have a reward of 0 and stay in $\alpha^1 = (k_1, T)$, or player 1 can play $B$ which will also give him a reward of 0 but the new state will be $\alpha^3 = (k_1, B)$ or $\alpha^4 = (k_2, B)$ with even probability. Starting from $\alpha^4 = (k_2, B)$, player 2 will play $2/3L + 1/3R$, hence player 1 can either play $T$, have a reward of 0 and go to $\alpha^1$ and $\alpha^2$ with even probability, or player 1 can play $B$, have a (expected) reward of $1/3$ and go

to $\alpha^3$ and $\alpha^4$ with even probability. One can check that in this MDP, player 1 has a stationary optimal strategy given by : play $B$ in $\alpha^1$, $\alpha^2$ and $\alpha^4$, and play $T$ in $\alpha^3$. The average reward of this strategy is $1/3$, which implies that the strategy (play $R$ if $p^1 > 1/2$, and $2/3L + 1/3R$ if $p = p^*$) is indeed optimal for player 2 in the auxiliary game.

What about the original repeated game? Our optimal strategy of player 1 in the auxiliary game naturally induces (see proposition 4.23 later) an optimal strategy in the original game : player 1 plays $T$ if the state is $k_1$ and his last action was $B$, and plays $B$ otherwise, i.e. if the state is $k_2$, or if the state is $k_1$ and the last action of player 1 was $T$. Regarding player 2, we have here a specific and simple example because player 2 can roughly determine if his "belief" is $p^*$ or not, *without* knowing the strategy of player 1. Consequently, an optimal strategy of player 2 in the original game is given by : play $R$ if the last action of player 1 was $T$, and play $(2/3L + 1/3T)$ if the last action of player 1 was $B$. In general, it is not possible to deduce an optimal strategy for player 2 in the original game from an optimal strategy of this player in the auxiliary stochastic game.

**Remark 2.1.** The above game belongs to the following family of repeated games with an informed controller. Keeping the same payoff functions, we define for every parameter $\alpha$, $\beta$, $\gamma$, $\delta$ in $[0, 1]$ the repeated game with the same set of states $K = \{k_1, k_2\}$, the same initial distribution $p^* = (1/2, 1/2)$, the same sets of pure actions for the players and the following transitions :

$$
\begin{array}{c}
\begin{array}{cc} L & R \end{array} \\
\begin{array}{c} T \\ B \end{array}
\begin{pmatrix} 1_{(\alpha,1-\alpha)} & 0_{(\alpha,1-\alpha)} \\ 0_{(\beta,1-\beta)} & 0_{(\beta,1-\beta)} \end{pmatrix} \\
\text{state } k_1
\end{array}
\qquad
\begin{array}{c}
\begin{array}{cc} L & R \end{array} \\
\begin{array}{c} T \\ B \end{array}
\begin{pmatrix} 0_{(\gamma,1-\gamma)} & 0_{(\gamma,1-\gamma)} \\ 0_{(\delta,1-\delta)} & 1_{(\delta,1-\delta)} \end{pmatrix} \\
\text{state } k_2
\end{array}
$$

For example, if the state is $k_1$ and $(T, L)$ is played, then player 1's payoff is 0, and the next state is $k_1$ with probability $\alpha$ and $k_2$ with probability $1 - \alpha$. The above example corresponds to the case : $\alpha = 1$, $\beta = \gamma = \delta = 1/2$. The case $\alpha = \beta = 1 - \gamma = 1 - \delta = 1$ corresponds to a standard repeated game with lack of information on one side introduced by Aumann and Maschler (1995) in the sixties. They proved that the value is $1/4$ in this case. The cases where $\alpha = \beta$ and $\gamma = \delta$ correspond to Markov chain games with lack of information on one side, as introduced in Renault 2006 : the sequence of states follows an exogeneous Markov chain uncontrolled by the players. The symmetric subcase is when $\alpha = \beta = 1 - \gamma = 1 - \delta$, and in this case Hörner *et al.* (2010) proved that for $\alpha$ in $[1/2, 2/3]$ the value is $\frac{\alpha}{4\alpha-1}$. The value is difficult to compute in general, and it is still an open problem in many cases, e.g. $\alpha = \beta = 1 - \gamma = 1 - \delta = 0.9$

# 3 A certain class of stochastic games

## 3.1 Model

We consider in this section 2-player zero-sum stochastic games with complete information and standard observation, *played with pure strategies*. We assume that after each stage, the new state is selected according to a probability with finite support.

If $X$ is a non empty set, we denote by $\Delta_f(X)$ the set of probabilities on $X$ with finite support.

We consider :

- three non empty sets : a set of states $X$, a set $A$ of actions for player 1, and a set $B$ of actions for player 2,
- an element $u$ in $\Delta_f(X)$, called the initial distribution on states,
- a mapping $g$ from $X \times A \times B$ to $[0,1]$, called the payoff function of player 1, and
- a mapping $l$ from $X \times A \times B$ to $\Delta_f(X)$, called the transition function.

The interpretation is the following. The initial state $p_1$ in $X$ is selected according to $u$, and is announced to both players. Then simultaneously, player 1 chooses $a_1$ in $A$, and player 2 chooses $b_1$ in $B$. The stage payoff is $g(p_1, a_1, b_1)$ for player 1, and $-g(p_1, a_1, b_1)$ for player 2, then $a_1$ and $b_1$ are publicly observed, and a new state $p_2$ is selected according to $l(p_1, a_1, b_1)$, etc... At any stage $t \geq 2$, the state $p_t$ is selected according to $l(p_{t-1}, a_{t-1}, b_{t-1})$, and announced to both players. Simultaneously, player 1 chooses $a_t$ in $A$ and player 2 chooses $b_t$ in $B$. The stage payoffs are $g(p_t, a_t, b_t)$ for player 1 and the opposite for player 2. Then $a_t$ and $b_t$ are publicly announced, and the play proceeds to stage $t + 1$.

From now on we fix $\Gamma = (X, A, B, g, l)$, and for every $u$ in $\Delta_f(X)$ we denote by $\Gamma(u) = (X, A, B, g, l, u)$ the corresponding stochastic game induced by $u$. For the moment we make no assumption on $\Gamma$. We start with elementary definitions and notations.

A strategy for player 1 is a sequence $\sigma = (\sigma_n)_{n \geq 1}$, where for each $n$, $\sigma_n$ is a mapping from $(X \times A \times B)^{n-1} \times X$ to $A$, with the interpretation that $\sigma_n(p_1, a_1, b_1, ..., p_{n-1}, a_{n-1}, b_{n-1}, p_n)$ is the action prescribed by player 1 at stage $n$ after $(p_1, a_1, b_1, ..., p_{n-1}, a_{n-1}, b_{n-1}, p_n)$ occurred. $\sigma_1$ simply is a mapping from $X$ to $A$ giving the first action played by player 1 depending on the initial state. Similarly, a strategy for player 2 is a sequence $\tau = (\tau_n)_{n \geq 1}$, where for each $n$, $\tau_n$ is a mapping from $(X \times A \times B)^{n-1} \times X$ to $B$. We denote by $\Sigma$ and $\mathcal{T}$ the sets of strategies of player 1 and player 2, respectively.

Fix for a while $(u, \sigma, \tau)$, and assume that player 1 plays $\sigma$ whereas player 2 plays $\tau$ in the game $\Gamma(u)$. The initial state $p_1$ is selected according to $u$, then the first actions are $a_1 = \sigma_1(p_1)$ and $b_1 = \tau_1(p_1)$. $p_2$ is selected according to $l(p_1, a_1, b_1)$, then $a_2 = \sigma_2(p_1, a_1, b_1, p_2)$, $b_2 = \tau_2(p_1, a_1, b_1, p_2)$, etc... By induction

this defines, for every positive $N$, a probability measure with finite support on the set $(X \times A \times B)^N$ corresponding to the set of the first $N$ states and actions. It is standard that these probabilities can be uniquely extended to a probability $\mathbb{P}_{u,\sigma,\tau}$ on the set of plays $\Omega = (X \times A \times B)^\infty$, endowed with the $\sigma$-algebra generated by the cylinders (one can apply, e.g., theorem 2.7.2. p.109 in Ash, 1972). Expectations with respect to this probability are written $\mathbb{E}_{u,\sigma,\tau}$.

**Definition 3.1.** *The expected average payoff for player 1 induced by $(\sigma, \tau)$ over the first $N$ stages in the game $\Gamma(u)$ is denoted by :*

$$\gamma_N^u(\sigma, \tau) = \mathbb{E}_{u,\sigma,\tau} \left( \frac{1}{N} \sum_{n=1}^{N} g(p_n, a_n, b_n) \right).$$

**Definition 3.2.** *For $u$ in $\Delta_f(X)$ and $N \geq 1$, the game $\Gamma_N(u)$ is the zero-sum game with normal form $(\Sigma, \mathcal{T}, \gamma_N^u)$.*

$\Gamma_N(u)$ is called the $N$-stage game with initial distribution $u$. This is the normal form game where player 1's strategy set is $\Sigma$, player 2's strategy set is $\mathcal{T}$, and $\gamma_N^u$ is the payoff function for player 1. It has a value if : $\sup_{\sigma \in \Sigma} \inf_{\tau \in \mathcal{T}} \gamma_N^u(\sigma, \tau) = \inf_{\tau \in \mathcal{T}} \sup_{\sigma \in \Sigma} \gamma_N^u(\sigma, \tau)$. A strategy $\sigma$ achieving the supremum on the LHS (if any) is then called an optimal strategy for player 1. Similarly, a strategy $\tau$ achieving the infimum on the RHS (if any) is then called an optimal strategy for player 2.

**Notations 3.3.**
*For $p$ in $X$, we denote by $\delta_p \in \Delta_f(X)$ the Dirac measure on $p$. A probability $u$ in $\Delta_f(X)$ is written $u = \sum_{p \in X} u(p)\delta_p$, where $u(p)$ is the probability of $p$ under $u$.*

*For $u$ in $\Delta_f(X)$, if $\Gamma_N(u)$ has a value, we denote it by $\tilde{v}_N(u) \in [0, 1]$. For $p$ in $X$, if $\Gamma_N(\delta_p)$ has a value, we denote it by $v_N(p)$ and we have $v_N(p) = \tilde{v}_N(\delta_p)$.*

Notice that whenever $X$ is convex and $p \neq p'$, we have $\delta_{1/2 p + 1/2 p'} \neq 1/2\, \delta_p + 1/2\, \delta_{p'}$, so we will *not* identify a state $p$ with the measure $\delta_p$. When the value of the $N$-stage game exists for every initial distribution, $\tilde{v}_N$ is a mapping from $\Delta_f(X)$ to $\mathbb{R}$, whereas $v_N$ is a mapping from $X$ to $\mathbb{R}$. It is easy to see that : $\forall u \in \Delta_f(X), \forall N \geq 1, \forall \sigma \in \Sigma, \forall \tau \in \mathcal{T}$,

$$\mathbb{P}_{u,\sigma,\tau} = \sum_{p \in X} u(p)\mathbb{P}_{\delta_p,\sigma,\tau} \text{ and } \gamma_N^u(\sigma, \tau) = \sum_{p \in X} u(p)\gamma_N^{\delta_p}(\sigma, \tau).$$

**Claim 3.4.** *If $v_N(p)$ exists for each $p$ in $X$, then $\tilde{v}_N(u)$ exists for every $u$ in $\Delta_f(X)$ and $\tilde{v}_N(u) = \sum_{p \in X} u(p)v_N(p)$.*

We now consider an infinite time horizon.

**Definition 3.5.** *Let $u$ be in $\Delta_f(X)$.*

*The lower (or maxmin) value of $\Gamma(u)$ is : $\underline{v}(u) = \sup_{\sigma \in \Sigma} \liminf_n \left( \inf_{\tau \in \mathcal{T}} \gamma_n^u(\sigma, \tau) \right).$*

*The upper (or minmax) value of $\Gamma(u)$ is :* $\overline{v}(u) = \inf_{\tau \in \mathcal{T}} \limsup_n \left( \sup_{\sigma \in \Sigma} \gamma_n^u(\sigma, \tau) \right).$

$\underline{v}(u) \leq \overline{v}(u)$. $\Gamma(u)$ *is said to have a uniform value if and only if* $\underline{v}(u) = \overline{v}(u)$, *and in this case the uniform value is* $\underline{v}(u) = \overline{v}(u)$.

An equivalent definition of the uniform value is as follows. Given a real number $v$, we say that player 1 can guarantee $v$ in $\Gamma(u)$ if for any $\varepsilon > 0$, player 1 has a strategy that gives him an expected average payoff greater than $v - \varepsilon$ in any game with sufficiently many stages : $\forall \varepsilon > 0, \exists \sigma \in \Sigma, \exists N_0, \forall N \geq N_0, \forall \tau \in \mathcal{T}, \gamma_N^u(\sigma, \tau) \geq v - \varepsilon$. Similarly, player 2 can guarantee $v$ in $\Gamma(u)$ if for any $\varepsilon > 0$, he has a strategy that gives player 1 an expected average payoff lower than $v + \varepsilon$ in any game with sufficiently many stages : $\forall \varepsilon > 0, \exists \tau \in \mathcal{T}, \exists N_0, \forall N \geq N_0, \forall \sigma \in \Sigma, \gamma_N^u(\sigma, \tau) \leq v + \varepsilon$. If player 1 can guarantee $v$ and player 2 can guarantee $w$ then clearly $v \leq w$. We also have :

**Claim 3.6.** $\underline{v}(u) = \max\{v \in \mathbb{R}, \text{player 1 can guarantee } v \text{ in } \Gamma(u) \}$,
$\qquad\quad \overline{v}(u) = \min\{v \in \mathbb{R}, \text{player 2 can guarantee } v \text{ in } \Gamma(u) \}$.
$v$ *can be guaranteed by both players if and only if* $v$ *is the uniform value of* $\Gamma(u)$.

**Definition 3.7.** *Whenever the uniform value $v$ of $\Gamma(u)$ exists, a strategy $\sigma$ of player 1 satisfying :* $\exists N_0, \forall N \geq N_0, \forall \tau \in \mathcal{T}, \gamma_N^u(\sigma, \tau) \geq v - \varepsilon$ *is called $\varepsilon$-optimal, and similarly a strategy $\tau$ of player 2 such that :* $\exists N_0, \forall N \geq N_0, \forall \sigma \in \Sigma, \gamma_N^u(\sigma, \tau) \leq v + \varepsilon$ *is called $\varepsilon$-optimal. A 0-optimal strategy is simply called optimal.*

Assume now that $\tilde{v}_N(u)$ exists for each $N$. If player 1 (resp. player 2) can guarantee $v$ then $\liminf_N \tilde{v}_N(u) \geq v$ (resp. $\limsup_N \tilde{v}_N(u) \leq v$). As a consequence we have the following :

**Claim 3.8.** *Assume that $\tilde{v}_N(u)$ exists for each $N$.*

$$\underline{v}(u) \leq \liminf_N \tilde{v}_N(u) \leq \limsup_N \tilde{v}_N(u) \leq \overline{v}(u).$$

So the existence of the uniform value $\underline{v}(u) = \overline{v}(u)$ implies the existence of the "limit value" $\lim_N \tilde{v}_N(u)$, and whenever the uniform value exists, all notions coincide : $\underline{v}(u) = \overline{v}(u) = \lim_N \tilde{v}_N(u)$.

## 3.2 An existence result for the uniform value

We define now some hypotheses on $\Gamma$ that will ensure the existence of the uniform value for each initial distribution $u$.

**Remark 3.9.** In view of the games of section 4, we have in mind repeated games where player 1 is informed and controls the transition (such as the game of section 2). In these games, there is an underlying finite set of parameters $K$, and $X$ is

the set of probabilities over $K$. Initially, a parameter $k$ is selected according to $p$, and is announced to player 1 only. Then the parameter may change from stage to stage, but it is always known by player 1 and its evolution is independent of player 2's actions. It will be possible to check the following hypotheses H1 to H7 in this model. Our point here is to be more general and simpler : we want to write a model without any reference to the underlying set of parameters, and where players use pure strategies.

We first make the important assumption that player 1 controls the transitions :

**Hypothesis H1** : the transition $l$ does not depend on player 2's actions, i.e. $\forall p \in X, \forall a \in A, \forall b \in B, \forall b' \in B, l(p, a, b) = l(p, a, b')$.

In the sequel we consider $l$ as a mapping from $X \times A$ to $\Delta_f(X)$, and we write $l(p, a)$ for the distribution on the next state if the actual state is $p$ and player 1 plays $a$.

**Hypothesis H2** : $X$ is a compact convex subset of a normed vector space.

We denote by $\Delta(X)$ the set of Borel probability measures on $X$, and we will use the weak* topology and the Choquet order on $\Delta(X)$. $\Delta_f(X)$ is now seen as a subset of $\Delta(X)$. We first fix notations and recall some definitions. We start with the topological aspect : $X$ is in particular a compact metric space, and we denote by $d(p, q)$ the distance between two elements $p$ and $q$ of $X$.

**Notations 3.10.** *We denote by $C(X)$ the set of continuous mappings from $X$ to the reals, and by $Lip_1(X)$ the set of non expansive, i.e. Lipschitz with constant 1, elements of $C(X)$. For $u$ in $\Delta(X)$ and $f$ in $C(X)$ we write $u(f) = \int_{p \in X} f(p) du(p)$. Given $f$ in $C(X)$, we extend $f$ by duality to an affine mapping $\tilde{f} : \Delta(X) \longrightarrow \mathrm{I\!R}$ by $\tilde{f}(u) = u(f)$.*

In the following, $\Delta(X)$ will *always* be endowed with the weak* topology : a sequence $(u_n)_n$ converges to $u$ in $\Delta(X)$ if and only if $u_n(f) \longrightarrow_{n \to \infty} u(f)$ for every $f$ in $C(X)$. $\Delta(X)$ is itself compact, the weak* topology can be metrized, and the set $\Delta_f(X)$ of probabilities on $X$ with finite support is dense in $\Delta(X)$ (see for example Doob, 1994, Ch.VIII, section 5, and Malliavin, 1995, p.99).

**Remark 3.11.** An important distance on $\Delta(X)$ which metrizes the weak* topology is the following (Fortet-Mourier-)Wasserstein distance, defined by :

$$\forall u \in \Delta(X), \forall v \in \Delta(X), \quad d(u, v) = \sup_{f \in Lip_1(X)} |u(f) - v(f)|.$$

One can check that this distance has the following nice properties. For every $p$, $q$ in $X$, $d(p, q) = d(\delta_p, \delta_q)$. Moreover, for $f$ in $C(X)$ and $C \geq 0$, $f$ is $C$-Lipschitz if and only if $\tilde{f}$ is $C$-Lipschitz.

We will also use the convexity of $X$. In zero-sum games with lack of information on the side of player 2, it is well known that the value is a concave function of the parameter $p$ : this fundamental property represents the advantage for player 1 to be informed (see for example, Sorin 2002, proposition 2.2 p. 16). In our setup, we want the initial distribution $\delta_{1/2p+1/2p'}$ to be more advantageous for player 1 than the initial distribution $1/2\,\delta_p + 1/2\,\delta_{p'}$. This is perfectly represented by the following order :

**Definition 3.12.** *For $u$ and $v$ in $\Delta(X)$, we say that $u$ is better than $v$, or that $v$ is a sweeping of $u$, and we write $u \succeq v$, if :*
$$\text{for every concave mapping } f \text{ in } C(X),\ u(f) \geq v(f).$$

This order was introduced by Choquet[2](1960). It is actually an order on $\Delta(X)$, the maximal elements are the Dirac measures, and Choquet proved that the minimal elements are the measures with support in the set of extreme points of $X$ (see P.A. Meyer, 1966, theorem 24 p. 282). For every $f$ in $C(X)$, we easily have the equivalence :

**Claim 3.13.** *$f$ is concave if and only if $\tilde{f}$ is non decreasing.*

We now define hypotheses H3 to H7.

**Hypothesis H3** : $A$ and $B$ are compact convex subsets of topological vector spaces.

**Hypothesis H4** : For every $(p,b)$ in $X \times B$, $(a \longrightarrow g(p,a,b))$ is concave and upper semi-continuous. For every $(p,a)$ in $X \times A$, $(b \longrightarrow g(p,a,b))$ is convex and lower semi-continuous.

We will prove in the sequel a natural dynamic programming principle *a la* Shapley (or Bellmann).

**Notation 3.14.** *For $f$ in $C(X)$ and $\alpha$ in $[0,1]$, we define $\Phi(\alpha, f) : X \longrightarrow \mathbb{R}$ with :*

$$\forall p \in X,\quad \Phi(\alpha, f)(p) = \sup_{a \in A} \inf_{b \in B} \left( \alpha\, g(p,a,b) + (1-\alpha)\, \tilde{f}(l(p,a)) \right).$$

**Hypothesis H5** : There exists a subset $\mathcal{D}$ of $Lip_1(X)$ containing $\Phi(1,0)$, and such that $\Phi(\alpha, f) \in \mathcal{D}$ for every $f$ in $\mathcal{D}$ and $\alpha$ in $[0,1]$.

We state the following hypothesis in full generality, but we will assume H1 so that $l(p,a,b)$ will not depend on $b$.

---

2. For convenience, we reverse here Choquet's order, i.e. we write $u \succeq v$ instead of $u \preceq v$. For $u$ and $v$ in $\Delta_f(X)$, a simple characterization of $u \succeq v$ will be stated later, see proposition 3.34.

**Hypothesis H6** : For every $(p, b)$ in $X \times B$, $(a \mapsto l(p, a, b))$ is continuous and concave.

**Hypothesis H7** : "Splitting" Consider a convex combination $p = \sum_{s=1}^{S} \lambda_s p_s$ in the set of states $X$, and a family of actions $(a_s)_{s \in S}$ in $A^S$. Then there exists $a$ in $A$ such that :

$$l(p, a) \succeq \sum_{s \in S} \lambda_s l(p_s, a_s) \quad \text{and} \quad \inf_{b \in B} g(p, a, b) \geq \sum_{s \in S} \lambda_s \inf_{b \in B} g(p_s, a_s, b).$$

H3 and H4 are standard and, by Sion's theorem, will lead to the existence of the value of the stage game. H5 is very important and will ensure that all value functions are 1-Lipschitz. We will provide later a simple condition implying H5, see remark 3.43. H6 is the unique hypothesis where the topology on $\Delta(X)$ appears, and does not depend on a particular distance metrizing the weak* topology. H7 is the generalization of the well known splitting lemma for games with lack of information on one side. Under the hypotheses H1,..., H7, our main result in theorem 3.17 will be the existence of the uniform value. We will also obtain several other properties, which are expressed via the following notions.

**Definition 3.15.** *A strategy $\sigma = (\sigma_t)_{t \geq 1}$ of player 1 is Markov if for each stage $t$, $\sigma_t$ only depends on the current state $p_t$. A Markov strategy for player 1 will be seen as a sequence $\sigma = (\sigma_t)_{t \geq 1}$, where for each $t$ $\sigma_t$ is a mapping from $X$ to $A$ giving the action to be played on stage $t$ depending on the current state. We denote the set of Markov strategies for player 1 by : $\Sigma^M = \{\sigma = (\sigma_t)_{t \geq 1}, \text{with } \forall t, \sigma_t : X \longrightarrow A\}$. The set of Markov strategies for player 2 is defined similarly and is denoted by $\mathcal{T}^M$.*

**Definition 3.16.** *For $m \geq 0$ and $n \geq 1$, the expected average payoff for player 1 induced by a strategy pair $(\sigma, \tau)$ in $\Sigma \times \mathcal{T}$ from stage $m + 1$ to stage $m + n$ is denoted by :*

$$\gamma_{m,n}^u(\sigma, \tau) = I\!E_{u, \sigma, \tau} \left( \frac{1}{n} \sum_{t=m+1}^{m+n} g(p_t, a_t, b_t) \right).$$

**Theorem 3.17.** *Assume that H1,..., H7 hold.*
*Then for every initial distribution $u$, the game $\Gamma(u)$ has a uniform value $v^*(u)$.*

*In $\Gamma(u)$, every player can guarantee $v^*(u)$ with a Markov strategy, i.e. for each positive $\varepsilon$ each player has a Markov strategy which is $\varepsilon$-optimal.*

$$We\ have : v^*(u) = \inf_{n \geq 1} \sup_{m \geq 0} \tilde{v}_{m,n}(u) = \sup_{m \geq 0} \inf_{n \geq 1} \tilde{v}_{m,n}(u)$$
$$= \inf_{n \geq 1} \sup_{m \geq 0} w_{m,n}(u) = \sup_{m \geq 0} \inf_{n \geq 1} w_{m,n}(u),$$

*where $\tilde{v}_{m,n}(u) = \sup_{\sigma \in \Sigma} \inf_{\tau \in \mathcal{T}} \gamma_{m,n}^u(\sigma, \tau) = \inf_{\tau \in \mathcal{T}} \sup_{\sigma \in \Sigma} \gamma_{m,n}^u(\sigma, \tau)$, and $w_{m,n}(u) = \sup_{\sigma \in \Sigma} \inf_{\tau \in \mathcal{T}} \min_{t \in \{1, \dots, n\}} \gamma_{m,t}^u(\sigma, \tau)$.*
*For every $m$ and $n$, $\tilde{v}_{m,n}$ and $w_{m,n}$ are non expansive, and $(v_n)_n$ uniformly converges to $v^*$.*

13

## 3.3 Proof of Theorem 3.17

We now prove theorem 3.17, and assume that H1,..., H7 hold. In the proof, we endow $\Delta(X)$ with the Wasserstein distance. We denote by $\mathbb{N}^*$ the set of positive integers. By H3 and H4, for every $(p, a) \in X \times A$, the infimum is achieved in $\inf_{b \in B} g(p, a, b)$, and we will simply write $g(p, a)$ for $\min_{b \in B} g(p, a, b)$. For each $p$, $(a \longrightarrow g(p, a))$ is still concave and upper semi-continuous.

**Lemma 3.18.** *For every concave $f$ in $C(X)$ and $\alpha$ in $[0, 1]$, $\Phi(\alpha, f)$ is concave.*

**Proof :** Fix a convex combination $p = \sum_{s=1}^{S} \lambda_s p_s$ in $X$, and consider for each $s$ an element $a_s$ in $A$. By the splitting hypothesis H7, one can find $a$ in $A$ such that $l(p, a) \succeq \sum_{s \in S} \lambda_s l(p_s, a_s)$ and $g(p, a) \geq \sum_{s \in S} \lambda_s g(p_s, a_s)$. $f$ is concave so $\tilde{f}$ is non decreasing and $\tilde{f}(l(p, a)) \geq \sum_{s \in S} \lambda_s \tilde{f}(l(p_s, a_s))$. We obtain :

$$
\begin{aligned}
\Phi(\alpha, f)(p) &\geq \alpha \sum_{s \in S} \lambda_s g(p_s, a_s) + (1 - \alpha) \sum_{s \in S} \lambda_s \tilde{f}(l(p_s, a_s)), \\
&= \sum_{s \in S} \lambda_s \left( \alpha g(p_s, a_s) + (1 - \alpha) \tilde{f}(l(p_s, a_s)) \right).
\end{aligned}
$$

This holds for every $(a_s)_{s \in S}$, so $\Phi(\alpha, f)(p) \geq \sum_{s \in S} \lambda_s \Phi(\alpha, f)(p_s)$. $\qquad\square$

The proof is divided into 4 parts. In subsection 3.3.1, we study the value of the games with finite horizon and derive an appropriate recursive formula. In subsection 3.3.2 we introduce a specific function $v^*$ and show that Player 2 can guarantee $v^*(u)$ in the game $\Gamma(u)$. The rest of the proof shows that also Player 1 can guarantee $v^*(u)$ in the game $\Gamma(u)$. Subsection 3.3.3 is technical and shows that there is no harm in restricting player 1 to use Markov strategies and assuming that player 2 plays a best-reply at every stage. Moreover, it introduces a family of mappings $(w_{m,n})$ (see theorem 3.17 for one expression of $w_{m,n}$) and shows that each of these mappings is non decreasing and non expansive. Finally, subsection 3.3.4 introduces an equivalent dynamic programming problem where the role of player 2 has disappeared. We use a result from Renault, 2011 to solve the dynamic programming problem and conclude the proof of theorem 3.17.

### 3.3.1 Value of finite games and the recursive formula.

**Lemma 3.19.** *For every state $p$ in $X$, the game $\Gamma_1(\delta_p)$ has a value which is :*

$$
v_1(p) = \max_{a \in A} \min_{b \in B} g(p, a, b) = \min_{b \in B} \max_{a \in A} g(p, a, b) = \Phi(1, 0)(p).
$$

*$v_1$ is concave and belongs to $\mathcal{D}$. $\tilde{v}_1$ is non decreasing and non expansive.*

**Proof :** Fix $p$ in $X$, and consider the game with normal form $(A, B, g(p, ., .))$. By H3 and H4, we can apply Sion's theorem (see e.g. Sorin 2002 p.156, thm A.7) and obtain that this game has a value and both players have optimal strategies.

14

By lemma 3.18, we get that $v_1$ is concave, and by H5, we have $v_1 \in \mathcal{D}$. By claim 3.4, for every distribution $u$ the game $\Gamma_1(u)$ has a value which is precisely $\tilde{v}_1$. By concavity of $v_1$, $\tilde{v}_1$ is non decreasing. Since $v_1 \in Lip_1(X)$ and we use the Wasserstein distance, $\tilde{v}_1$ is non expansive. $\qquad\square$

We will need to consider not only the $n$-stage games $\Gamma_n(u)$, but a larger family of games with initial distribution $u$.

**Definition 3.20.** *Let* $\theta = \sum_{t \geq 1} \theta_t \delta_t$ *be in* $\Delta_f(\mathbb{N}^*)$, *i.e.* $\theta$ *is a probability with finite support over positive integers. For $u$ in* $\Delta_f(X)$, *the game* $\Gamma_{[\theta]}(u)$ *is the game with normal form* $(\Sigma, \mathcal{T}, \gamma_{[\theta]}^u)$, *where :*

$$\gamma_{[\theta]}^u(\sigma, \tau) = \mathbb{E}_{u, \sigma, \tau} \left( \sum_{t=1}^{\infty} \theta_t \ g(p_t, a_t, b_t) \right).$$

If $\theta = 1/n \sum_{t=1}^{n} \delta_t$, $\Gamma_{[\theta]}(u)$ is nothing but $\Gamma_n(u)$. $\Gamma_{[\theta]}(u)$ can be seen as the game where after the play, a stage $t^*$ is selected according to $\theta$ and then only the payoff of stage $t^*$ matters. If $\theta = \sum_{t \geq 1} \theta_t \delta_t$, define $\theta^+$ as the law of $t^* - 1$ given that $t^* \geq 2$. Define arbitrarily $\theta^+ = \theta$ if $\theta_1 = 1$, and otherwise we have $\theta^+ = \sum_{t \geq 1} \frac{\theta_{t+1}}{1 - \theta_1} \delta_t$. We now write a recursive formula for the value of the games $\Gamma_{[\theta]}(u)$.

**Proposition 3.21.** *For* $\theta = \sum_{t \geq 1} \theta_t \delta_t$ *in* $\Delta_f(\mathbb{N}^*)$ *and $u$ in* $\Delta_f(X)$, *the game* $\Gamma_{[\theta]}(u)$ *has a value* $\tilde{v}_{[\theta]}(u)$ *such that :*

$$
\begin{aligned}
\forall p \in X, \quad v_{[\theta]}(p) &= \Phi(\theta_1, v_{[\theta^+]})(p), \\
&= \max_{a \in A} \theta_1 g(p, a) + (1 - \theta_1) \tilde{v}_{[\theta^+]}(l(p, a)), \\
&= \min_{b \in B} \max_{a \in A} \theta_1 g(p, a, b) + (1 - \theta_1) \tilde{v}_{[\theta^+]}(l(p, a)).
\end{aligned}
$$

*In* $\Gamma_{[\theta]}(u)$, *both players have optimal Markov strategies.* $v_{[\theta]}$ *is concave and belongs to* $\mathcal{D}$. $\tilde{v}_{[\theta]}$ *is non decreasing and non expansive.*

**Proof :** by induction. If $\theta = \delta_1$, lemma 3.19 gives the result.

Fix now $n \geq 2$, and assume that the proposition is true for every $\theta$ with support included in $\{1, ..., n-1\}$. Fix a probability $\theta = \sum_{t=1}^{n} \theta_t \delta_t$, and notice that $\theta^+$ has a support included in $\{1, ..., n-1\}$. Fix also $p$ in $X$.

Consider the auxiliary zero-sum game $\Gamma_{[\theta]}'(p)$ with normal form $(A, B, f_{[\theta]}^p)$, where $f_{[\theta]}^p(a, b) = \theta_1 g(p, a, b) + (1 - \theta_1) \tilde{v}_{[\theta^+]}(l(p, a))$. We will apply Sion's theorem to this game. By H3, $A$ and $B$ are compact convex subsets of topological vector spaces. For every $a$, $(b \mapsto f_{[\theta]}^p(a, b))$ is convex l.s.c. by H4. Consider now a convex combination $\lambda a + (1 - \lambda) a'$ in $A$. By H6, we have $l(p, \lambda a + (1 - \lambda) a') \succeq \lambda l(p, a) + (1 - \lambda) l(p, a')$. By the induction hypothesis, $\tilde{v}_{[\theta^+]}$ is non decreasing, so :

$$
\begin{aligned}
\tilde{v}_{[\theta^+]} \left( l(p, \lambda a + (1 - \lambda) a') \right) &\geq \tilde{v}_{[\theta^+]} \left( \lambda l(p, a) + (1 - \lambda) l(p, a') \right), \\
&= \lambda \tilde{v}_{[\theta^+]}(l(p, a)) + (1 - \lambda) \tilde{v}_{[\theta^+]}(l(p, a')).
\end{aligned}
$$

Since $g$ is concave in $a$, we obtain that $f_{[\theta]}^p$ is concave in $a$. Regarding continuity, by H6 and the induction hypothesis, $(a \longrightarrow \tilde{v}_{[\theta^+]}(l(p,a)))$ is continuous. By H3, $(a \mapsto g(p,a,b))$ is u.s.c., so $(a \mapsto f_{[\theta]}^p(a,b))$ is u.s.c.. By Sion's theorem $\Gamma'_{[\theta]}(p)$ has a value which is :

$$
\begin{aligned}
v'_{[\theta]}(p) &= \max_{a \in A} \min_{b \in B} \theta_1 g(p,a,b) + (1-\theta_1)\tilde{v}_{[\theta^+]}(l(p,a)) \\
&= \min_{b \in B} \max_{a \in A} \theta_1 g(p,a,b) + (1-\theta_1)\tilde{v}_{[\theta^+]}(l(p,a)).
\end{aligned}
$$

Consider now the original game $\Gamma_{[\theta]}(p)$, and a strategy pair $(\sigma, \tau)$ in $\Sigma \times \mathcal{T}$. Write $a = \sigma_1(p)$, resp. $b = \tau_1(p)$, for the first action played by player 1, resp. player 2, in $\Gamma_{[\theta]}(p)$. Denote by $\sigma_{p,a,b}^+$ the continuation strategy issued from $\sigma$ after $(p,a,b)$ has occurred at stage 1. $\sigma_{p,a,b}^+$ belongs to $\Sigma$, and plays at stage $n$ after $(p_1,a_1,b_1,...,p_n)$ what $\sigma$ plays at stage $n+1$ after $(p,a,b,p_1,a_1,b_1,...,p_n)$. Similarly denote by $\tau_{p,a,b}^+$ the continuation strategy issued from $\tau$ after $(p,a,b)$ has occurred at stage 1. It is easy to check that :

$$
\gamma_{[\theta]}^p(\sigma,\tau) = \theta_1 g(p,a,b) + (1-\theta_1)\gamma_{[\theta^+]}^{l(p,a)}(\sigma_{p,a,b}^+, \tau_{p,a,b}^+).
$$

Consequently, in the game $\Gamma_{[\theta]}(p)$ player 1 can guarantee $\max_{a \in A} \min_{b \in B} \theta_1 g(p,a,b)+ (1-\theta_1)\tilde{v}_{[\theta^+]}(l(p,a))$ by playing a Markov strategy. Similarly player 2 has a Markov strategy which guarantees $\min_{b \in B} \max_{a \in A} \theta_1 g(p,a,b) + (1-\theta_1)\tilde{v}_{[\theta^+]}(l(p,a))$. Since the two quantities coincide, $\Gamma_{[\theta]}(p)$ has a value $v_{[\theta]}(p) = v'_{[\theta]}(p)$, and both players have Markov optimal strategies.

This implies that for every $u$ in $\Delta_f(X)$, the game $\Gamma_{[\theta]}(u)$ has a value which is the affine extension $\tilde{v}_{[\theta]}(u)$, and both players have Markov optimal strategies in $\Gamma_{[\theta]}(u)$. $v_{[\theta]} = \Phi(\theta_1, v_{[\theta^+]})$ and $v_{[\theta^+]}$ is concave, so by lemma 3.18 $v_{[\theta]}$ is concave, and hence, by claim 3.13, $\tilde{v}_{[\theta]}$ is non decreasing. By H5 $v_{[\theta]}$ is in $\mathcal{D}$, so $v_{[\theta]}$ is 1-Lipschitz, and finally $\tilde{v}_{[\theta]}$ is non expansive. $\qquad \square$

Among the games $\Gamma_{[\theta]}(u)$, the following family will play an important role and deserves a specific notation.

**Definition 3.22.** *For $m \geq 0$ and $n \geq 1$, $\Gamma_{m,n}(u)$ is the game with normal form $(\Sigma, \mathcal{T}, \gamma_{m,n}^u)$.*

Recall that in definition 3.16, we put $\gamma_{m,n}^u(\sigma,\tau) = \mathbb{E}_{u,\sigma,\tau}\left(\frac{1}{n}\sum_{t=m+1}^{m+n} g(p_t,a_t,b_t)\right)$ for each $(\sigma, \tau)$. So $\Gamma_{m,n}(u)$ is nothing but $\Gamma_{[\theta]}(u)$ with $\theta = 1/n \sum_{t=m+1}^{m+n} \delta_t$. We can apply the previous proposition and denote the value of $\Gamma_{m,n}(u)$ by $\tilde{v}_{m,n}(u)$. $v_{0,n}$ is just the value of the $n$-stage game $v_n$, and for convenience we put $v_0 = 0$. We have for all $p$ in $X$, and positive $m$ and $n$ :

$$
\begin{aligned}
v_n(p) &= \Phi(1/n, v_{n-1}) = \frac{1}{n} \max_{a \in A} \min_{b \in B} \left(g(p,a,b) + (n-1)\tilde{v}_{n-1}(l(p,a))\right), \\
&= \frac{1}{n} \min_{b \in B} \max_{a \in A} \left(g(p,a,b) + (n-1)\tilde{v}_{n-1}(l(p,a))\right), \\
v_{m,n}(p) &= \Phi(0, v_{m-1,n}) = \max_{a \in A} \tilde{v}_{m-1,n}(l(p,a)).
\end{aligned}
$$

16

In $\Gamma_{m,n}(u)$, the players first play $m$ stages to control the state, then they play $n$ stages for payoffs. Moreover, player 2 does not control the transitions, so he can play arbitrarily in the first $m$ stages. The next lemma formalizes the following idea : if $\tau$ is an optimal strategy for player 2 in the $n$-stage game for each initial state, then any strategy that coincides with $\tau$ between stages $m+1$ and $m+n$ is optimal in the game $\Gamma_{m,n}(u)$ for each $u$.

**Lemma 3.23.** *Fix $n \geq 1$. There exists a Markov strategy $\tau = (\tau_t)_{t \geq 1}$ for player 2 such that $\forall m \geq 0$, $\forall \tau' = (\tau'_t)_{t \geq 1}$ in $\mathcal{T}$ :*

*the condition $(\forall t = 1, ..., n, \ \forall p \in X, \tau'_{m+t}(., ..., p) = \tau_t(p))$ implies that for every $u$ in $\Delta_f(X)$, $\tau'$ is an optimal strategy for player 2 in $\Gamma_{m,n}(u)$.*

**Proof :** For each $t$ in $\{1, ..., n\}$, define $\tau_t$ as the mapping which plays, if the current state is $p \in X$, an element $b$ in $B$ achieving the minimum in :

$$\min_{b \in B} \ \max_{a \in A} \frac{1}{n - t + 1} \left( g(p, a, b) + (n - t)\tilde{v}_{n-t}(l(p, a)) \right) = v_{n+1-t}(p).$$

Using the previous recursive formula, one can show by induction that this construction of $\tau$ is appropriate. $\qquad\square$

### 3.3.2 Player 2 can guarantee $\inf_{n \geq 1} \sup_{m \geq 0} \tilde{v}_{m,n}(u)$ in $\Gamma(u)$.

We now consider the game with infinitely many stages $\Gamma(u)$. The following results are similar to propositions 7.7 and 7.8 in Renault, 2006.

**Proposition 3.24.** *In $\Gamma(u)$, player 2 can guarantee with Markov strategies the quantity :*

$$\inf_{n \geq 1} \limsup_T \left( \frac{1}{T} \sum_{t=0}^{T-1} \tilde{v}_{nt,n}(u) \right).$$

**Proof :** Fix $n \geq 1$, and consider $\tau_1, ..., \tau_n$ given by lemma 3.23. Divide the set of stages $\mathbb{N}^*$ into consecutive blocks $B^1, B^2, ..., B^m, ...$ of equal length $n$. By lemma 3.23, the cyclic strategy $\tau' = (\tau_1, ..., \tau_n, \tau_1, ..., \tau_n, \tau_1, ..., \tau_n, ....)$ is optimal for player 2 in the game $\Gamma_{nm,n}(u)$, for each $m \geq 0$. $\tau'$ is a Markov strategy for player 2, and for every strategy $\sigma$ of player 1 in $\Sigma$ we have :

$$\forall m \geq 0, \ \mathbb{E}_{u,\sigma,\tau'} \left( \frac{1}{n} \sum_{t \in B^{m+1}} g(p_t, a_t, b_t) \right) \leq \tilde{v}_{nm,n}(u),$$

$$\text{so} \ \forall M \geq 1, \ \mathbb{E}_{u,\sigma,\tau'} \left( \frac{1}{nM} \sum_{t=1}^{nM} g(p_t, a_t, b_t) \right) \leq \frac{1}{M} \sum_{m=0}^{M-1} \tilde{v}_{nm,n}(u).$$

And since $n$ is fixed and payoffs are bounded, we obtain that player 2 can guarantee with Markov strategies : $\limsup_M \left( \frac{1}{M} \sum_{m=0}^{M-1} \tilde{v}_{nm,n}(u) \right)$ in $\Gamma(u)$. $\qquad\square$

It is actually possible to prove a slightly stronger result, not needed for the proof of theorem 3.17 but implying in the end the existence of a 0-optimal strategy for player 2, see remark 3.42. Anyway, the proof of proposition 3.24 also shows the following inequality.

**Lemma 3.25.** $\forall u \in \Delta_f(X), \forall n \geq 1, \forall T \geq 1, \quad \tilde{v}_{nT}(u) \leq \frac{1}{T} \sum_{t=0}^{T-1} \tilde{v}_{nt,n}(u).$

The following quantity will turn out to be the value of $\Gamma(u)$.

**Notation 3.26.** *For every $u$ in $\Delta_f(X)$, we put :*

$$v^*(u) = \inf_{n \geq 1} \sup_{m \geq 0} \tilde{v}_{m,n}(u).$$

Since $v^*(u) \geq \inf_{n \geq 1} \limsup_T \left( \frac{1}{T} \sum_{t=0}^{T-1} \tilde{v}_{nt,n}(u) \right)$, by proposition 3.24 player 2 can also guarantee $v^*(u)$ with Markov strategies in $\Gamma(u)$. By claim 3.6, $\overline{v}(u) = \min\{v \in \mathbb{R}, \text{player 2 guarantees } v \text{ in } \Gamma(u) \}$, so we now have the following inequality chain :

$$\boxed{\underline{v}(u) \leq \liminf_N \tilde{v}_N(u) \leq \limsup_N \tilde{v}_N(u) \leq \overline{v}(u) \leq v^*(u).}$$

### 3.3.3 Markov strategies for player 1.

By H1 player 2 does not control the transition, so a Markov strategy $\sigma$ induces, together with the initial distribution $u$, a probability distribution $\mathbb{P}_{u,\sigma}$ over $(X \times A)^\infty$, i.e. over sequences of states and actions for player 1. For $u$ in $\Delta_f(X)$ and $\sigma_1 : X \longrightarrow A$, we denote by $H(u, \sigma_1)$ the law of the state of stage 2 if the initial distribution is $u$ and player 1 plays at stage 1 according to $\sigma_1$. We denote by $G(u, \sigma_1)$ the payoff guaranteed by $\sigma_1$ at stage 1. And we also define the continuation strategy $\sigma^+$.

**Notations 3.27.**
$G(u, \sigma_1) = \sum_{p \in X} u(p) g(p, \sigma_1(p))$, *and* $H(u, \sigma_1) = \sum_{p \in X} u(p) l(p, \sigma_1(p)) \in \Delta_f(X)$.
*If $\sigma = (\sigma_t)_{t \geq 1}$ is in $\Sigma^M$, we write $\sigma^+$ for the Markov strategy $(\sigma_t)_{t \geq 2}$.*

We now concentrate on what player 1 can achieve in $\Gamma(u)$ and completely forget player 2. We use similar notations as in definition 3.20.

**Definition 3.28.** *For $\theta = \sum_{t \geq 1} \theta_t \delta_t$ in $\Delta_f(\mathbb{N}^*)$, $u$ in $\Delta_f(X)$ and $\sigma$ in $\Sigma^M$, we put :*

$$\gamma_{[\theta]}^u(\sigma) = \mathbb{E}_{u,\sigma} \left( \sum_{t \geq 1} \theta_t \, g(p_t, a_t) \right) = \sum_{t \geq 1} \theta_t \, \gamma_{[\delta_t]}^u(\sigma).$$

For simplicity, we write $\gamma_{[t]}$ instead of $\gamma_{[\delta_t]}$ for the payoff induced at stage $t$. Clearly, we have for every $t \geq 2$ :

$$\gamma_{[t]}^u(\sigma) = \gamma_{[t-1]}^{H(u,\sigma_1)}(\sigma^+).$$

**Lemma 3.29.** $\gamma_{[\theta]}^u(\sigma) = \min_{\tau \in \mathcal{T}} \gamma_{[\theta]}^u(\sigma, \tau).$

18

The proof is easy, the minimum on the RHS being achieved by a Markov strategy $\tau$ such that for every $t$ and $p_t$, $\tau_t(p_t)$ achieves the minimum in $b$ of the quantity $g(p_t, \sigma_t(p_t), b)$. As a corollary of lemma 3.29, we obtain that $\sup_{\sigma \in \Sigma^M} \gamma^u_{[\theta]}(\sigma) = \sup_{\sigma \in \Sigma^M} \inf_{\tau \in \mathcal{T}} \gamma^u_{[\theta]}(\sigma, \tau)$. By proposition 3.21, $\Gamma_{[\theta]}(u)$ has a value, so we get :

**Corollary 3.30.**
*For every $\theta$ in $\Delta_f(I\!N^*)$ and $u$ in $\Delta_f(X)$,* $\quad \sup_{\sigma \in \Sigma^M} \gamma^u_{[\theta]}(\sigma) = \tilde{v}_{[\theta]}(u)$.

As in definition 3.22, we now specify notations for a particular class of probabilities.

**Definition 3.31.** *For $n \geq 1$, $m \geq 0$, $u$ in $\Delta_f(X)$ and $\sigma$ in $\Sigma^M$, we put :*

$$\gamma^u_{m,n}(\sigma) = I\!E_{u,\sigma} \frac{1}{n} \left( \sum_{t=1}^n g(p_{m+t}, a_{m+t}) \right) = \frac{1}{n} \sum_{t=1}^n \gamma^u_{[m+t]}(\sigma).$$

Finally, we consider a situation where player 1 does not precisely know the length of the game.

**Definition 3.32.** *Fix $m \geq 0$, $n \geq 1$, and $u$ in $\Delta_f(X)$. We define :*

$$w_{m,n}(u) = \sup_{\sigma \in \Sigma^M} \min\{\gamma^u_{m,t}(\sigma), t \in \{1, .., n\}\}.$$

The mappings $w_{m,n}$ will play an important role in the sequel, while applying corollary 3.8 of Renault, 2011. We will show in corollary 3.39 that they are non expansive. To prove corollary 3.39 we will use the following lemma 3.35 and propositions 3.34, 3.36 and 3.37. We start by defining an auxiliary game [3].

**Definition 3.33.** *For $m \geq 0$, $n \geq 1$, and $u$ in $\Delta_f(X)$, we define $\mathcal{A}(m, n, u)$ as the zero-sum game with normal form $(\Sigma^M, \Delta(\{1, ..., n\}), f)$, where :*

$$\forall \sigma \in \Sigma^M, \forall \theta \in \Delta(\{1, ..., n\}), \ f(\sigma, \theta) = \sum_{t=1}^n \theta_t \gamma^u_{m,t}(\sigma).$$

In $\mathcal{A}(m, n, u)$ player 1 chooses a Markov strategy and an extra player 2 chooses which stage payoffs will be considered. We will prove later that $\mathcal{A}(m, n, u)$ has a value which is $w_{m,n}(u)$ (see proposition 3.37 below). Notice that in general $f(., \theta)$ is not concave in $\sigma$. However, we will show that it is *concave-like* in $\sigma$, i.e. that : $\forall \sigma', \sigma'', \forall \lambda \in [0, 1]$, there exists $\sigma$ such that $\forall \theta, f(\sigma, \theta) \geq \lambda f(\sigma', \theta) + (1 - \lambda) f(\sigma'', \theta)$. We start with a characterization for the partial order $\succeq$.

**Proposition 3.34.** *Let $u$ and $v$ be in $\Delta_f(X)$. Write $u = \sum_{p \in X} u(p) \delta_p$. The following conditions are equivalent :*
    *(i) $u \succeq v$, and*

---

3. We proceed similarly as in section 6.2. of Renault, 2011. However the situation is more technical here, and it will not be possible to apply a standard minmax theorem to the game $\mathcal{A}(m, n, u)$.

*(ii) For every $p$ such that $u(p) > 0$, there exist $S(p) \geq 1$, $\lambda_1^p, ..., \lambda_{S(p)}^p \geq 0$ and $q_1^p, ..., q_{S(p)}^p$ in $X$ such that : $\sum_{s=1}^{S(p)} \lambda_s^p = 1$, $\sum_{s=1}^{S(p)} \lambda_s^p q_s^p = p$, and $v = \sum_{p \in X} u(p) \left( \sum_{s=1}^{S(p)} \lambda_s^p \delta_{q_s^p} \right)$.*

The proof can be easily deduced from a theorem of Loomis (see Meyer, 1966, T26 p.283), which deals with positive measures on $X$. Notice that condition $(ii)$ can be seen as follows : $u$ is the law of some random variable $X_1$ with values in $X$, $v$ is the law of some random variable $X_2$ with values in $X$, and we have the martingale condition : $\mathbb{E}(X_2|X_1) = X_1$.

In general, $\gamma_{[t]}^u(\sigma)$ is not a non decreasing function of $u$. However, we have the following property.

**Lemma 3.35.** *Let $n \geq 1$, $u$ and $v$ be in $\Delta_f(X)$ such that $u \succeq v$. For every $\sigma \in \Sigma^M$, there exists $\sigma' \in \Sigma^M$ such that :*

$$\forall t \in \{1, ..., n\}, \ \gamma_{[t]}^u(\sigma') \geq \gamma_{[t]}^v(\sigma).$$

**Proof :** by induction on $n$.

If $n = 1$, there exists $\sigma'$ such that $\gamma_{[1]}^u(\sigma') = \tilde{v}_1(u)$. Since $\tilde{v}_1$ is non decreasing, $\tilde{v}_1(u) \geq \tilde{v}_1(v) \geq \gamma_{[1]}^v(\sigma)$.

Fix now $n \geq 1$, and assume that the lemma is proved for $n$. Fix $u$ and $v$ in $\Delta_f(X)$ with $u \succeq v$, and $\sigma$ in $\Sigma^M$. We have $u = \sum_{p \in X} u(p)\delta_p$, and by proposition 3.34 it is possible to write $v = \sum_{p \in X} u(p) \left( \sum_{s=1}^{S(p)} \lambda_s^p \delta_{q_s^p} \right)$, with $\lambda_s^p \geq 0$, $\sum_{s=1}^{S(p)} \lambda_s^p = 1$, and $\sum_{s=1}^{S(p)} \lambda_s^p q_s^p = p$ for each $p$ such that $u(p) > 0$. Define for every such $p$ and $s$, $a_s^p = \sigma_1(q_s^p)$. By the splitting hypothesis H7, for every $p$ one can find $a^p \in A$ such that :

$$l(p, a^p) \succeq \sum_{s=1}^{S(p)} \lambda_s^p \, l(q_s^p, a_s^p) \quad \text{and} \quad g(p, a^p) \geq \sum_{s=1}^{S(p)} \lambda_s^p \, g(q_s^p, a_s^p).$$

We define what $\sigma'$ plays at stage 1 if the state is $p$ as : $\sigma_1'(p) = a^p$.

We have : $\gamma_{[1]}^u(\sigma') = \sum_p u(p)g(p, a^p) \geq \sum_p u(p) \sum_{s=1}^{S(p)} \lambda_s^p g(q_s^p, a_s^p) = \gamma_{[1]}^v(\sigma)$, and $H(u, \sigma_1') = \sum_p u(p)l(p, a^p) \succeq \sum_p u(p) \sum_{s=1}^{S(p)} \lambda_s^p \, l(q_s^p, a_s^p) = H(v, \sigma_1)$.

Since $H(u, \sigma_1') \succeq H(v, \sigma_1)$, we apply the induction hypothesis to the continuation strategy $\sigma+$. We obtain the existence of some Markov strategy $\tau = (\tau_t)_{t \geq 1}$ such that : $\forall t \in \{1, ..., n\}$, $\gamma_{[t]}^{H(u,\sigma_1')}(\tau) \geq \gamma_{[t]}^{H(v,\sigma_1)}(\sigma^+)$. Define $\sigma_t' = \tau_{t-1}$ for each $t \geq 2$. $\sigma' = (\sigma_t')_{t \geq 1}$ is a Markov strategy for player 1, and satisfies : $\gamma_{[1]}^u(\sigma') \geq \gamma_{[1]}^v(\sigma)$, and for $t \in \{2, ..., n+1\}$ :

$$\gamma_{[t]}^u(\sigma') = \gamma_{[t-1]}^{H(u,\sigma_1')}(\sigma'^+) \geq \gamma_{[t-1]}^{H(v,\sigma_1)}(\sigma^+) = \gamma_{[t]}^v(\sigma). \qquad \square$$

Lemma 3.35 is now improved.

**Proposition 3.36.** *Let $n \geq 1$, $\lambda \in [0,1]$, $u$, $u'$ and $u''$ in $\Delta_f(X)$ be such that $u \succeq \lambda u' + (1-\lambda)u''$. For every $\sigma'$ and $\sigma''$ in $\Sigma^M$, there exists $\sigma \in \Sigma^M$ such that :*

$$\forall t \in \{1, ..., n\}, \ \gamma_{[t]}^u(\sigma) \geq \lambda \gamma_{[t]}^{u'}(\sigma') + (1-\lambda)\gamma_{[t]}^{u''}(\sigma'').$$

**Proof :** by induction on $n$.

If $n = 1$, there exists $\sigma$ such that $\gamma_{[1]}^u(\sigma) = \tilde{v}_1(u) \geq \tilde{v}_1(\lambda u' + (1-\lambda)u'')$
$= \lambda \tilde{v}_1(u') + (1-\lambda)\tilde{v}_1(u'') \geq \lambda \gamma_{[1]}^{u'}(\sigma') + (1-\lambda)\gamma_{[1]}^{u''}(\sigma'')$.

Assume the proposition is proved for some $n \geq 1$, and fix $u$, $u'$, $u''$, $\lambda$, $\sigma'$ and $\sigma''$ with $u \succeq \lambda u' + (1-\lambda)u''$. Write $v = \lambda u' + (1-\lambda)u''$. By lemma 3.35 it is enough to find $\sigma$ in $\Sigma^M$ such that : $\forall t \in \{1, ..., n+1\}, \ \gamma_{[t]}^v(\sigma) \geq \lambda \gamma_{[t]}^{u'}(\sigma') + (1-\lambda)\gamma_{[t]}^{u''}(\sigma'')$.

We have $v = \sum_p \left(\lambda u'(p) + (1-\lambda)u''(p)\right)\delta_p$, and $v(p) = \lambda u'(p) + (1-\lambda)u''(p)$ for each $p$. For every $p$ such that $v(p) > 0$, we define :

$$\sigma_1(p) = \frac{\lambda u'(p)}{v(p)}\sigma_1'(p) + \frac{(1-\lambda)u''(p)}{v(p)}\sigma_1''(p).$$

$\sigma_1(p)$ belongs to $A$ by convexity. Now,

$$
\begin{aligned}
\lambda \gamma_{[1]}^{u'}(\sigma') + (1-\lambda)\gamma_{[1]}^{u''}(\sigma'') &= \lambda \sum_p u'(p)g(p, \sigma_1'(p)) + (1-\lambda)\sum_p u''(p)g(p, \sigma_1''(p)) \\
&= \sum_p v(p)\left(\frac{\lambda u'(p)}{v(p)}g(p, \sigma_1'(p)) + \frac{(1-\lambda)u''(p)}{v(p)}g(p, \sigma_1''(p))\right) \\
&\leq \sum_p v(p)\, g(p, \sigma_1(p)) = \gamma_{[1]}^v(\sigma),
\end{aligned}
$$

where the inequality comes from the concavity of $g$ in the variable $a$ (see H4). Proceeding in the same way with distributions on the second state, we obtain via the concavity of $l$ in $a$ (see H6) :

$$
\begin{aligned}
\lambda H(u', \sigma_1') + (1-\lambda)H(u'', \sigma_1'') &= \lambda \sum_p u'(p)l(p, \sigma_1'(p)) + (1-\lambda)\sum_p u''(p)l(p, \sigma_1''(p)) \\
&= \sum_p v(p)\left(\frac{\lambda u'(p)}{v(p)}l(p, \sigma_1'(p)) + \frac{(1-\lambda)u''(p)}{v(p)}l(p, \sigma_1''(p))\right) \\
&\preceq \sum_p v(p)\, l(p, \sigma_1(p)) = H(v, \sigma_1)
\end{aligned}
$$

Consequently, we have $H(v, \sigma_1) \succeq \lambda H(u', \sigma_1') + (1-\lambda)H(u'', \sigma_1'')$, and by the induction hypothesis there exists $\sigma^+$ in $\Sigma^M$ such that $\forall t \in \{1, ..., n\}, \ \gamma_{[t]}^{H(v,\sigma_1)}(\sigma^+) \geq \lambda \gamma_{[t]}^{H(u',\sigma_1')}(\sigma'^+) + (1-\lambda)\gamma_{[t]}^{H(u'',\sigma_1'')}(\sigma''^+)$. We naturally define $\sigma = (\sigma_1, \sigma^+)$, and we have, for $t \in \{2, ..., n+1\}$ : $\gamma_{[t]}^v(\sigma) = \gamma_{[t-1]}^{H(v,\sigma_1)}(\sigma^+) \geq \lambda \gamma_{[t-1]}^{H(u',\sigma_1')}(\sigma'^+) + (1-\lambda)\gamma_{[t-1]}^{H(u'',\sigma_1'')}(\sigma''^+) = \lambda \gamma_{[t]}^{u'}(\sigma') + (1-\lambda)\gamma_{[t]}^{u''}(\sigma'')$. $\qquad\square$

**Proposition 3.37.** *For every $m \geq 0$, $n \geq 1$ and $u$ in $\Delta_f(X)$, the game $\mathcal{A}(m, n, u)$ has a value which is $w_{m,n}(u)$.*

**Proof :** Recall that the payoff function in $\mathcal{A}(m, n, u)$ is : $f(\sigma, \theta) = \sum_{t=1}^{n} \theta_t \gamma_{m,t}^u(\sigma)$ for every $\sigma$ in $\Sigma^M$ and $\theta$ in $\Delta(\{1, ..., n\})$.

$\Delta(\{1, ..., n\})$ is convex and compact, and $f$ is affine continuous in $\theta$. We now show that $f$ is concave-like in $\sigma$. Let $\sigma', \sigma''$ be in $\Sigma^M$, and $\lambda \in [0, 1]$. By the previous proposition, there exists $\sigma$ in $\Sigma^M$ such that : $\forall t \in \{1, ..., m + n\}$, $\gamma_{[t]}^u(\sigma) \geq \lambda \gamma_{[t]}^u(\sigma') + (1 - \lambda) \gamma_{[t]}^u(\sigma'')$. So $f(\sigma, \theta) = \sum_{t=1}^{n} \frac{\theta_t}{t} \sum_{t'=1}^{t} \gamma_{[m+t']}^u(\sigma) \geq \sum_{t=1}^{n} \frac{\theta_t}{t} \sum_{t'=1}^{t} \left( \lambda \gamma_{[m+t']}^u(\sigma') + (1 - \lambda) \gamma_{[m+t']}^u(\sigma'') \right) = \lambda f(\sigma', \theta) + (1 - \lambda) f(\sigma'', \theta)$. By a theorem of Fan (1953, see proposition A.13 p.160 in Sorin 2002), $\mathcal{A}(m, n, u)$ has a value which is : $\sup_{\sigma \in \Sigma^M} \inf_{\theta \in \Delta(\{1, ..., n\})} f(\sigma, \theta) = w_{m,n}(u)$. □

**Notation 3.38.** *For $\theta = \sum_{t=1}^{n} \theta_t \delta_t \in \Delta(\{1, ..., n\})$, and $m \geq 0$, we define $\theta^{m,n}$ in $\Delta(\{1, ..., m + n\})$ by :*
$$\theta_s^{m,n} = 0 \text{ if } s \leq m, \text{ and } \theta_s^{m,n} = \sum_{t=s-m}^{n} \frac{\theta_t}{t} \text{ if } m < s \leq n + m.$$

**Corollary 3.39.** *For every $m \geq 0$, $n \geq 1$ and $u$ in $\Delta_f(X)$,*

$$w_{m,n}(u) = \inf_{\theta \in \Delta(\{1, ..., n\})} \tilde{v}_{[\theta^{m,n}]}(u).$$

*The mapping $w_{m,n}$ is non decreasing and non expansive.*

**Proof :** Fix $m$, $n$ and $u$. $w_{m,n}(u)$ is the value of $\mathcal{A}(m, n, u)$, so we have : $w_{m,n}(u) = \inf_{\theta \in \Delta(\{1, ..., n\})} \sup_{\sigma \in \Sigma^M} f(\sigma, \theta)$. But $f(\sigma, \theta) = \sum_{t=1}^{n} \theta_t \gamma_{m,t}^u(\sigma) = \gamma_{[\theta^{m,n}]}^u(\sigma)$. So we obtain : $w_{m,n}(u) = \inf_{\theta \in \Delta(\{1, ..., n\})} \sup_{\sigma \in \Sigma^M} \gamma_{[\theta^{m,n}]}^u(\sigma)$. Corollary 3.30 gives $w_{m,n}(u) = \inf_{\theta \in \Delta(\{1, ..., n\})} \tilde{v}_{[\theta^{m,n}]}(u)$. For each $\theta$, $\tilde{v}_{[\theta^{m,n}]}$ is non decreasing and non expansive by proposition 3.21, hence the result. □

### 3.3.4 A dynamic programming problem.

We will conclude the proof of theorem 3.17 and show that the uniform value exists. Fix the initial distribution $u$, our stochastic game is defined by $\Gamma(u) = (X, A, B, g, l, u)$. We now define an auxiliary Markov Decision Process as follows.

**Definition 3.40.** *The MDP $\Psi(z_0)$ associated to $\Gamma(u)$ is defined as $\Psi(z_0) = (Z, F, r, z_0)$, where :*

• *$Z = \Delta_f(X) \times [0, 1]$ is the set of states of the MDP,*

• *$z_0 = (u, 0)$ is the initial state in $Z$,*

• *$r$ is the payoff function from $Z$ to $[0, 1]$ defined by $r(u, y) = y$ for each $(u, y)$ in $Z$,*

• *and the transition function $F$ is the correspondence from $Z$ to $Z$ such that :*
$\forall z = (u, y) \in Z$,

$$F(z) = \left\{ \left( \sum_{p \in X} u(p) l(p, a(p)), \sum_{p \in X} u(p) g(p, a(p)) \right), \forall p \; a(p) \in A \right\},$$
$$= \{(H(u, f), G(u, f)), f \text{ is a mapping from } X \text{ to } A\}.$$

Notice that $F(u, y)$ does not depend on $y$, hence the value functions will not depend on $y$. $F$ has non empty values. Even with strong assumptions on $l$ and $g$, $F$ may not have a compact graph, because in the definition of $F(z)$ we have a unique $a(p)$ for each $p$. So even if $q$ is close to $p$, the image by $F$ of $(1/2\delta_p + 1/2\delta_q, 0)$ may be quite larger than $F(\delta_p, 0)$.

As in Renault, 2011, we denote by $S(z_0) = \{s = (z_1, ..., z_t, ...) \in Z^\infty, \forall t \geq 1, z_t \in F(z_{t-1})\}$ the set of plays at $z_0$. The next proposition shows the strong links between the stochastic game $\Gamma(u)$ and the MDP $\Psi(z_0)$.

**Proposition 3.41.** *a) For every Markov strategy $\sigma$ in $\Sigma^M$, there exists a play $s = (z_1, ..., z_t, ...) \in S(z_0)$ such that $\forall t \geq 1$, $\gamma_{[t]}^u(\sigma) = r(z_t)$.*

*b) Reciprocally, for every play $s = (z_1, ..., z_t, ...) \in S(z_0)$, there exists a Markov strategy $\sigma$ in $\Sigma^M$ such that : $\forall t \geq 1$, $\gamma_{[t]}^u(\sigma) = r(z_t)$.*

**Proof :** a) Take a Markov strategy $\sigma = (\sigma_t)_{t \geq 1}$ in $\Sigma^M$. Put $u_1 = u$ and $y_0 = 0$, so that $z_0 = (u_1, y_0)$. Define by induction, for every $t \geq 1$, $u_{t+1} = H(u_t, \sigma_t)$, $y_t = G(u_t, \sigma_t)$, and $z_t = (u_{t+1}, y_t) \in F(z_{t-1})$. $s = (z_t)_{t \geq 1}$ is a play at $z_0$.

$\gamma_{[1]}^u(\sigma) = G(u, \sigma_1) = y_1 = r(z_1)$, and for $t \geq 2$, $\gamma_{[t]}^u(\sigma) = \gamma_{[t-1]}^{H(u,\sigma_1)}(\sigma^+) = \gamma_{[t-1]}^{u_2}(\sigma^+) = \gamma_{[t-2]}^{u_3}(\sigma^{++}) = ... = \gamma_{[1]}^{u_t}((\sigma_{t'})_{t' \geq t}) = G(u_t, \sigma_t) = y_t = r(z_t)$.

b) Take a play $s = (z_t)_{t \geq 1}$ at $z_0$. Write for each $t \geq 0$, $z_t = (u_{t+1}, y_t) \in \Delta_f(X) \times [0, 1]$. For every $t \geq 1$, there exists a mapping $f_t$ from $X$ to $A$ which defines $z_t = (u_{t+1}, y_t)$ in terms of $z_{t-1}$, i.e. that $u_{t+1} = H(u_t, f_t)$, and $y_t = G(u_t, f_t)$. Simply define $\sigma = (f_t)_{t \geq 1}$. As in point a), one can check that $\gamma_{[1]}^u(\sigma) = r(z_t)$ for each positive $t$. $\qquad\square$

For any $m \geq 0$, $n \geq 1$, and $s = (z_t)_{t \geq 1}$ in $Z^\infty$, we put as in definitions 3.1., 3.2. of Renault, 2011 :

$\gamma_{m,n}(s) = \frac{1}{n} \sum_{t=1}^{n} r(z_{m+t})$,

$\nu_{m,n}(s) = \min\{\gamma_{m,t}(s), t \in \{1, ..., n\}\}$,

$v_{m,n}(z_0) = \sup_{s \in S(z_0)} \gamma_{m,n}(s)$, and

$w_{m,n}(z_0) = \sup_{s \in S(z_0)} \nu_{m,n}(s)$.

By proposition 3.41 and corollary 3.30, it is easy to obtain for every $m$ and $n$, the equality of the values in the stochastic game $\Gamma(u)$ and in the MDP $\Psi(z_0)$ : $\tilde{v}_{m,n}(u) = v_{m,n}(z_0)$. As a consequence, we also have $v^*(u) = \inf_{n \geq 1} \sup_{m \geq 0} \tilde{v}_{m,n}(u) = v^*(z_0)$ (see definitions 3.6 of Renault, 2011 and notation 3.26 here). Similarly, we have $w_{m,n}(z_0) = \sup_{\sigma \in \Sigma^M} \min\{\gamma_{m,t}^u(\sigma), t \in \{1, .., n\}\} = w_{m,n}(u)$ (see definition 3.32).

Define now $\underline{v}^M(u)$ as the maximal quantity that can be guaranteed by player 1 in $\Gamma(u)$ with Markov strategies :

$$\underline{v}^M(u) = \sup_{\sigma \in \Sigma^M} \liminf_n \left(\inf_{\tau \in \mathcal{T}} \gamma_n^u(\sigma, \tau)\right) = \sup_{\sigma \in \Sigma^M} \liminf_n \gamma_{0,n}^u(\sigma).$$

Recall that the lower value of the MDP is defined by :

$\underline{v}(z_0) = \sup_{(z_t)_{t \geq 1} \in S(z_0)} \left(\liminf_n \frac{1}{n} \sum_{t=1}^n r(z_t)\right).$ Again, proposition 3.41 gives the equality $\underline{v}^M(u) = \underline{v}(z_0)$, so that we have the following relations :

$$\underline{v}(z_0) = \underline{v}^M(u) \leq \underline{v}(u) \leq \liminf_N \tilde{v}_N(u) = \liminf_N v_N(z_0)$$
$$\leq \limsup_N v_N(z_0) = \limsup_N \tilde{v}_N(u) \leq \overline{v}(u) \leq v^*(u) = v^*(z_0).$$

We can now conclude. We use the Wasserstein distance on $\Delta_f(X)$, so $Z$ naturally is a precompact metric space. For every $m$ and $n$, by corollary 3.39, $(u \mapsto w_{m,n}(u))$ is a non expansive mapping from $\Delta_f(X)$ to $[0,1]$. This implies that $(z_0 \mapsto w_{m,n}(z_0))$ is a non expansive mapping from $Z$ to $[0,1]$. As a consequence the family $(w_{m,n})_{m \geq 0, n \geq 1}$ is uniformly continuous. By corollary 3.8 of Renault, 2011, we obtain that the MDP $\Psi(z_0)$ has a uniform value which is :

$$v^*(z_0) = \underline{v}(z_0) = \lim_N v_N(z_0) = \sup_{m \geq 0} \inf_{n \geq 1} w_{m,n}(z_0) = \sup_{m \geq 0} \inf_{n \geq 1} v_{m,n}(z_0).$$

And the convergence from $(v_n)$ to $v^*$ is uniform. Back to our stochastic game $\Gamma(u)$, we obtain that $(v_n)_n$ uniformly converges to $v^*$, and

$$v^*(u) = \underline{v}^M(u) = \underline{v}(u) = \lim_N \tilde{v}_N(u) = \overline{v}(u).$$

So $\underline{v}(u) = \overline{v}(u)$, which implies that $\Gamma(u)$ has a uniform value. Moreover, $v^*(u) = \sup_{m \geq 0} \inf_{n \geq 1} w_{m,n}(u) = \sup_{m \geq 0} \inf_{n \geq 1} \tilde{v}_{m,n}(u)$, and using definition 3.6 of Renault, 2011, $v^*(u) = \inf_{n \geq 1} \sup_{m \geq 0} w_{m,n}(u) = \inf_{n \geq 1} \sup_{m \geq 0} \tilde{v}_{m,n}(u)$. This concludes the proof of theorem 3.17.

$$\square \quad \square \quad \square \quad \square$$

## 3.4 Comments.

**Remark 3.42.** Player 2 has 0-optimal strategies.

Under the same hypotheses H1,..., H7, it is possible to slightly modify the proof of Proposition 3.24 and obtain that player 2 has a Markov strategy $\tau$ which is 0-optimal in $\Gamma(u)$, i.e. such that : $\forall \varepsilon > 0, \exists N_0, \forall N \geq N_0, \forall \sigma \in \Sigma, \gamma_N^u(\sigma, \tau) \leq v + \varepsilon$.

Divide the set of stages into consecutive blocks $B^1,..., B^m,...$, such that $B^m$ has cardinal $m$ for each $m$. By lemma 3.23, there exists a Markov strategy $\tau = (\tau_t)_{t \geq 1}$ with the property that for every $m > 0$, $\tau$ plays optimally within $B^m$, in the sense that $\tau$ is an optimal strategy for player 2 in $\Gamma_{m(m-1)/2,m}(u)$. For every strategy $\sigma$ of player 1, we have :

$$\forall m \geq 1, \ \mathbb{E}_{u,\sigma,\tau}\left(\frac{1}{m(m+1)/2} \sum_{t=1}^{\max(B^m)} g(p_t, a_t, b_t)\right) \leq \frac{1}{m} \sum_{i=1}^m \tilde{v}_{i(i-1)/2,i}(u).$$

We have seen in subsubsection 3.3.4 that the values in the stochastic game $\Gamma(u)$ are the values of the MDP $\Psi(z_0)$, so we can apply lemma 3.4 of Renault, 2011 : $\forall i \geq 1, \forall k \geq 1, \tilde{v}_{i(i-1)/2,i}(u) \leq \sup_{l\geq 0} w_{l,k}(u) + \frac{k-1}{i}$.

Fix now $\varepsilon > 0$. One can find $k$ such that $\sup_{l\geq 0} w_{l,k}(u) \leq v^*(u) + \varepsilon$. Since $\frac{k-1}{i} \longrightarrow_{i\to\infty} 0$, one can find $m_0$ such that for every $m \geq m_0$,

$$\mathbb{E}_{u,\sigma,\tau}\left(\frac{1}{m(m+1)/2} \sum_{t=1}^{\max(B^m)} g(p_t, a_t, b_t)\right) \leq \frac{1}{m} \sum_{i=1}^{m}\left(v^*(u) + \varepsilon + \frac{k-1}{i}\right) \leq v^*(u) + 2\varepsilon.$$

Looking at the size of the blocks, one can show that $\tau$ is 0-optimal for player 2. $\square$

**Remark 3.43.** A simple hypothesis implying H5.

Recall that hypothesis H5 requires the existence of some subset $\mathcal{D}$ of $Lip_1(X)$ which contains $\Phi(1,0)$, and is stable under any $\Phi(\alpha, .)$. The following hypothesis is stated in terms of the mappings $g$ and $l$. The distance $d$ on $X$ is extended to $\Delta(X)$ by the Wasserstein distance.

Hypothesis H5' : $\forall p \in X, \forall a \in A, \forall p' \in X, \exists a' \in A$ such that :
$d(l(p,a), l(p',a')) \leq d(p,p')$, and $\inf_{b\in B} g(p', a', b) \geq \inf_{b\in B} g(p, a, b) - d(p, p')$.

It is easy to check that H5' implies that $Lip_1(X)$ itself is stable under any $\Phi(\alpha, .)$ so H5' implies H5. Consequently, the conclusions of theorem 3.17 are true if H1,H2,H3,H4,H5',H6, H7 hold.

# 4 Repeated games with an informed controller

## 4.1 Model

We first consider a general model of zero-sum repeated game. We have :

- five non empty finite sets : a set of states or parameters $K$, a set $I$ of actions for player 1, a set $J$ of actions for player 2, a set $C$ of signals for player 1 and a set $D$ of signals for player 2.
- an initial distribution $\pi \in \Delta(K \times C \times D)$,
- a mapping $g$ from $K \times I \times J$ to $[0,1]$, called the payoff function of player 1, and
- a mapping $q$ from $K \times I \times J$ to $\Delta(K \times C \times D)$, called the transition function.

Initially, $(k_1, c_1, d_1)$ is selected according to $\pi$, player 1 learns $c_1$ and player 2 learns $d_1$. Then simultaneously player 1 chooses $i_1$ in $I$ and player 2 chooses $j_1$

in $J$. The payoff for player 1 is $g(k_1, i_1, j_1)$, then $(k_2, c_2, d_2)$ is selected according to $q(k_1, i_1, j_1)$, etc... At any stage $t \geq 2$, $(k_t, c_t, d_t)$ is selected according to $q(k_{t-1}, i_{t-1}, j_{t-1})$, player 1 learns $c_t$ and player 2 learns $d_t$. Simultaneously, player 1 chooses $i_t$ in $I$ and player 2 chooses $j_t$ in $J$. The stage payoffs are $g(k_t, i_t, j_t)$ for player 1 and the opposite for player 2, and the play proceeds to stage $t + 1$.

From now on we fix $\Gamma = (K, I, J, C, D, g, q)$, and for every $\pi$ in $\Delta(K \times C \times D)$ we denote by $\Gamma(\pi) = (K, I, J, C, D, \pi, g, q)$ the corresponding repeated game induced by $\pi$. For the moment we make no assumption on $\Gamma$, so we have a general model including stochastic games, repeated games with incomplete information and imperfect monitoring (signals). We start with elementary definitions and notations.

Players are allowed to select their actions randomly. A (behavior) strategy for player 1 is a sequence $\sigma = (\sigma_t)_{t \geq 1}$, where for each $t$, $\sigma_t$ is a mapping from $(C \times I)^{t-1} \times C$ to $\Delta(I)$, with the interpretation that $\sigma_t(c_1, i_1, ..., c_{t-1}, i_{t-1}, c_t)$ is the lottery on actions used by player 1 at stage $n$ after $(c_1, i_1, ..., c_{t-1}, i_{t-1}, c_t)$. $\sigma_1$ is just a mapping from $C$ to $\Delta(I)$ giving the first action played by player 1 depending on his initial signal. Similarly, a strategy for player 2 is a sequence $\tau = (\tau_t)_{t \geq 1}$, where for each $t$, $\tau_t$ is a mapping from $(D \times J)^{t-1} \times D$ to $\Delta(J)$. We denote by $\Sigma$ and $\mathcal{T}$ the sets of strategies of player 1 and player 2, respectively.

It is standard that a pair of strategies $(\sigma, \tau)$ induces a probability $I\!\!P_{\pi, \sigma, \tau}$ on the set of plays $\Omega = (K \times C \times D \times I \times J)^\infty$, endowed with the $\sigma$-algebra generated by the cylinders.

**Definition 4.1.** *The payoff for player 1 induced by $(\sigma, \tau)$ in the first $N$ stages is denoted by :*

$$\gamma_N^\pi(\sigma, \tau) = I\!\!E_{\pi, \sigma, \tau} \left( \frac{1}{N} \sum_{n=1}^{N} g(k_n, i_n, j_n) \right).$$

For $\pi$ in $\Delta(K \times C \times D)$ and $N \geq 1$, the $N$-stage game game $\Gamma_N(\pi)$ is the zero-sum game with normal form $(\Sigma, \mathcal{T}, \gamma_N^\pi)$. By Kuhn 's theorem, $\Gamma_N(\pi)$ can be seen as the mixed extension of a finite game, so it has a value $v_n(\pi)$.

The following definitions are similar to those of subsection 3.1 or section 2 of Renault, 2011.

**Definition 4.2.** *Let $\pi$ be in $\Delta(K \times C \times D)$.*

*The liminf value of $\Gamma(\pi)$ is : $v^-(\pi) = \liminf_n v_n(\pi)$.*

*The limsup value of $\Gamma(\pi)$ is : $v^+(\pi) = \limsup_n v_n(\pi)$.*

*The lower (or maxmin) value of $\Gamma(\pi)$ is :*

$$\underline{v}(\pi) = \sup_{\sigma \in \Sigma} \liminf_n \left( \inf_{\tau \in \mathcal{T}} \gamma_n^\pi(\sigma, \tau) \right).$$

*The upper (or minmax) value of $\Gamma(u)$ is :*

$$\overline{v}(\pi) = \inf_{\tau \in \mathcal{T}} \limsup_n \left(\sup_{\sigma \in \Sigma} \gamma_N^\pi(\sigma, \tau)\right).$$

*We have $\underline{v}(\pi) \leq v^-(\pi) \leq v^+(\pi) \leq \overline{v}(\pi).$*

$\Gamma(\pi)$ *is said to have a uniform value if and only if $\underline{v}(\pi) = \overline{v}(\pi)$, and in this case the uniform value is $\underline{v}(\pi) = \overline{v}(\pi)$.*

An equivalent definition of the uniform value is as follows. Given a real number $v$, we say that player 1 can guarantee $v$ in $\Gamma(\pi)$ if : $\forall \varepsilon > 0, \exists \sigma \in \Sigma, \exists N_0, \forall N \geq N_0, \forall \tau \in \mathcal{T}, \gamma_N^\pi(\sigma, \tau) \geq v - \varepsilon$. Player 2 can guarantee $v$ in $\Gamma(\pi)$ if : $\forall \varepsilon > 0, \exists \tau \in \mathcal{T}, \exists N_0, \forall N \geq N_0, \forall \sigma \in \Sigma, \gamma_N^\pi(\sigma, \tau) \leq v + \varepsilon$. If player 1 can guarantee $v$ and player 2 can guarantee $w$ then clearly $w \geq v$. We also have, exactly as in claim 3.6 :

**Claim 4.3.** $\underline{v}(\pi) = \max\{v \in \mathbb{R}, \text{player 1 can guarantee } v \text{ in } \Gamma(\pi) \},$
$\overline{v}(\pi) = \min\{v \in \mathbb{R}, \text{player 2 can guarantee } v \text{ in } \Gamma(\pi) \}.$
*A real number $v$ can be guaranteed by both players if and only if $v$ is the uniform value of $\Gamma(\pi)$.*

Whenever the uniform value exists, a strategy $\sigma$ of player 1 satisfying : $\exists N_0, \forall N \geq N_0, \forall \tau \in \mathcal{T}, \gamma_N^\pi(\sigma, \tau) \geq v - \varepsilon$ is called $\varepsilon$-optimal for player 1 in $\Gamma(\pi)$, and a 0-optimal strategy is simply called optimal. Optimal strategies for player 2 are defined similarly. We now consider hypotheses on $q$ and $\pi$.

**Hypothesis HA** : Player 1 is informed of everything, i.e. at any stage $t \geq 1$, he can deduce from his signal $c_t$ : the state $k_t$, player 2's signal $d_t$, and if $t \geq 2$, he can also deduce from $c_t$ the action $j_{t-1}$ previously played by player 2.

**Hypothesis HB** : Player 1 fully controls the transition, i.e. $q(k, i, j)$ does not depend on $j$ for each $(k, i)$ in $K \times I$.

HA and HB are very strong hypotheses, and they are incompatible as soon as $J$ has several elements. We will use weaker hypotheses.

**Hypothesis HA'** : Player 1 is informed, in the sense that he can always deduce the state and player 2's signal from his own signal. Formally, there exists two mappings $\hat{k} : C \longrightarrow K$ and $\hat{d} : C \longrightarrow D$ such that, if $E$ denotes $\{(k, c, d) \in K \times C \times D, \hat{k}(c) = k \text{ and } \hat{d}(c) = d\}$, then :

$$\pi(E) = 1, \text{ and } q(k, i, j)(E) = 1, \forall (k, i, j) \in K \times I \times J.$$

Notice that HA' does not mean that player 1 knows everything, see the example of section 2. Since we did not include the moves in the signals, knowing the signal $d_t$ of player 2 at some stage $t$ does not imply knowing the action $j_t$ by player 2. However, not knowing this action will not be a problem for player 1

because we will also assume that player 2 does not really influence the transitions.

**Hypothesis HB'** : Player 1 controls the transition, in the sense that the marginal of the transition $q$ on $K \times D$ does not depend on player 2's action. For $k$ in $K$, $i$ in $I$ and $j$ in $J$, we denote by $\bar{q}(k, i)$ the marginal of $q(k, i, j)$ on $K \times D$.

Assume that HA' and HB' hold. The pair (new state, signal of player 2) is selected according to a distribution depending on the current state and player 1's action, but not depending on player 2's action. Player 2 may influence the distribution of player 1's signal, but still player 1 will be able to deduce the state and player 2's new information on the state. So essentially player 2 can influence player 1's knowledge about player 2's action. But this information is not relevant because it does not affect player 2's belief on the future states.

**Theorem 4.4.** *Under the hypotheses HA' and HB', the repeated game $\Gamma(\pi)$ has a uniform value.*

The next subsection is devoted to the proof of theorem 4.4. See subsection 4.3 for other comments on hypotheses, applications and open questions.

## 4.2   Proof of theorem 4.4

We assume in this subsection that HA' and HB' are satisfied. Keeping fixed all other quantities, increasing the set $C$ of signals for player 1 has no influence on the existence of the uniform value, so in the sequel we will assume w.l.o.g. that : *The mapping $( c \longrightarrow (\hat{k}(c), \hat{d}(c)) )$ is a surjection from $C$ to $K \times D$.*

We put $X = \Delta(K)$. An element $u$ in $\Delta_f(X)$ is written $u = \sum_{p \in X} u(p)\delta_p$. As in the previous section, we use the Wasserstein distance, and the (reverse of ) the Choquet order on $\Delta(X)$. $\forall u \in \Delta(X), \forall v \in \Delta(X),\ \ d(u, v) = \sup_{f: E \to \mathbb{R}, 1-Lip}|u(f) - v(f)|$. And we write $u \succeq v$ iff for every continuous *concave* real valued mapping $f$ defined on $X$ , $u(f) \geq v(f)$.

If $\mathcal{S}$ is a finite set, we use the norm $\|.\|_1$ on $\mathbb{R}^{\mathcal{S}}$. The set of probability distributions $\Delta(\mathcal{S})$ is viewed [4] as a subset of $\mathbb{R}^{\mathcal{S}}$.

### 4.2.1   Value of finite games.

As in definition 3.20, we need to consider a large family of finite games.

**Definition 4.5.** *Let $\theta = \sum_{t \geq 1} \theta_t \delta_t$ be in $\Delta_f(\mathbb{N}^*)$, i.e. $\theta$ is a probability with finite support over positive integers. For $\pi$ in $\Delta(K \times C \times D)$, the game $\Gamma_{[\theta]}(\pi)$ is the game with normal form $(\Sigma, \mathcal{T}, \gamma_{[\theta]}^{\pi})$, where :*

$$\gamma_{[\theta]}^{\pi}(\sigma, \tau) = \mathbb{E}_{\pi,\sigma,\tau}\left(\sum_{t \geq 1} \theta_t\, g(k_t, i_t, j_t)\right).$$

---

4. Notice that if we put $d(s, s') = 2$ for any distinct elements of $\mathcal{S}$, then for every $p$ and $q$ in $\Delta(\mathcal{S})$ we have $\sup_{f: \mathcal{S} \to \mathbb{R}, 1-Lip}|\sum_s p^s f(s) - \sum_s q^s f(s)| = \|p - q\|_1$.

*Particular cases : if $\theta = 1/n \sum_{t=1}^{n} \delta_t$, $\Gamma_{[\theta]}(\pi)$ is nothing but $\Gamma_n(\pi)$.*

*For $m \geq 0$ and $n \geq 1$, we denote by $\Gamma_{m,n}(\pi)$ the game $\Gamma_{[\theta]}(\pi)$ where $\theta = 1/n \sum_{t=m+1}^{m+n} \delta_t$. The payoff function is written in this case : $\gamma_{m,n}^{\pi}(\sigma, \tau)$. The value of $\Gamma_{m,n}(\pi)$ will be denoted by $v_{m,n}(\pi)$.*

Notice that $v_{0,n}$ is just $v_n$, the value of the $n$-stage game. The following lemma is true without the hypotheses $HA'$ and $HB'$.

**Lemma 4.6.** *For every $\theta \in \Delta_f(\mathbb{N}^*)$ and $\pi \in \Delta(K \times C \times D)$, the game $\Gamma_{[\theta]}(\pi)$ has a value, denoted by $v_{[\theta]}(\pi)$, and both players have optimal strategies. Moreover, $v_{[\theta]}$ is a non expansive mapping from $\Delta(K \times C \times D)$ to $\mathbb{R}$.*

**Proof :** The existence of the value and optimal strategies is standard. Notice that for every $\theta$, $\pi$, and strategy pair $(\sigma, \tau)$ :

$$\gamma_{[\theta]}^{\pi}(\sigma, \tau) = \sum_{k_1, c_1, d_1} \pi(k_1, c_1, d_1) \, \gamma_{[\theta]}^{\delta_{(k_1, c_1, d_1)}}(\sigma, \tau).$$

Since we use $\|.\|_1$, $v_{[\theta]}$ is 1-Lipschitz. $\qquad\qquad\square$

**Definition 4.7.** *We define a mapping $\Psi$ from $\Delta(K \times D)$ to $\Delta_f(X)$ by : for each probability $\pi$ on $K \times D$, $\Psi(\pi) = \sum_{d \in D} \pi(d) \delta_{\pi^d}$, where for each $d$, $\pi^d$ is the conditional probability on $K$ given $d$ issued from $\pi$.*

**Notation 4.8.** *Let $\pi$ be in $\Delta(K \times C \times D)$. We denote by $\bar{\pi}$ the marginal of $\pi$ on $K \times D$, and denote by $\hat{\pi}$ the probability induced by $\pi$ (or $\bar{\pi}$) on $X$, i.e. $\hat{\pi} = \Psi(\bar{\pi}) = \sum_{d \in D} \pi(d) \delta_{\pi^d} \in \Delta_f(X)$, where for each $d$, $\pi^d$ is the conditional probability on $K$ given $d$ issued from $\pi$ (or $\bar{\pi}$).*

*We also put $\Delta^E = \{\pi \in \Delta(K \times C \times D), \pi(E) = 1\}$, where $E = \{(k, c, d) \in K \times C \times D, \hat{k}(c) = k \text{ and } \hat{d}(c) = d\}$.*

**Lemma 4.9.** *Let $\pi$ and $\pi'$ be in $\Delta^E$ such that $\hat{\pi} = \hat{\pi}'$. Then $v_{[\theta]}(\pi) = v_{[\theta]}(\pi')$ for each $\theta$.*

**Proof :** Fix $\theta$ in $\Delta_f(\mathbb{N}^*)$, $\pi$ in $\Delta^E$, and a strategy pair $(\sigma, \tau)$.

$$\gamma_{[\theta]}^{\pi}(\sigma, \tau) = \sum_{d_1} \pi(d_1) \sum_{k_1} \pi(k_1 | d_1) \sum_{c_1} \pi(c_1 | d_1, k_1) \, \gamma_{[\theta]}^{\delta_{(k_1, c_1, d_1)}}(\sigma, \tau).$$

Since $\pi(E) = 1$, player 1 can deduce $d_1$ and $k_1$ from $c_1$, so :

$$\sup_{\sigma \in \Sigma} \gamma_{[\theta]}^{\pi}(\sigma, \tau) = \sum_{d_1} \pi(d_1) \sum_{k_1} \pi(k_1 | d_1) \sum_{c_1} \pi(c_1 | d_1, k_1) \sup_{\sigma \in \Sigma} \gamma_{[\theta]}^{\delta_{(k_1, c_1, d_1)}}(\sigma, \tau).$$

But $\sup_{\sigma \in \Sigma} \gamma_{[\theta]}^{\delta_{(k_1, c_1, d_1)}}(\sigma, \tau)$ does not depend on $c_1$, so for any fixed $c^*$ in $C$,

$$\sup_{\sigma \in \Sigma} \gamma_{[\theta]}^{\pi}(\sigma, \tau) = \sum_{d_1} \pi(d_1) \sum_{k_1} \pi(k_1 | d_1) \sup_{\sigma \in \Sigma} \gamma_{[\theta]}^{\delta_{(k_1, c^*, d_1)}}(\sigma, \tau).$$

Consequently, $\sup_{\sigma \in \Sigma} \gamma_{[\theta]}^{\pi}(\sigma, \tau)$ only depends on $\hat{\pi}$, $\tau$ and $\theta$, and $v_{[\theta]}(\pi) = \inf_{\tau \in \mathcal{T}} \sup_{\sigma \in \Sigma} \gamma_{[\theta]}^{\pi}(\sigma, \tau)$ only depends on $\hat{\pi}$ and $\theta$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remember that we assumed w.l.o.g. that the function $(\hat{k}, \hat{d})$ appearing in hypothesis HA' is surjective. It will be convenient in the sequel to use the following notation.

**Notation 4.10.** *For any $(k, d)$ in $K \times D$, we fix an element $c(k, d)$ in $C$ such that $\hat{k}(c(k, d)) = k$ and $\hat{d}(c(k, d)) = d$.*

$v_{[\theta]}$ has been defined as a mapping from $\Delta(K \times C \times D)$ to $\mathbb{R}$. We now define value functions with domain $X$ and $\Delta(X)$. Let $p$ be in $X$. We define $\pi$ in $\Delta^E$ as follows : fix $d^*$ in $D$, and $\pi$ chooses, for each $k$ in $K$, the element $(k, c(k, d^*), d^*)$ with probability $p^k$. Then $\hat{\pi} = \delta_p$, so by the previous lemma $v_{[\theta]}(\pi)$ only depends on $p$. We thus define :

$$v_{[\theta]}(p) = v_{[\theta]}(\pi).$$

With a slight abuse of notation, $v_{[\theta]}$ now also denotes a mapping from $\Delta(K)$ to $\mathbb{R}$. And we have for each $\pi$ in $\Delta^E$ :

$$
\begin{aligned}
v_{[\theta]}(\pi) &= \inf_{\tau \in \mathcal{T}} \sup_{\sigma \in \Sigma} \gamma_{[\theta]}^{\pi}(\sigma, \tau), \\
&= \sum_{d_1 \in D} \pi(d_1) \inf_{\tau \in \mathcal{T}} \left( \sum_{k_1} \pi(k_1 | d_1) \sup_{\sigma \in \Sigma} \gamma_{[\theta]}^{\delta_{(k_1, c(k_1, d_1), d_1)}}(\sigma, \tau) \right), \\
&= \sum_{d_1 \in D} \pi(d_1) v_{[\theta]}(\pi^{d_1}).
\end{aligned}
$$

So we have obtained the following.

**Lemma 4.11.**

$$\forall \pi \in \Delta^E, \forall \theta \in \Delta_f(\mathbb{N}^*), \quad v_{[\theta]}(\pi) = \sum_{d_1 \in D} \pi(d_1) v_{[\theta]}(\pi^{d_1}).$$

**Notation 4.12.** *$\tilde{v}_{[\theta]}$ denotes the affine extension of $v_{[\theta]}$ on $\Delta(X)$, i.e. : $\forall u \in \Delta(X), \tilde{v}_{[\theta]}(u) = \int_{p \in X} v_{[\theta]}(p) du(p)$.*

From the previous computations, $\tilde{v}_{[\theta]}$ is clearly linked to the original value function $v_{[\theta]}$.

**Claim 4.13.**
$$\forall \theta \in \Delta_f(\mathbb{N}^*), \forall \pi \in \Delta^E, v_{[\theta]}(\pi) = \tilde{v}_{[\theta]}(\hat{\pi}).$$

So from the knowledge of $v_{[\theta]}$ on $X$, one can deduce its extension $\tilde{v}_{[\theta]}$ on $\Delta(X)$ and then the original value function $v_{[\theta]}$ on $\Delta^E$. We have, for each $p$ in $\Delta(K)$ (and for any $d^*$ in $D$) :

$v_{[\theta]}(p) = \inf_{\tau \in \mathcal{T}} \left( \sum_{k_1} p^{k_1} \sup_{\sigma \in \Sigma} \gamma_{[\theta]}^{\delta_{(k_1, c(k, d^*), d^*)}}(\sigma, \tau) \right)$. So $v_{[\theta]}$ is a concave and non expansive mapping from $\Delta(K)$ to $\mathbb{R}$. Consequently, $\tilde{v}_{[\theta]}$ is a non decreasing

and non expansive mapping from $\Delta(X)$ to $I\!R$.

We finally define :

**Definition 4.14.** *For $\pi$ in $\Delta(K \times C \times D)$, we put :*

$$v^*(\pi) = \inf_{n \geq 1} \sup_{m \geq 0} v_{m,n}(\pi).$$

### 4.2.2 An auxiliary stochastic game.

We now introduce a stochastic game with complete information, to be played in pure strategies, as in section 3.

**Definition 4.15.** *Recall that $X = \Delta(K)$. We put $A = \Delta(I)^K$ and $B = \Delta(J)$, and define for every $p$ in $X$, $a$ in $A$ and $b$ in $B$ :*

$$
\begin{aligned}
g(p,a,b) &= \sum_{k \in K} p^k \sum_{i \in I} \sum_{j \in J} a^k(i)b(j)g(k,i,j) \in [0,1], \\
g(p,a) &= \inf_{b \in B} g(p,a,b), \\
Q(p,a,b) &= \sum_{(k,i,j) \in K \times I \times J} p^k a^k(i)b(j)q(k,i,j) \in \Delta(K \times C \times D), \\
\bar{Q}(p,a) &= \sum_{(k,i) \in K \times I} p^k a^k(i)\bar{q}(k,i) \in \Delta(K \times D), \\
l(p,a) &= \Psi(\bar{Q}(p,a)) \in \Delta_f(X).
\end{aligned}
$$

*For $u$ in $\Delta_f(X)$, we write $\hat{\Gamma}(u)$ for the stochastic game $(X, A, B, g, l, u)$ with initial distribution $u$.*

By hypothesis HB', it is easy to see that the marginal of $Q(p,a,b)$ on $\Delta(K \times D)$ does not depend on $b$, and precisely is $\bar{Q}(p,a)$. $l(p,a)$ is nothing but $\sum_{d \in D} \bar{Q}(p,a)(d) \, \delta_{\bar{Q}(p,a)^d}$, where for each $d$, $\bar{Q}(p,a)^d$ is the conditional probability on $K$ given $d$ issued from $\bar{Q}(p,a)$. Notice also that for every $(p,a,b)$, $Q(p,a,b)$ belongs to the convex set $\Delta^E$.

Suppose that in the original game $\Gamma(\pi)$, the current state is selected according to $p$, player 1 knows $k$ and plays the mixed action $a^k \in \Delta(I)$, whereas player 2 just knows $p$ and plays the mixed action $b \in \Delta(J)$. Then $g(p,a,b)$ is the (ex-ante) expected payoff for player 1, and $l(p,a)$ is the (ex-ante) distribution of player 2's future belief on the next state.

We will eventually apply theorem 3.17 to $\hat{\Gamma}(\hat{\pi})$, so we have to check the hypotheses H1 to H7 of section 3. H1, H2, H3 and H4 are clearly true. We now need the following properties of the mapping $\Psi$.

**Lemma 4.16.** *$\Psi$ is concave :*

$$\forall \pi', \pi'' \in \Delta(K \times D), \forall \lambda \in [0,1], \ \Psi(\lambda \pi' + (1-\lambda)\pi'') \succeq \lambda \Psi(\pi') + (1-\lambda)\Psi(\pi'').$$

**Proof :** Write $\pi = \lambda\pi' + (1-\lambda)\pi''$. Notice that for each $d$ in $D$, $\pi^d = \frac{1}{\pi(d)}(\lambda\pi'(d)\pi'^d + (1-\lambda)\pi''(d)\pi''^d)$. Let $f$ be a concave continuous mapping from $X$ to $\mathbb{R}$, we have to show that $\Psi(\pi)(f) \geq \lambda\Psi(\pi')(f) + (1-\lambda)\Psi(\pi'')(f)$.

$$
\begin{aligned}
\lambda\Psi(\pi')(f) + (1-\lambda)\Psi(\pi'')(f) &= \lambda\sum_d \pi'(d)f(\pi'^d) + (1-\lambda)\sum_d \pi''(d)f(\pi''^d), \\
&= \sum_d \pi(d)\left(\frac{\lambda\pi'(d)}{\pi(d)}f(\pi'^d) + \frac{(1-\lambda)\pi''(d)}{\pi(d)}f(\pi''(d))\right), \\
&\leq \sum_d \pi(d)f\left(\frac{\lambda\pi'(d)}{\pi(d)}\pi'^d + \frac{(1-\lambda)\pi''(d)}{\pi(d)}\pi''(d)\right), \\
&= \sum_d \pi(d)f(\pi^d) = \Psi(\pi)(f). \qquad \square
\end{aligned}
$$

For any $p$ in $X$, the marginal $\bar{Q}(p,a)$ is affine in $a$, so we obtain that $l(p,a) = \Psi(\bar{Q}(p,a))$ is concave in $a$. Hypothesis H6 will then immediately follow from the next lemma.

**Lemma 4.17.** $\Psi$ *is continuous.*

**Proof :** Let $(\pi_n)_n$ be a sequence in $\Delta(K \times D)$ converging for the norm $\|.\|_1$ to $\pi$. It is easy to see that for every $f$ continuous, $\Psi(\pi_n)(f) = \sum_d \pi_n(d)f(\pi_n^d)$ converges as $n$ goes to infinity to $\sum_d \pi(d)f(\pi^d) = \Psi(\pi)(f)$. $\qquad\square$

**Remark 4.18.** One can show that $\Psi$ is Lipschitz, but it is not 1-Lipschitz when $\|.\|_1$ and the Wasserstein distance are used.

**Lemma 4.19.** *"Splitting hypothesis" H7. Consider a convex combination* $p = \sum_{s=1}^{S} \lambda_s p_s$ *in $X$, and a family of actions $(a_s)_{s\in S}$ in $A^S$. Then there exists $a$ in $A$ such that :*

$$
l(p,a) \succeq \sum_{s\in S} \lambda_s l(p_s, a_s) \quad \text{and} \quad g(p,a) \geq \sum_{s\in S} \lambda_s g(p_s, a_s).
$$

**Proof :** Define $a : K \longrightarrow \Delta(I)$ with the well known splitting procedure of Aumann and Maschler : observe $k$ in $K$ which has been selected according to $p$, then choose $s$ with probability $\lambda_s p_s^k / p^k$, and finally play $a_s^k$. Formally, put $a^k = \sum_{s\in S} \frac{\lambda_s p_s^k}{p^k} a_s^k \in \Delta(I)$ if $p^k > 0$, and define arbitrarily $a^k$ if $p^k = 0$. We have :

$$
\begin{aligned}
\bar{Q}(p,a) &= \sum_{k\in K}\sum_{i\in I} p^k a^k(i)\bar{q}(k,i), \\
&= \sum_{s\in S}\sum_{k\in K}\sum_{i\in I} \lambda_s p_s^k a_s^k(i)\bar{q}(k,i), \\
&= \sum_{s\in S} \lambda_s \bar{Q}(p_s, a_s)
\end{aligned}
$$

So by concavity of $\Psi$, we have $l(p,a) \succeq \sum_s \lambda_s l(p_s, a_s)$.

Regarding payoffs, we have for each $b$ in $B$, $g(p,a,b) = \sum_s \lambda_s g(p_s, a_s, b)$, so $\inf_{b\in B} g(p,a,b) \geq \sum_{s\in S} \lambda_s \inf_{b\in B} g(p_s, a_s, b)$. $\qquad\square$

Up to now, only H5 remains to be proved.

### 4.2.3  The recursive formula.

We now prove a standard recursive formula for the value functions. As in definition 3.20, for $\theta = \sum_{t \geq 1} \theta_t \delta_t$ we define $\theta^+$ as the law of $t^* - 1$ given that $t^* \geq 2$, so that $\theta^+ = \sum_{t \geq 1} \frac{\theta_{t+1}}{1-\theta_1} \delta_t$ if $\theta_1 \neq 1$, and $\theta^+$ is defined arbitrarily if $\theta_1 = 1$.

**Proposition 4.20.** *For $\theta$ in $\Delta_f(\mathbb{N}^*)$ and $p$ in $X$,*

$$
\begin{aligned}
v_{[\theta]}(p) &= \max_{a \in A} \min_{b \in B} \left( \theta_1 g(p,a,b) + (1-\theta_1)\tilde{v}_{[\theta^+]}(l(p,a)) \right), \\
&= \min_{b \in B} \max_{a \in A} \left( \theta_1 g(p,a,b) + (1-\theta_1)\tilde{v}_{[\theta^+]}(l(p,a)) \right).
\end{aligned}
$$

*For every $\pi$ in $\Delta^E$, in the game $\Gamma_{[\theta]}(\pi)$ both players have optimal strategies only depending on $\bar{\pi} \in \Delta(K \times D)$.*

**Proof :** By the proof of lemma 4.9, we know that for any $\tau$ in $\mathcal{T}$ and fixed $c^*$ in $C$,

$$
\sup_{\sigma \in \Sigma} \gamma_{[\theta]}^\pi(\sigma,\tau) = \sum_{d_1} \pi(d_1) \sum_{k_1} \pi(k_1|d_1) \sup_{\sigma \in \Sigma} \gamma_{[\theta]}^{\delta_{(k_1,c^*,d_1)}}(\sigma,\tau).
$$

And $\tau$ is optimal in $\Gamma_{[\theta]}(\pi)$ if and only if $\sum_{d_1} \pi(d_1) \sum_{k_1} \pi(k_1|d_1) \sup_{\sigma \in \Sigma} \gamma_{[\theta]}^{\delta_{(k_1,c^*,d_1)}}(\sigma,\tau)$ $= \sum_{d_1} \pi(d_1) \inf_{\tau' \in \mathcal{T}} \sum_{k_1} \pi(k_1|d_1) \sup_{\sigma \in \Sigma} \gamma_{[\theta]}^{\delta_{(k_1,c^*,d_1)}}(\sigma,\tau')$. Hence player 2 has an optimal strategy in $\Gamma_{[\theta]}(\pi)$ that only depends on $\bar{\pi}$.

We now show that player 1 has an optimal strategy in $\Gamma_{[\theta]}(\pi)$ that only depends on $\bar{\pi}$. For every $d_1$ in $D$, fix an optimal strategy $\sigma(d_1)$ of player 1 in the game $\Gamma_{[\theta]} \left( \sum_{k_1 \in K} \pi(k_1|d_1) \delta_{(k_1,c(k_1,d_1),d_1)} \right)$. Define now $\sigma$ that plays after each initial signal $c_1$ exactly what $\sigma(\hat{d}(c_1))$ plays after the initial signal $c(\hat{k}(c_1), \hat{d}(c_1))$. One can check that $\sigma$ is optimal in any game $\Gamma_{[\theta]}(\pi')$, if $\pi' \in \Delta^E$ and $\bar{\pi}' = \bar{\pi}$.

We now prove the recursive formula by induction on the greatest element in the support of $\theta$. $v_1(p) = \max_{a \in A} \min_{b \in B} g(p,a,b) = \min_{b \in B} \max_{a \in A} g(p,a,b)$ is easy by Sion's theorem.

Fix now $n \geq 2$, and assume that the proposition is true for every $\theta$ with support included in $\{1,...,n-1\}$. Fix a probability $\theta = \sum_{t=1}^{n} \theta_t \delta_t$, and notice that $\theta^+$ has a support included in $\{1,...,n-1\}$. Fix also $p$ in $X$. The equality $\max_{a \in A} \min_{b \in B} (\theta_1 g(p,a,b) + (1-\theta_1)\tilde{v}_{[\theta^+]}(l(p,a))) = \min_{b \in B} \max_{a \in A} (\theta_1 g(p,a,b) + (1-\theta_1)\tilde{v}_{[\theta^+]}(l(p,a)))$ is standard and similar to the proof of proposition 3.21, so the proof is omitted here.

By definition, we have $v_{[\theta]}(p) = v_{[\theta]}(\pi)$, where $\pi = \sum_k p^k \delta_{(k,c(k,d^*),d^*)} \in \Delta(K \times C \times D)$, and $d^*$ is an arbitrary element of $D$. We thus consider the game $\Gamma_{[\theta]}(\pi)$. Let $(\sigma,\tau)$ be a strategy pair. Write $a = (a^k)_k = (\sigma_1(c(k,d^*)))_k$ in $A = \Delta(I)^K$ for the action played by player 1 at stage 1, and write $b = \tau_1(d^*) \in B = \Delta(J)$ for the first stage action of player 2. We have :

$$
\gamma_{[\theta]}^\pi(\sigma,\tau) = \theta_1 g(p,a,b) + (1-\theta_1) \sum_{k,i_1,j_1} p^k a^k(i_1) b(j_1) \gamma_{[\theta^+]}^{q(k,i_1,j_1)}(\sigma_{c(k,d^*),i_1}^+, \tau_{d^*,j_1}^+),
$$

where $\sigma^+_{c(k,d^*),i_1}$ and $\tau^+_{d^*,j_1}$ are continuation strategies.

$a$ being fixed, we can choose $\sigma^+$ an optimal strategy for player 1 in the game $\Gamma_{[\theta+]}(Q(p,a,j_1))$, and this choice can be made independently of $j_1$ since the marginal $\bar{Q}(p,a,j_1)$ does not depend on $j_1$. We have for every $j_1$ in $J$ : $\sum_{k,i_1} p^k a^k(i_1) \gamma^{q(k,i_1,j_1)}_{[\theta+]}(\sigma^+,\tau^+_{d^*,j_1}) = \gamma^{Q(p,a,j_1)}_{[\theta+]}(\sigma^+,\tau^+_{d^*,j_1}) \geq v_{[\theta+]}(Q(p,a,j_1)) = \tilde{v}_{[\theta+]}(l(p,a))$. By playing $a$ at stage 1, and then according to $\sigma^+$ (for any first signal $c(k,d^*)$ and first action $i_1$), player 1 can thus guarantee : $\inf_b \theta_1 g(p,a,b) + (1-\theta_1)\tilde{v}_{[\theta+]}(l(p,a))$. So $v_{[\theta]}(p) \geq \max_{a\in A} \min_{b\in B}(\theta_1 g(p,a,b) + (1-\theta_1)\tilde{v}_{[\theta+]}(l(p,a)))$.

We finally show that player 2 can defend $\max_{a\in A} \min_{b\in B}(\theta_1 g(p,a,b) + (1-\theta_1)\tilde{v}_{[\theta+]}(l(p,a)))$ in $\Gamma_{[\theta]}(\pi)$. Fix a strategy $\sigma$ of player 1. $a = (\sigma_1(c(k,d^*)))_k$ being fixed, choose $b$ in $B$ achieving $\inf_b g(p,a,b)$. We also choose $\tau^+$ in $\mathcal{T}$ an optimal strategy for player 2 in the game $\Gamma_{[\theta+]}(Q(p,a,j_1))$, and this choice can be made independently of $j_1$. $\tau^+$ now being fixed, notice that there exists a strategy $\sigma'$ of player 1 which is a best reply to $\tau^+$ in any game $\Gamma_{[\theta+]}(\pi')$, with $\pi' \in \Delta^E$. We now have :

$$
\begin{aligned}
\gamma^\pi_{[\theta]}(\sigma,\tau) &= \theta_1 g(p,a,b) + (1-\theta_1)\sum_{j_1} b(j_1)\sum_{k,i_1} p^k a^k(i_1)\gamma^{q(k,i_1,j_1)}_{[\theta+]}(\sigma^+_{c(k,d^*),i_1},\tau^+), \\
&\leq \theta_1 g(p,a) + (1-\theta_1)\sum_{j_1} b(j_1)\sum_{k,i_1} p^k a^k(i_1)\gamma^{q(k,i_1,j_1)}_{[\theta+]}(\sigma',\tau^+), \\
&= \theta_1 g(p,a) + (1-\theta_1)\sum_{j_1} b(j_1)\gamma^{Q(p,a,j_1)}_{[\theta+]}(\sigma',\tau^+), \\
&= \theta_1 g(p,a) + (1-\theta_1)\sum_{j_1} b(j_1)v_{[\theta+]}(l(p,a)), \\
&\leq \sup_{a\in A}\left(\theta_1 g(p,a) + (1-\theta_1)\tilde{v}_{[\theta+]}(l(p,a))\right)
\end{aligned}
$$

So $v_{[\theta]}(p) \leq \max_{a\in A} \min_{b\in B}\left(\theta_1 g(p,a,b) + (1-\theta_1)\tilde{v}_{[\theta+]}(l(p,a))\right)$. $\qquad\square$

### 4.2.4 Player 2 can guarantee $v^*(\pi)$ in $\Gamma(\pi)$.

We first have an analog of lemma 3.23. Recall that a strategy for player 2 is a sequence $\tau = (\tau_1,\tau_2,...,\tau_t,...)$, where for each $t$, $\tau_t$ is a mapping from $(D\times J)^{t-1}\times D$ to $\Delta(J)$.

**Lemma 4.21.** *For every $\pi \in \Delta^E$, $n \geq 1$ and $m \geq 0$, $\forall \tau_1,...,\tau_m$, $\exists \tau_{m+1},...,\tau_{m+n}$ such that any strategy of player 2 starting by $\tau_1,...,\tau_m,...,\tau_{m+n}$ is optimal in the game $\Gamma_{m,n}(\pi)$.*

**Proof :** Fix $\pi$, $m$, $n$, and $\tau_1,...,\tau_m$, with $\tau_t : (D\times J)^{t-1}\times D \longrightarrow \Delta(J)$ for every $t \leq m$. Define $\mathcal{T}^\dagger$ as the set of strategies of player 2 that start with $\tau_1,...,\tau_m$. Let us now consider the zero-sum game $\Gamma^\dagger_{m,n}(\pi)$ with strategy set $\Sigma$ for player 1, $\mathcal{T}^\dagger$ for player 2, and payoff function (the restriction of) $v_{m,n}$. Stages greater than $m+n$ do not care, so $\Gamma^\dagger_{m,n}(\pi)$ can be seen as the mixed extension of a finite game, where nature plays $\tau_1,...,\tau_m$ instead of player 2 for the first $m$

stages. Consequently, $\Gamma^{\dagger}_{m,n}(\pi)$ has a value which we denote by $v^{\dagger}_{m,n}(\pi)$. Clearly, $v^{\dagger}_{m,n}(\pi) \geq v_{m,n}(\pi)$. Now, for any strategy $\sigma$ of player 1, it is easy to construct, by the recursive formula of proposition 4.20, a strategy $\tau$ that defends $v_{m,n}(\pi)$ against $\sigma$. So $v^{\dagger}_{m,n}(\pi) = v_{m,n}(\pi)$, and considering an optimal strategy of player 2 in $\Gamma^{\dagger}_{m,n}(\pi)$ concludes the proof. $\qquad\square$

**Proposition 4.22.** *For each $\pi$ in $\Delta^E$, player 2 can guarantee $v^*(\pi)$ in the game $\Gamma(\pi)$.*

**Proof :** Divide the set of stages into consecutive blocks $B^1$, $B^2$,..., $B^m$ of equal length $n$. Define a strategy $\tau$ of player 2 as follows. At block $B^1$, pick $\tau_1, \tau_2, ..., \tau_n$ in order to get an optimal strategy in $\Gamma_{0,n}(\pi)$. At block $B^2$, use lemma 4.21 to construct $\tau_{n+1}, ..., \tau_{2n}$ and get an optimal strategy also in $\Gamma_{n,n}(\pi)$, etc... At block $B^{m+1}$, given $\tau_1,...,\tau_{nm}$, use lemma 4.21 to define $\tau_{nm+1},..., \tau_{n(m+1)}$ to get an optimal strategy in $\Gamma_{nm,n}(\pi)$.

For any $\sigma$ and $M \geq 1$, we have :

$\mathbb{E}_{\pi,\sigma,\tau} \left( \frac{1}{Mn} \sum_{t=1}^{Mn} g(k_t, i_t, j_t) \right) \leq \frac{1}{M} \sum_{m=0}^{M-1} \tilde{v}_{mn,n}(\pi) \leq \sup_m \tilde{v}_{m,n}(\pi)$. So player 2 can guarantee $v^*(\pi)$. $\qquad\square$

We now have the following inequality chain, for $\pi$ in $\Delta^E$ :

$$\boxed{\underline{v}(\pi) \leq v^-(\pi) \leq v^+(\pi) \leq \overline{v}(\pi) \leq v^*(\pi).}$$

### 4.2.5 Player 1 can guarantee $v^*(\pi)$ via the auxiliary game.

For $f$ continuous and $\alpha$ in $[0,1]$, we defined in notation 3.14 the mapping $\Phi(\alpha, f)$ as follows : $\forall p \in X$,

$\Phi(\alpha, f)(p) = \sup_{a \in A} \inf_{b \in B} \left( \alpha\, g(p, a, b) + (1 - \alpha)\, \tilde{f}(l(p, a)) \right)$.

We now simply define the following subset of mappings from $X$ to $\mathbb{R}$ :

$$\mathcal{D} = \{v_{[\theta]},\ \theta \in \Delta_f(\mathbb{N}^*)\}.$$

By the recursive formula, we obtain that $\mathcal{D}$ is stable under $\Phi$ : $\forall f \in \mathcal{D}, \forall \alpha \in [0,1]$, $\Phi(\alpha, f) \in \mathcal{D}$. Since $v_1 = \Phi(1,0) \in \mathcal{D}$ and all elements of $\mathcal{D}$ are non expansive, the hypothesis H5 of section 3 is satisfied.

**Proposition 4.23.**

*a) For every $u$ in $\Delta_f(X)$, the auxiliary game $\hat{\Gamma}(u) = (X, A, B, g, l, u)$ defined in definition 4.15 satisfies the hypotheses H1,..., H7 of theorem 3.17.*

*b) For every $\theta$ in $\Delta_f(\mathbb{N}^*)$ and $u$ in $\Delta_f(X)$, the auxiliary game $\hat{\Gamma}_{[\theta]}(u)$ defined from $\hat{\Gamma}(u)$ in definition 3.20 has a value which corresponds to $\tilde{v}_{[\theta]}(u)$, as defined in notation 4.12.*

*c) For $\pi$ in $\Delta^E$, anything that can be guaranteed by player 1 with Markov strategies in $\hat{\Gamma}_{[\theta]}(\hat{\pi})$ can be guaranteed by player 1 in the original game $\Gamma(\pi)$.*

**Proof :** a) H1,...,H7 have been proved. b) The equality between the value functions of the original game and of the auxiliary game comes from propositions 3.21 and 4.20. c) A Markov strategy of player 1 is a sequence $(\sigma_t)_{t\geq 1}$, where for each $t$ $\sigma_t$ is a mapping from $X$ to $A$ giving the action to be played on stage $t$ depending on the current state in $X = \Delta(K)$. It induces a probability distribution on $(X \times A)^\infty$, regardless of player 2's actions (see subsection 3.3.3). Any such strategy can be mimicked by player 1 in the original game, since this player can compute at each stage the belief of player 2 on the state in $K$. □

Notice that the analog of point c) is not true regarding player 2, because in the original game this player can compute posterior beliefs on $K$ only if he knows player 1's strategy.

We can now conclude the proof of theorem 4.4. Fix $\pi$ in $\Delta^E$, and put $u = \hat{\pi} \in \Delta_f(X)$. For every $\theta$, we have $v_{[\theta]}(\pi) = \tilde{v}_{[\theta]}(u)$ by proposition 4.23 b) and claim 4.13. So $v^*(\pi) = \inf_n \sup_m v_{m,n}(\pi) = \inf_n \sup_m \tilde{v}_{m,n}(u) = v^*(u)$. By theorem 3.17, $(\tilde{v}_n(u))_n$ converges to $v^*(u)$, and player 1 can guarantee $v^*(u)$ in $\hat{\Gamma}(u)$ with a Markov strategy. So we obtain that $(v_n(\pi))_n$ converges to $v^*(\pi)$, and player 1 can guarantee this quantity in the original game $\Gamma(\pi)$. Finally $\Gamma(\pi)$ has a uniform value which is :

$$\boxed{\underline{v}(\pi) = v^-(\pi) = v^+(\pi) = \overline{v}(\pi) = v^*(\pi).}$$

□  □  □  □

## 4.3  Comments and consequences

### 4.3.1  Byproducts of the proof.

The proof of theorem 4.4 shows, under the very same hypotheses HA' and HB', more than the existence of the uniform value.

In particular, the application of theorem 3.17 to the auxiliary game gives :

$$\begin{aligned}
v^*(\pi) &= \inf_{n\geq 1}\sup_{m\geq 0}v_{m,n}(\pi) = \sup_{m\geq 0}\inf_{n\geq 1}v_{m,n}(\pi), \\
&= \inf_{n\geq 1}\sup_{m\geq 0}w_{m,n}(\pi) = \sup_{m\geq 0}\inf_{n\geq 1}w_{m,n}(\pi).
\end{aligned}$$

where $v_{m,n}(\pi)$ is defined in definition 4.5, and $w_{m,n}(\pi) = \inf_{\theta\in\Delta(\{1,...,n\})}v_{[\theta^{m,n}]}(\pi)$ (see definition 3.32 and corollary 3.39).

And $(v_n)_n$ uniformly converges to $v^*$ on $\Delta^E$.

Concerning $\varepsilon$-optimal strategies, we have seen that player 1 can guarantee $v^*(u)$ with Markov strategies, i.e. with strategies that play at each stage a mixed action determined by player 2's current belief on the current state of nature in $K$.

Regarding player 2, one can strengthen the construction of proposition 4.22 and show as in remark 3.42 that player 2 has 0-optimal strategies in the game $\Gamma(\pi)$. □

### 4.3.2  HA' or HB' can not be withdrawn in theorem 4.4.

An example of a game satisfying HA' and having no uniform value is given in Sorin, 1984. It is a particular case of a stochastic game with incomplete information, where after each stage the players perfectly observe the actions just chosen. (There are two possible stochastic games of "Big Match" type, and player 1 only knows which one is being played.)

An example of a game satisfying HB' and having no uniform value is given in Sorin and Zamir, 1985. It belongs to a class of games called repeated games with incomplete information on one and a half side : at each stage, both players will play the same matrix game. Player 1 initially receives a signal which tells him which matrix game will be played, but does not know the initial signal of player 2, so can not deduce from his signal the belief of player 2 on the selected matrix game. The transition function $q$ is particularly simple there : $q(k, i, j)$ is the Dirac measure on $(k, j, i)$.

### 4.3.3  Applications.

Consider the following model of repeated games with standard monitoring. There are : a finite set of states $K$, an initial probability $p$ on $\Delta(K)$, finite action sets $I$ and $J$, for each state $k$ a payoff matrix $(G^k(i,j))_{(i,j)}$, and for each state $k$ in $K$ and action $i$ in $I$ there is a probability $l(k, i) \in \Delta(K)$. At stage 1, a state $k_1$ is selected according to $p$ and told to player 1 only. Then simultaneously player 1 chooses $i_1$ in $I$ and player 2 chooses $j_1$ in $J$. The payoff for player 1 is $G^{k_1}(i_1, j_1)$, and $(i_1, j_1)$ is publicly announced. At stage $t \geq 2$, $k_t$ is selected according to $l(k_{t-1}, i_{t-1})$ and told to player 1 only. Then the players choose $i_t$ and $j_t$. The stage payoff for player 1 is $G^{k_t}(i_t, j_t)$, $(i_t, j_t)$ is publicly announced, and the play goes to stage $t + 1$.

This model is a generalization of the model of Markov chain repeated games with lack of information on one side introduced in Renault, 2006. Here, player 1 is not only informed of the sequence of actions, but also he can influence the state process. Studying this model has lead to the present paper, and some ideas developed here already come from Renault 2006. It also contains stochastic games with a single controller and incomplete information on the side of his opponent, as studied in Rosenberg *et al.*, 2004. So the present paper generalizes both theorem 2.3 in Renault 2006, and theorem 6 in Rosenberg *et al.* 2004, and as a consequence it also generalizes the original existence result of Aumann and Maschler (1995) for the value of (non stochastic) repeated games with incomplete information on one side and perfect monitoring.

Notice that our result does not imply the existence of the value for models when player 1 receives signals without having a perfect knowledge of the belief of player 2 on the state (see Aumann Maschler 1995, or Zamir, 1992 for repeated with lack of information on one side, or Neyman 2008 for Markov chain repeated games with lack of information on one side). When the state is uncontrolled, more

flexibility on the signalling structure can be allowed.

### 4.3.4  Open problems.

1. We have seen that hypotheses HA' and HB' can not be withdrawn in theorem 4.4. However, strengthening HB' into HB may allow to weaken HA' into the following hypothesis.

Hypothesis HA" : Player 1 is more informed than player 2, i.e. there exists a mapping $\hat{d} : C \longrightarrow D$ such that : if $E$ denotes $\{(k, c, d) \in K \times C \times D, \hat{d}(c) = d\}$, we have : $\pi(E) = 1$, and $q(k, i, j)(E) = 1$, $\forall (k, i, j) \in K \times I \times J$.

If we only assume that HA" and HB' hold, it may be the case that player 2 controls player 1's signal, hence in some sense player 2 may "manipulate" player 1's knowledge of the state. So our proof does not apply here, and in our opinion most likely the value may fail to exist.

The situation is different if we assume HA" and HB. Player 1 always have more information than player 2 about the state, but the set $X = \Delta(K)$ is not sufficient to characterize, after each stage, the difference of information from player 1 to player 2. The natural state space here may rather be the set $\{(u, v) \in \Delta_f(X) \times \Delta_f(X), u \preceq v\}$. Does the uniform value exist in this case ?

2. In general, recall that lemma 4.6 is true without hypotheses, so in particular the $n$ stage values $v_n$ always exist. There is no known example of a zero-sum repeated game (defined with finite data exactly as in subsection 4.1) where $\lim_n v_n$ does not exist.

3. Assume that player 1 always has more information than player 2, i.e. that player 1 can deduce from his signal both the signal and the action of player 2. This is the case, e.g., when HA holds. It has been conjectured by Mertens, Sorin and Zamir (see Mertens, 1986, p.1572, Sorin, 2002, 6.5.8. p.147, or Mertens et al, 1994, Part C, p. 451) that for such repeated games, the limit of $(v_n(\pi))_n$ exists and can be guaranteed by player 1 in $\Gamma(\pi)$.

The approach used here might help to prove the conjecture. An important step would be to obtain an analog of the result on dynamic programming (Renault, 2011) for two-player stochastic games with deterministic transitions and action-independent payoffs. More precisely, let $Z$ be a state space, $A$ and $B$ be action sets, $r$ be a payoff function from $Z$ to $[0, 1]$, and $f$ be a transition from $Z \times A \times B$ to $Z$. At each stage, if the current state is $z$ simultaneously player 1 chooses $a$ and player 2 chooses $b$. Player 1's payoff is $r(z)$, and $(a, b)$ and the new state $f(z, a, b)$ are publicly announced. $Z$ being a precompact metric space, can we find "nice" uniform equicontinuity conditions on some auxiliary value functions that will ensure the existence [5] of the uniform value ?

---

5. Mertens and Neyman (1981) already provided conditions on the discounted values which

4. Existence of optimal/stationary strategies. Our proof mainly deals with the value functions and does not tell much regarding the existence of optimal and/or stationary strategies for the players. We just obtained the existence of 0-optimal strategies of player 2 in repeated games with an informed controller and in the stochastic games of section 3. Regarding player 1, since our existence result in dynamic programming does not provide the existence of a 0-optimal strategy for player 1 (see counter example 5.2. in Renault, 2011), we do not obtain such an existence result in the stochastic games of section 3 and consequently in repeated games with an informed controller. One just obtains the existence of stationary $\varepsilon$-optimal strategies for player 1, for each $\varepsilon > 0$ (see remark 5.3 in Renault 2011 or the proof of theorem 2 in Rosenberg *et al.*, 2002).

However, several positive results in this direction can be found in the litterature. For stochastic games with complete information and perfect information in which player 1 controls the transitions, Filar (1981) as well as Parathasarathy and Raghavan (1981) proved that if the number of states is finite then both players have optimal stationary strategies. This is also clearly the case if one considers the stochastic games of level 2 obtained from standard repeated games with lack of information on one side *a la* Aumann and Maschler. Moreover, Rosenberg *et al.* proved the existence of a 0-optimal strategy for player 1 in stochastic games with a single controller and incomplete information on the side of his opponent (2004). Neyman (2008) proved the existence of a 0-optimal strategy for player 1 in Markov chain repeated games with lack of information on one side. Consequently, it may well be the case that player 1 has a stationary 0-optimal strategy for the stochastic games of section 3, and a 0-optimal strategy for repeated games with an informed controller. And player 2 may have a stationary optimal strategy in the stochastic games of section 3.

―――――――――――――――――

guarantee the existence of the uniform value in stochastic games, but it seems difficult to apply them here.

# References.

Aumann, R.J. and M. Maschler (1995) : Repeated games with incomplete information. With the collaboration of R. Stearns. Cambridge, MA : MIT Press.

Ash, R.B. (1972) : Real Analysis and Probability, Probability and Mathematical Statistics, Academic Press.

Coulomb, J.M. (2003) : Games with a recursive structure. based on a lecture of J-F. Mertens. Chapter 28, Stochastic Games and Applications, A. Neyman and S. Sorin eds, Kluwer Academic Publishers.

Doob, J.L. (1994) : Measure Theory. Springer-Verlag, New-York.

Filar, J. (1981) : Ordered field property for stochastic games when the player who controls transitions changes from state to state. Journal of Optimization Theory and Applications, 4, 513-515.

Hörner J., D. Rosenberg, E. Solan and N. Vieille (2010) : Markov Games with One-Sided Information. Operations Research, 58, 1107-1115.

Malliavin, P. (1995) : Integration and probability. Springer-Verlag, New-York.

Mertens, J-F. (1986) Repeated games. *Proceedings of the International Congress of Mathematicians, Berkeley 1986*, 1528–1577. *American Mathematical Society*, 1987.

Mertens, J-F. and A. Neyman (1981) : Stochastic games. International Journal of Game Theory, 1, 39-64.

Mertens, J-F., S. Sorin and S. Zamir (1994) : Repeated Games. Parts A, B and C. CORE Discussion Papers 9420, 9421 and 9422.

Meyer, P.A. (1966) : Probabilités et potentiel. Hermann.

Neyman, A. (2008) : Existence of Optimal Strategies in Markov Games with Incomplete Information. International Journal of Game Theory, 581-596.

Parthasarathy, T and T.E.S. Raghavan (1981) : An orderfield property for stochastic games when one player controls transition probabilities. Journal of Optimization Theory and Applications, 33, 375-392.

Renault, J. (2006) : The value of Markov chain games with lack of information on one side. Mathematics of Operations Research, 31, 490-512.

Renault, J. (2011) : Uniform value in Dynamic Programming. Journal of the European Mathematical Society, vol. 13, p.309-330.

Rosenberg, D., Solan, E. and N. Vieille (2002) : Blackwell Optimality in Markov Decision Processes with Partial Observation. The Annals of Statisitics, 30, 1178-1193.

Rosenberg, D., Solan, E. and N. Vieille (2004) : Stochastic games with a single controller and incomplete information. SIAM Journal on Control and Optimization, 43, 86-110.

Sorin, S. (1984) : Big match with lack of information on one side (Part I), International Journal of Game Theory, 13, 201-255.

Sorin, S. (2002) : A first course on Zero-Sum repeated games. Mathématiques et Applications, SMAI, Springer.

Sorin, S. and S. Zamir (1985) : A 2-person game with lack of information on 1 and 1/2 sides. Mathematics of Operations Research, 10, 17-23.

Zamir, S. (1992) : Repeated Games of Incomplete Information : Zero-Sum. Handbook of Game Theory with Economic Applications, edited by Aumann and Hart. Vol I, chapter 6.