A Folk Theorem For Stochastic Games with Finite Horizon.

Chantal Marlats*

May 22 2009

Abstract

This paper gives a limit characterization of the set of perfect equilibrium payoffs for a class of stochastic games with finite horizon. If asymptotic conditions à la Dutta (1995) hold and if at some particular states the set of SPE payoffs in a finite truncation of the stochastic game is sufficiently rich, then any payoff vector in asymptotic game Pareto-Dominating the minimax point is approximated by a perfect equilibrium payoff in the finitely repeated game, provided that the horizon is sufficiently large.

1 Introduction

In stochastic games stage game payoffs are determined by the realization of a state. The distribution over states is determined by the actions of the players. This class of games has natural applications in economic environments with history dependent preferences. Furthermore, they find a new source of interest in behavioral economics : they provide a framework to model interactions with non-standard preferences, for instance addiction or intrinsic reciprocity. An agent is said to have intrinsic reciprocity when he chooses a *costly* action to increase (resp. decrease) the payoff of an opponent who kindly (resp. unkindly) behaves. In static settings this type of preferences may be modeled by an utility function that places a weight on an opponent's payoffs which increases if and only if the opponent chooses a nicer strategy (Segal and Sobel (2007)). In dynamic settings, one

*Paris School of Economics; Université Paris Panthéon-Sorbonne, CES 106-112 boulevard de l'Hopital 75013 Paris. Tel : (0033) 1 44 07 82 31. E-mail: cmarlats@univ-paris1.fr.

may consider this weight depending on the number of times where the opponent chooses to cooperate. In other words, the preferences are endogenous with respect to the past actions. Consequently, if cooperation have been played sufficiently often, the incentive to deviate from this very action is reduced. Then, a stochastic game whose matrix payoffs is initially a prisoner dilemma may evolve in such a way that after a certain date the stage games are coordination games, i.e. games where cooperation is effectively an equilibrium.

In this paper we address the question of the characterization of the set of Subgame Perfect Equilibrium (SPE) payoffs in finitely repeated stochastic games. Dutta (1995) identifies a class of stochastic games with infinite horizon for which the folk theorem holds. The proof of this result raises an additional difficulty in comparison to non-stochastic games : Not only players may want to deviate in order to improve their current gains but also to influence the probability distribution on states. For this reason, asymptotic state invariance of feasible long-run payoffs and dynamic minimax level assumptions are needed. Dutta constructs a punishment strategy such that for all individually rational payoffs there is a subgame perfect strategy whose payoff approaches the former. The equilibrium punishment strategy consists of minmaxing the deviator during several periods and then playing a strategy that rewards players who punished during the previous phase.

In the finite horizon case, these punishment strategies may not be credible. As the rewarding phase must be sufficiently long to compensate the loss endured during the minmax phase, if a player deviates at a late date, the punishment does not occur. Indeed the punishment strategy is credible if there exists SPE strategies played at the end of the game that give a perfect threat. Under the sole assumptions made in Dutta (1995) the characterization of equilibrium payoffs for infinitely repeated stochastic game does not hold when time is finite. The reason is that in some stochastic games it is impossible to construct perfect threat strategies. Consequently, late deviations may occur. Benoit and Krishna (1985) show that, for non-stochastic games, if the stage game admits Nash equilibria that give distinct payoffs then there exist perfect threat strategies. The idea is the following : under this condition it is possible for each player to play two series of Nash equilibrium that give distinct payoffs. We simply call them good Nash and bad Nash strategies. In non stochastic games, the conjunction of SPE strategies is still a SPE. Let M

be the maximal gain generated by a late deviation during the first punishment phases. By repeating these Nash strategies a sufficient large number of time, the difference between the payoffs increase by more than M. Consequently, the play of a series of bad Nash strategies is a credible threat to late deviations during the reward phase. Nevertheless in stochastic game the strategies consisting of playing a Nash equilibrium at each stage game may not be a perfect threat strategies. The reason is that players may have incentives to deviate from a Nash equilibrium at some states in order to change, in a profitable way, the probability distribution on states. Indeed, conditions under which the conjunction of SPE strategies is still a SPE and then the existence of perfect threat strategies are delicate issues in endogenously changing environments.

The purpose of this paper is to identify conditions for a limit folk theorem for stochastic games with finite horizon to hold. The main result shows that under the asymptotic state invariance assumptions of Dutta (1995) and an additional assumption on the richness of the set SPE payoffs feasible at some *particular* initial states, the set of the individually rational feasible payoffs of the asymptotic game (i.e. infinitely repeated game) can be approximated by a SPE payoffs in the finitely repeated game. First, under the previous hypothesis it is proved that in stochastic games with large but finite horizon of T, a payoff in the convex hull of rational outcomes generated by a strategy that only depends on states (i.e. stationary Markov strategy) can be maintained as a SPE payoff in all but the last L periods. Indeed, these last periods correspond to the perfect threat. This result settles on the existence of punishment stategy, called a four phases punishments. The length of the last periods can be fixed independently of the overall length of the game. Using the result due to Dutta (1995) according to which if the length of the game is sufficiently large then a feasible payoff can be approximated by a payoff in the convex hull of stationary Markovian outcomes, then we show that any individually rational feasible payoffs of the asymptotic game is approximated by an SPE payoffs. Furthermore, when the horizon is sufficiently large the set of SPE payoffs is contained in a neighborhood of the set if the rational and feasible payoffs in the asymptotic game.

The rest of the paper is organized as follows. Section 2 presents the basic setting of the model. Section 3 defines the notion of individually rational feasible payoff in stochastic

games and states related preliminary results. Section 4 describes the main assumptions and their implications. The main result is presented in section 5. Section 6 and section 7 respectively describe the first and the second steps of the proof discussed above.

2 The Model.

Let Γ^T be a stochastic game of length T with n players (indexed by $i = 1, ..., n$). At each period t a state s in S is drawn and each player i chooses an action a_i in action space A_i which is the same in every state. The sets S and $A = \prod_{i=1}^{n} A_i$ are finite. The set of player i's mixed strategies is denoted by $\Delta(A_i)$. The one period utility function is $u_i : A \times S \to \mathbb{R}$. The probability distribution over the set of the states depends on the current state and on the action profile. The dynamic of the game is described by transition function $q : S \times A \to \Delta(S)$. The initial state is denoted by s. A $t-$history is vector of the realized states and the realized actions until date t and is denoted $z^t = (s^1, a^1, ..., s^t, , a^t) \in Z^t$ where $s = s^1$ and Z^t is the set of all $t-$histories. In a game of length T behavioral strategy is denoted by $\sigma_i = (\sigma_i^t)_{t=1}^T$, where $\sigma_i^t : Z^t \to \Delta(A_i)$. Let Σ_i be player i's set of strategies and Σ the set of strategy profiles. Let $Z^t(z^\tau)$ be the set of all the possible histories at date t given that z^τ is realized. A strategy and an initial state induce a probability on the set of histories donoted by $\Pr_{\sigma,s}$. A continuation strategy at z^τ is $\sigma_i|_{z^\tau} = (\sigma_i^t|_{z^\tau})_{t=\tau+1}^T$, where $\sigma_i^t|_{z^\tau} : Z^t(z_\tau) \to \Delta(A_i)$. A pure stationary Markov strategy is a strategy that depends only on the current state, whatever the date. Let G be the set of pure stationary Markov strategies. An almost stationary Markov strategy consists of playing a stationary Markov strategy until deviation occurs and then a possibly history dependent punishment strategy. If σ is an almost stationary Markov strategy then we denote by x_σ the stationary Markov strategy played until deviation. The set $\mathcal{A}_\sigma^t(s)$ contains all the states that can be reached with a positive probability (or equivalently all the accessible states) at t given an initial state s and a strategy σ.

Let $v_i^t(\sigma, s)$ be the expected returns at $t \leq T$ given the strategy profile σ and the initial state s. Given an initial state s and a strategy σ, the undiscounted time average payoff

(or $T-$average payoff) is :

$$U_i^T(\sigma, s) = \frac{1}{T}\sum_{t=1}^{t=T} \upsilon_i^t(\sigma, s).$$

For a given z^τ, the continuation time average payoffs from date τ is

$$U_i^{T-\tau}(\sigma|_{z^\tau}, s^\tau) = \frac{1}{T-\tau}\sum_{t=\tau+1}^{T} \upsilon_i^t(\sigma|_{z^\tau}, s^\tau)$$

The concatenation of a strategy σ_T and $\sigma_{T'}$ is denoted $\sigma_T\sigma_{T'}$. In words the notation $\sigma_T\sigma_{T'}$ refers to the strategy that consists of playing σ_T over T periods and then $\sigma_{T'}$ over T' periods. The notation $\Gamma^{T+T'}(s)$ refers to the stochastic game repeated $T+T'$ times with payoffs given by $U_i^{T+T'}(\sigma_T\sigma_{T'}, s) = \frac{1}{T+T'}\sum_{t=1}^{t=T+T'} \upsilon_i^t(\sigma_T\sigma_{T'}, s)$. A Nash Equilibrium in $\Gamma^T(s)$ (the stochastic game Γ^T with initial state s) is a strategy profile σ such that for all $i \in I$,all $\sigma_i' \in \Sigma_i, U_i^T(\sigma, s) \geq U_i^T(\sigma_i', \sigma_{-i}, s)$. A strategy profile is a SPE if for all $\tau \leq T$, all $i = (1, ...n)$,all $\sigma_i' \in \Sigma_i$,for all $z_\tau \in Z_\tau$,

$$U_i^{T-t}(\sigma|_{z_t}, s) \geq U_i^{T-t}(\sigma_i'|_{z_t}, \sigma_{-i}|_{z_t}, s)$$

Notice that the set on SPE may change across initial states and length of play. Let $P^T(s)$ be the set of perfect equilibrium payoffs of $\Gamma^T(s)$. The notation $\sigma \in P^T(s)$ means the equilibrium path is given by σ. By a slight abuse of notation we write $\sigma \in P^T(C)$ when the strategy whose equilibrium path is given by σ is an equilibrium in $\Gamma^T(s)$ for all $s \in C$. For all T let $w_i^T(.)$ be the worst SPE payoffs from the i's point view in the stochastic game $\Gamma^T(.)$. This payoff is called the optimal $T-$periods punishment for i (c.f. Abreu(1983)). Let $\omega_i = \max_{t\in\mathbb{N}} \max_{s\in S} w_i^t(s)$ be the best optimal punishment across time and states. Let $||.||$ be any one the equivalent norm in $\mathbb{R}^n$. Players are assumed to play according to public random device.

3 Individually rational feasible Payoffs.

In stochastic games, payoffs and continuation payoffs vary from a state to an other. Consequently, the set of feasible payoffs and the minimax may vary across states and game horizons. For these reasons, the set of feasible payoffs and the minmax will be expressed

in terms of time average payoffs. Typically, in stochastic games the set of strategies is very large. This complexity is relaxed when one assume that players use stationary Markov strategies. Several results in the stochastic game literature show that it is not restrictive asymptotically (cf. Blackwell (1965), Dutta (1995)). In this section individual rationality will be defined in terms of behavioral strategies and stationary Markov strategies. Some results on the relationship between these two notions will be presented.

3.1 Feasible payoffs.

The set of feasible time average payoffs with initial publicly randomization of the stochastic game of length T when the initial state is s is given by :

$$F^T(s) = \{v \in \mathbb{R}^n : \forall i = (1,..,n) \exists \beta^k \geq 0, \sum_k \beta^k = 1 \text{ and } \sigma^k \in \Sigma_T \text{ s.t. } \sum_k \beta^k U_i^T(\sigma^k, s) = v_i\}$$

In a stochastic game of length T and initial state s let $\phi^T(s)$ denote the set of feasible time average payoffs when only stationary Markov strategies are played and $co\phi^T(s)$ be his convex hull.

$$co\phi^T(s) = \{v \in \mathbb{R}^n : \forall i = (1,..,n) \exists \beta^k \geq 0, \sum_k \beta^k = 1 \text{ and } g^k \in G \text{ s.t. } \sum_k \beta^k U_i^T(g^k, s) = v_i\}$$

The asymptotic version of a stochastic game $\Gamma^T(s)$, that is when $T \to \infty$, is denoted by $\Gamma(s)$. Let $F(s)$ denote the set of publicly randomized long run feasible average payoffs of the a stochastic game :

$$F(s) = \{v \in \mathbb{R}^n : \forall i = (1,..,n) \exists \beta^k \geq 0, \sum_k \beta^k = 1 \text{ and } \sigma^k \in \Sigma^T \text{s.t.} \liminf_{T\to\infty} \sum_k \beta^k U_i^T(\sigma^k, s) = v_i\}$$

In the rest of the paper the payoff given by $\liminf_{T\to\infty} U_i^{T_r}(\sigma_T, s)$ is denoted by $U_i(\sigma, s)$.

Lemma 1 *For all $s \in S, F(s)$ is non empty.*

PROOF : It suffices to show that $\forall\ s \in S$, $\forall \sigma \in \Sigma_T$ there exists a sequence $\{T_r\}_{r>0}$ such that $U_i^{T_r}(s,\sigma) \to v$ as $r \to \infty$. Recall that S and A are finite so $\min_{s,a} \underline{u}_i(a,s) = \underline{u}_i$ and $\max_{s,a} \bar{u}_i(a,s) = \bar{u}_i$. Fix T and s and remark that $\frac{1}{T}\sum_{t=1}^{t=T} \min_\sigma v_i^t(\sigma,s) \geq \min_{a\in A} \frac{1}{T}\sum_{t=1}^{t=T} \underline{u}_i(a,s) \geq \underline{u}_i$. Similarly $\frac{1}{T}\sum_{t=1}^{t=T} \max_\sigma v_i^t(\sigma,s) \leq \bar{u}_i$. Fix $s \in S$ and $\sigma \in \Sigma_T$. By the previous argument, $\bar{u}_i \geq U_i^T(\sigma,s) \geq \underline{u}_i$ for all $T < \infty$. Then the sequence $\{U_i^T(\sigma,s)\}_{T>0}$ take values in the compact set $[\underline{u}_i, \bar{u}_i]$. So there exist a subsequence of finite horizons $\{T_r\}_{r>0}$ and a $v \in \mathbb{R}^n$ such that $U_i^{T_r}(\sigma,s) \to v$ as $r \to \infty$.

■

Let $\phi(s)$ denotes the set of long run feasible average payoffs of the infinitely repeated stochastic game when only pure stationary Markov strategies are played.

$$co\phi(s) = \{v \in \mathbb{R}^2 : \forall i = (1,..,n) \exists \beta^k \geq 0, \sum_k \beta^k = 1 \text{ and } g^k \in G \text{ s.t.} \sum_k \beta^k U_i(g^k,s) = v_i\}$$

The following lemmata recall some useful result in stochastic game. The first states that for all $s \in S$, $\lim_{T\to\infty} 1/T \times U_i^T(g,s)$ exists. This fact is used by Dutta (1995) to prove that a feasible long run average payoffs can be realized by a public randomization over pure stationary Markov strategies.

Lemma 2 *$\forall\ s \in S, \forall v \in \phi(s)$, there is a $g \in G$ such that $\lim_{T\to\infty} U_i^T(g,s) = v$. Consequently, the set $\phi(s)$ is non empty.*

PROOF : (c.f. Dutta (1995)).

Lemma 3 *For all $\forall\ s \in S$, $F(s) = co\phi(s)$*

PROOF : (c.f. Dutta (1995)).

A direct implication of this lemma is the following corollary :

Corollary 1 *$\forall\ s \in S, \forall\ v \in F(s)$, $\forall\ \epsilon > 0$, there exists a $T(\epsilon)$ such that for all $T \geq T(\epsilon)$, there is a $v^T \in co\phi^T(s)$ such that $||v - v^T|| \leq \epsilon$.*

PROOF: Fix s, $\epsilon > 0$ and v. By lemma3 if $v \in F(s)$ then $v \in co\phi^T(s)$. Take $\beta^k \geq 0, \sum_{k=1}^{K} \beta^k = 1$ and $\{g^k\}_{k=1}^{k=K}$ such that $v = \sum_k \beta^k U_i(g^k, s)$. A consequence of lemma2 is that $\lim_{T\to\infty} \frac{1}{T} \sum_k \beta^k U_i^T(g^k, s)$ exists. By definition of a limit, there exists a $T(\epsilon)$ such that if $T \geq T(\epsilon)$ then there is a $v^T \in co\phi^T(s)$ that gives within ϵ of v.

■

3.2 Individual rationality.

Because the minmax depend on the initial state and on the length of the game it seems reasonable to define it in terms of time average payoffs : $m_i^T(s) = \inf_{\sigma_{-i}} \sup_{\sigma_i} U_i^T(\sigma_i, \sigma_j, s)$. Also when T tends to infinity the minmax is given by $m_i(s) = \inf_{\sigma_{-i,}} \sup_{\sigma_{i,}} \liminf_{T\to\infty} U_i^T(\sigma_{i,}, \sigma_{-i}, s)$. The following lemma states the relationship between the minmax in finitely and infinitely repeated game : it says that when player i's is minmaxed by his opponent for a finite period T he can not expect more than his long term minmax plus a gain that decreases with T.

Lemma 4 *For $\eta > 0$, there is a σ_{-i} and a $T(\eta) < \infty$ so that if $T \geq T(\eta)$ then for all σ_i*

.

$$m_i^T(s) \leq U_i^T(\sigma_i, \sigma_{-i}, s) \leq \eta + m_i(s) \text{ for all } s \in S$$

PROOF: First, $U_i^T(\sigma_i, \sigma_{-i}, s)$ is necessarily greater or equal to $m_i^T(s)$ because it is the minimal payoff that i can expect when he plays a best response. Fix $\eta > 0$. By the Proposition 3 in Dutta (1995), there exists a σ_{-i} and a $T(\eta)$, such that if $T \geq T(\eta)$, it gives for all σ_i : $U_i^T(\sigma_i, \sigma_{-i}, s) - \lim_{\delta\to 1} m_i(s, \delta) \leq \eta$ where $m_i(s, \delta)$ is the minmax calculated with discounted infinite average payoff (the parameter $\delta \in (0, 1)$ is the discount factor). Dutta also shows that $\lim_{\delta\to 1} m_i(\delta, s) = m_i(s)$. Then $U_i^T(\sigma_i, \sigma_{-i}, s) - m_i(s) \leq \eta$.

■

Dutta (1995) points out that the definition of an individually rational payoffs is subtle because the continuation payoffs for given strategy and initial state vary across histories. Two definitions, one in ex-ante terms and the other in ex-post terms, make sense. In a stochastic game Γ^T with an initial state s a $T-$average payoff will be said to be ex-ante individually rational if it is greater than $m_i^T(s)$. The set of ex-ante individually rational

payoffs is

$$F^T(s)^* = \{v \in F^T(s) : \forall i = (1,..,n), \exists \sigma \text{ s.t. } U_i^T(\sigma, s) = v_i > m_i^T(s)\}$$

In the asymptotic version of the stochastic game, this set is given by : $F(s)^* = \{v \in F(s) :$ for all $\forall i = (1,..,n)$, $\exists$ σ such that $\liminf_{T\to\infty} U^T(\sigma, s) = v_i > m_i(s)\}$. The set $co\phi^T(s)^*$ (resp. $co\phi(s)^*$) defines in a similar fashion the set of individually rational $T-$average payoffs realizable by pure stationary Markov strategies in the finite (resp. asymptotic) stochastic game. On the other hand if for a given strategy the continuation payoffs after every histories are ex-ante individually rational, the $T-$average payoff associated with this strategy is ex-post individually rational. Formally the set of ex-post payoffs is denoted by $F^T(s)^{**} = \{v \in F^T(s) : \forall i = (1,..,n),\ \exists\ \sigma$ s.t. $\forall 0 < t < T, \forall z_t;\ U_i^{T-t}(\sigma|_{z^t}, s^t) = v_i > m_i^{T-t}(s^t)\}$. Let $F(s)^{**}$ be the ex-post individually rational payoffs in the asymptotic game.

4 Assumptions on payoffs and transition function.

Under widely different strategic stakes across states the folk theorem may fail in infinite stochastic game. Dutta (1995) overcomes this problem by assuming asymptotic invariance across states of individually rational feasible payoffs and minmax level. Nevertheless in a finitely repeated stochastic games, this problem is more stringent because these asymptotic conditions over payoffs are not sufficient. Furthermore, as we discussed in the introduction the folk theorem doesn't hold for finite horizon if the payoffs possibilities are such that it is impossible to play a perfect threat strategy at the end of the game. An additional assumption will allow us to bypass these difficulties.

4.1 Asymptotic invariance assumptions.

We start to present asymptotic state independence assumptions and their consequences for stochastic game with finite but large horizon. First, long run feasible payoffs are supposed to be invariant across states.

Assumption 1 *For all* $s, s' \in S$, $F(s) = F(s') = F$.

Recall that all feasible long run payoffs can be replicated by an initial one shot public randomization over pure stationary Markov strategies. Nevertheless, it is possible that for some histories the continuation payoffs differ considerably. Dutta shows that if $A1$ holds, then for all one shot randomization, there exists a cycle of pure stationary Markov strategies that approximates it. An interesting feature for this pure strategy give an approximation of a payoff in F whatever the initial state. The following lemma is a variation of lemma 6 on Dutta (1995) and proves the existence on the strategies presented above.

Lemma 5 *Assume that A1 holds.* $\forall\ v \in F, \forall\ s \in S$ *and* $\forall\ \epsilon > 0$, *there is a* $T(\epsilon) < \infty$, *so that if* $T \geq T(\epsilon)$ *then there exists a strategy whose* $T-$*average payoffs* $v^T \in co\phi^{*T}(s)$ *are within* ϵ *of* v *after all the histories.*

PROOF: Pick $v \in F$. We know by lemma 3 that $v \in co\phi(s)$ for all s. In words there exist publicly randomized pure stationary Markov strategies that duplicates long run feasible payoffs. Let $\{\lambda_k\}$ be the probability weights and g^k a pure stationary Markov strategy with payoffs v_k such that $v = \sum_{k=1}^{K} \lambda_k v_k$. A time cycle stationary Markov strategy σ^*consists of playing g^1 for $\gamma_1$periods, then g^2 for $\gamma_2$periods and ending by playing g^K for γ_K periods. The strategy spreads over $\gamma = \sum_{k=1}^{K} \gamma_k$ periods. Choose g^1,..., g^K the pure stationary Markov strategies that give the payoffs $v_1, ..., v_K$ defined above. Let $\bar{\epsilon}$ be the smaller real positive number $v_i + \bar{\epsilon} > m_i + \bar{\epsilon}$.

Let $\epsilon \leq \bar{\epsilon}$. By corollary 1 we know that for all $i = 1, ..., n$ and $s \in S$ there exist $\gamma_1(\epsilon), ...,\ \gamma_K(\epsilon)$ such that for all $K \geq k \geq 1$ if $\gamma_k \geq \gamma_k(\epsilon)$ then: $|U_i^{\gamma_k}(g^k, s)\ - v_k| \leq \epsilon$ (where $U_i^{\gamma_k}(g^k, s) \in \phi^{\gamma_k}(s)$denotes the γ_k-average payoffs received when g^k is played). Consider first the case where: $U_i^{\gamma_k}(g^k, s) > v_k\ - \epsilon$. For all $K \geq k \geq 1$ choose $\gamma_k \geq \gamma_k(\epsilon)$, so that γ_k/γ such that $\gamma_k/\gamma \geq \lambda_k -\ \epsilon'$with $\epsilon' \to 0$. Then $\sum_{k=1}^{K} \gamma_k/\gamma U_i^{\gamma_k}(g^k, s) \geq \sum_{k=1}^{K} (\lambda_k - \epsilon') U_i^{\gamma_k}(g^k, s)$. As $\epsilon' \to 0$ we have $\sum_{k=1}^{K} \lambda_k U_i^{\gamma_k}(g^k, s) > \sum_{k=1}^{K} \lambda_k (U_i^{\gamma_k}(g^k, s) - \epsilon) > \sum_{k=1}^{K} \lambda_k v_k - \epsilon$.

Second consider the case where: $U_i^{\gamma_k}(g^k, s) \leq v_k + \epsilon$. For all $K \geq k \geq 1$ take $\gamma_k \geq \gamma_k(\epsilon)$, so that γ_k/γ such that $\gamma_k/\gamma \leq \lambda_k +\ \epsilon'$. Similarly we show that $\sum_{k=1}^{K} \lambda_k U_i^{\gamma_k}(g^k, s) \leq \sum_{k=1}^{K} \lambda_k (U_i^{\gamma_k}(g^k, s) + \epsilon) \leq \sum_{k=1}^{K} \lambda_k v_k + \epsilon$ with $\epsilon' \to 0$. By construction the payoff given by this strategy is in $co\phi^T(s)$. Because ϵ is small the payoff is greater than $v_i + \epsilon$ and so

$m_i + \epsilon$. By lemma 4 it is also a rational payoff ($m_i + \epsilon > m_i^T(s)$). Consequently it is in $co\phi^{T*}(s)$. For $\epsilon \geq \bar{\epsilon}$ take $\gamma_k(\epsilon) = \gamma_k(\bar{\epsilon})$ for all $K \geq k \geq 1$.

The strategy σ^* is repeated after γ periods and the difference between $U_i^{\gamma}(g^k, s)$ and v_i cannot be greater ϵ, whatever the initial state. Replace $\gamma(\epsilon) = \sum_{k=1}^{K} \gamma_k(\epsilon)$ by $T(\epsilon)$. To sum up we have shown that there is a $T(\epsilon) < \infty$, so that if $T \geq T(\epsilon)$ then the average payoffs associated with this strategy gives an approximation v and continuation payoffs don't largely vary across states. Under this strategy the only histories that matter are s^t, the state reached at date t. We conclude that this approximation holds after all histories.

■

The second assumption says that long run minmax payoffs are state independent.

Assumption 2 *For all $s \in S$, $m_i(s) = m_i$.*

The Dutta (1995)'s folk theorem for stochastic game with infinite horizon also relies on these assumptions. In his paper Dutta identifies particular general classes of game that verifies them. Communicating games (for any couple s, s' there is strategy so that $\Pr_{\sigma,s}(s^t = s' \text{ for some } t > \infty) > 0$), scrambling games (noisy law motion) or any game without non return states satisfy assumption 1. Assumption 2 is satisfied for common resource games with $(n-1)$ state controllability. To sum up the first condition says that the payoffs possibilities are the same at any initial state and the second that the support of the law motion is independent of the actions of any one player.

4.2 Distinct SPE payoffs.

We make an assumption of richness of the set of SPE payoffs in a finite truncation of the stochastic game starting at some particular states. It says the stochastic game has recurrent sets for strategies with distinct payoffs that are reacheable from any. As it will be make clear below these strategies can be repeated with worrying about deviation . This assumption will be used to construct strategies aimed at deterring deviation in late periods. Before presenting it formally we shall give further definitions. It is well known that when player plays stationary Markov strategy then the probability distribution on states is a Markov chain. A set of states S' is recurrent with respect to x if for all $s \in S'$:

- $\sum_{s' \in S'} Pr_{x,s}(s') = 1$
- $Pr_{x,s}(s^t = s' \text{for some } t) = 1$

Let $\mathcal{R}(x)$ be the collection of all the recurrent sets w.r.t. the stationary Markov strategy x.

Assumption 3 *There exist $n+1$ almost stationary Markov strategies, denoted by $\hat{\sigma}, \check{\sigma}^1, ..., \check{\sigma}^n$ and an integer $\tilde{T}$ so that:*

- *For all $s \in S$, $x = (x_{\hat{\sigma}}, x_{\check{\sigma}^1}, ..., x_{\check{\sigma}^n})$ there is $\tilde{s} \in \mathcal{R}(x)$, σ and t so that $\tilde{s} \in \mathcal{A}^t_\sigma(s)$.*
- *For all $s \in \mathcal{R}(x_{\hat{\sigma}})$, $s' \in \mathcal{R}(x_{\check{\sigma}^i})$ we have: $\hat{\sigma} \in P^{\tilde{T}}(s)$ and $\check{\sigma}^i \in P^{\tilde{T}}(s')$ where $U_i^{\tilde{T}}(\hat{\sigma}, s) > U_i^{\tilde{T}}(\check{\sigma}^i, s')$.*
- *For all $s \in \mathcal{R}(x_{\check{\sigma}^i})$, $U_i^{\tilde{T}}(\check{\sigma}^i, s) > (\geq)\omega_i$ (if $\mathcal{R}(x_{\hat{\sigma}}) = \mathcal{R}(x_{\check{\sigma}^i}) = S$)*

Notice that when the set of states is a singleton this assumption is weaker than the one of Benoit Krishna (1985). We give an example of a stochastic game that verifies it. Let the following 2×2 stage games define the stage game payoffs.

State s

	L	R
U	$-1, -1$	$-1, -1$
D	$2/3, 2/3^{(s')}$	$2/3, 2/3$

State s′

	L	R
U	$1, 1^{(s)}$	$0, 0$
D	$0, 0$	$2, 2^{(s)}$

The notation $^{(s')}$ means that when (D, L) is played, the subsequent stage game payoffs are given by s' with probability one. We show that, whatever the initial state, there are three SPE strategies $\hat{\sigma}$ and $\check{\sigma}^1$ and $\check{\sigma}^2$ in the 2 -periods stochastic game that give to each player a distinct payoff. Indeed, $\check{\sigma}^1$ and $\check{\sigma}^2$ will be identical. For this reason they will be denoted by $\check{\sigma}$ in the rest of the example. First, consider the strategy $\hat{\sigma}$: at $t = 1$ play $\hat{\sigma}(norm) : (s, s') \to ((D, L); (D, R))$. If deviation from $\hat{\sigma}(norm)$ then at $t = 2$ play $\hat{\sigma}$ $(dev) : (s, s') \to ((D, L); (U, L))$, otherwise play $\hat{\sigma}(norm)$. So the 2-average payoffs are:

$s^1 = s$

	$player1$	$Player2$
$Deviation$	$-1/6$	$2/3$
$No\ deviation$	$4/3$	$4/3$

$s^1 = s'$

	$player1$	$Player2$
$Deviation$	$1/2$	$1/2$
$No\ deviation$	$4/3$	$4/3$

We conclude that $\hat{\sigma}$ is effectively a SPE. Now consider the strategy $\check{\sigma} : \hat{\sigma}(norm) : (s, s') \to ((D, R); (U, L))$ after each history. The payoffs are :

$s^1 = s$

	$player1$	$Player2$
$Deviation$	$-1/6$	$2/3$
$No\ deviation$	$5/6$	$5/6$

$s^1 = s'$

	$player1$	$Player2$
$Deviation$	$1/2$	$1/2$
$No\ deviation$	$5/6$	$5/6$

We deduce that $\check{\sigma}$ an SPE strategies profile. We now check that the conditions states in the assumption hold. Condition (i) is trivially satisfied because $\mathcal{R}(x_{\hat{\sigma}}) = \mathcal{R}(x_{\check{\sigma}}) = S$. Also condition (ii) is satisfied $\hat{\sigma}$ and $\check{\sigma}$ are SPE strategies that give different payoffs to each player, whatever the initial state. Finally, for all t and all initial states, the worse $t-$periods equilibrium payoff cannot give more than $2/3$ to any player. Then the third requirement is satisfied.

To best of our knowledge this assumption doesn't refer to some particular class of stochastic games in the literature. It may be understood as general condition that ensure to any stochastic game to have perfect threats.

4.3 Dimensionality.

We assume that set of feasible long run average payoffs, namely F, satisfies the condition of full dimensionality.

Assumption 4 *The set F has a dimension of n.*

This assumption is the reformulation for stochastic settings of the standard full dimensionality condition (Fudenberg and Maskin (1994)). This assumption allows the construction of asymmetric punishment.

5 Equilibrium Characterization.

Let $w_i^T(s)$ be the worst SPE payoff from the i's point of view in the stochastic game $\Gamma^T(s)$. We start by a necessary and sufficient condition for a play path to be a SPE in Γ^T. We deduce from this result a interesting property of some particular SPE strategy

in the stochastic games begining at particular state. Finally we state the main result of this paper.

Lemma 6 *$\forall s^* \in S$ the profile $\tilde{\sigma} = \{\tilde{\sigma}^t\}_{t=1}^T \in P^T(s^*)$ if and only if for all $i = (1, ..., n)$ for all $t \leq T$, for all $s \in \mathcal{A}_{\tilde{\sigma}}^t(s^*)$ and all $d_i \in A_i$.*

$$(T-t)U_i^{T-t}(\tilde{\sigma}, s) \geq u_i(d_i, \tilde{\sigma}_{-i}^t, s) + (T-t-1)E(w_i^{T-t-1}(s')|d, s)(*)$$

PROOF: Fix a $t \leq T$ and a $s \in S$. The left hand side of the inequality is the player i's expected payoffs generated by $\tilde{\sigma}$ if no deviation occurs and the right hand side is his expected payoff when he deviates. The first element is the current gain from choosing d_i. The second can be interpreted as the expected utility when player i is punished. Recall that there are two motives of deviation in stochastic games : first, to improve the current gain and second to influence the probability distribution (a player may receive more by being punished at certain states than playing the normal path at other states). The current gain of deviation added to the expected punishment payoffs are smaller than the reward associated with the normal play, player i doesn't deviates at t. It follows that if the inequality holds for all i and all $t \leq T$ and all $s \in S$ then the normal play path $\tilde{\sigma}$ is a SPE.

■

The existence of a folk theorem for repeated game with finite horizon relies on a particulary property of these game, called by Benoit and Krishna (1985) the "conjunction property". This property says that for all $T_1$and T_2 if an outcome is an SPE in P^{T_1} and an other in P^{T_2} then the concatenation of these outcomes is still a SPE in P^T, where $T = T_1 + T_2$. Nevertheless, it does't hold in all the stochastic games. The failure of this property is due to incentives to deviate in order to improve distribution. For instance, suppose the following stochastic game with two players and two states, namely s and s'. Let the one shot payoffs defined by the the following matrix :

State s

	L	R
U	$1, 1$	$0, 0$
D	$0, 0^{(s')}$	$2, 2$

State s'

	L	R
U	$100, 1$	$0, 0$
D	$0, 0^{(s)}$	$200, 2$

Note that (U,L) is a SPE in $\Gamma^1(s)$ and (U,L) is an SPE payoffs in $\Gamma^1(s')$. We show that the 2-periods strategies profile ((U,L),(U,L)), is not an SPE payoffs in $\Gamma^2(s)$. If player 1 deviates then he can have at least a payoff of 100/2. If he plays ((U,L),(U,L)) then he receives a payoff of 1.

The following lemma gives a sufficient condition for the conjunction property to hold in stochastic game when the SPE path is given by a stationary Markov strategies. This result will be useful when we will construct perfect threat strategies.

Lemma 7 *Let σ be an almost stationary Markov strategy. Let C be a set of recurrent state w.r.t. x_σ. If there is an integer T such that $\sigma \in P^T(s)$ for all $s \in C$ and if*

$$U_i^T(\sigma, s) \geq \omega_i \text{ for all } s \in S$$

then $\sigma \in P^{T\times Q}(s)$ for all $s \in S$ and all $Q \in \mathbb{N}^*$..

PROOF: Fix a $\bar{Q} \in \mathbb{N}^*$ and denote by $\bar{T} = T\bar{Q}$. Take an integer $Q \leq \bar{Q}$ and suppose that at the beginning of the $Q - 1^{\text{th}}$ block the state is $s' \in C$. Take $t = TQ - m$ for all $m \in [1, T)$. As C is a recurrent set for x_σ, if players follow σ until date $\bar{T}$, player i gets at least:

$$mU_i^m(\sigma, s) + Q^* \min_{s\in C} TU_i^T(\sigma, s)$$

for all $s \in \mathcal{A}_\sigma^{T-m}(s')$. As $\sigma \in P^T(s')$ we have:

$$mU_i^m(\sigma, s) \geq u_i(d_i, a_{-i}(\sigma, s), s) + [m-1]E[w_i^m | d_i, a_{-i}(\sigma, s), s]$$

for all $d_i \neq a_i(\sigma, s)$. Also by assumption: $QTU_i^T(\sigma, s) \geq QT\omega_i$. Then:

$$\begin{aligned} mU_i^m(\sigma, s) + QT \min_{s\in C} U_i^T(\sigma, s) &\geq u_i(d_i, a_{-i}(\sigma, s), s) \\ &\quad + [m-1]E[w_i^m | d_i, a_{-i}(\sigma, s), s] + QT\omega_i \\ &\Leftrightarrow (\bar{T} - t)U_i^{\bar{T}-t}(\sigma, s) \geq u_i(d_i, a_{-i}(\sigma, s), s) \\ &\quad + (\bar{T} - t - 1)E(w_i^{\bar{T}-t-1}(s') | d, s) \end{aligned}$$

This proves that in the subgame that starts at date $t = TQ - m$ and at stage $s \in \mathcal{A}_\sigma^{T-m}(s')$, σ is a SPE strategy. As s', Q and m are arbitrary and because C is a recurrent set we conclude that it is an SPE in the game $\Gamma^{\bar{T}}(s)$ for all $s \in C$.

■

The main result gives a characterization of the equilibrium payoffs for a class of stochastic game with large but finite horizon.

Theorem 1 *Assume A1, A2, A3 and A4 hold. For any $v \in F^*, \forall s \in S, \forall\ \epsilon > 0$, there is a $T^* < \infty$ such that if $T \geq T^*$ then there exists a Subgame Perfect Equilibrium strategy whose payoff is within ϵ of v, after all the histories reached with a positive probability.*

Furthermore, $\forall s \in S$, $\forall \epsilon > 0$ and $\forall u \in \liminf_{T\to\infty} P(s)$ there exists a $v \in F^*$ within ϵ of u.

In words, if v is an individually rational payoff of an asymptotic stochastic game that satisfies the above assumptions then it can be approximated by a SPE strategy in a large but finite horizon version of this stochastic game. Let us give an intuition of the proof. We decompose a fixed T in two periods, T' and L, such that $T = T' + L$. Fix $v \in F^*$, σ a cycle of pure Markov strategies so that $U^{T'}(\sigma, s) = v^{T'}$ can be made arbitrarily close to v when the T' is large. First we show that if T' and L are sufficiently large then it is possible to find a punishment in the stochastic game $\Gamma^T(s)$ that maintained $v^{T'}$ as a SPE payoff in all but the last L periods. We show that the length of the last phases, L can be fixed independently of T'. Then, as $T' \to \infty$, the equilibrium payoffs will be arbitrarily close to v.

6 Four phase punishments

Let us start this section by giving an outline of the punishment strategy that will be used to prove our main result. To simplify the exposition assume that player i is the last player who has deviated. During the first phase, player i is minmaxed (in term of long run average payoffs) by the his opponents for R_1 periods. Player i is in best response during this phase but the other players may not. In order to reward their compliance, the R_1 periods are followed by a reward phase where the average payoffs are such that

each player receives more when he is punishing than if he were punished. This phase lasts R_2 periods. As the horizon is finite, players may want to deviate at the last periods because the remaining time is not sufficient for a punishment to be credible. In order to deter late deviations, the last periods will be dedicated to the play of a perfect threat strategy. Indeed if player j deviates from his punishment strategy during the first phases, he will endure a punishment during the last two phases whose cost will not be enough to be compensated by the gain from deviation. We show that under assumption 3 there exist perfect threat strategies. Because the state at the end of the reward phase may not be one of this set, a transition phase will ensure the connection. This section is aimed at proving the following proposition. This result constitutes the first main step of the proof of theorem 1.

Proposition 1 *Assume that A1-A4 hold. Fix* $s \in S, v \in F^*$, $i = (1, ..., n)$. *There are a* $\bar{T}'$ *and a* $\bar{L}$ *so that if* $T' \geq \bar{T}'$ *and* $L \geq \bar{L}$ *then there is pure strategy* $\tilde{\sigma}$, *so that* $\liminf_{T' \to \infty} U_i^{T'}(\tilde{\sigma}, s) = v$, *and strategy* α *such that* $U^{T'+L}(\tilde{\sigma}_{T'} \alpha_L, s) \in P^{T'+L}(s)$.

This proposition states that for sufficiently large values of T' and L there exists a SPE strategy (indeed for a four phase punishment) in $\Gamma^{T'+L}(s)$ that maintains $U_i^{T'}(\tilde{\sigma}, s)$, a payoffs given by a cycle of pure Markov strategy that can approximates $v \in F^*$, as an equilibrium payoff in all but the last L periods.

6.1 Reward strategies.

In a stochastic game Γ^T, let $\tilde{\sigma}_T$ denote the normal play path, i.e. the strategy played until player i deviates. As we discussed above the reward phase starts at the end of the R_1 periods of the minmax phase. Following Dutta, we construct strategies that give strictly different payoffs to player j according that the punishment is directed against him or not. More precisely, a cycle of stationary Markov strategies will be played. Recall that it consists of the repetition of a stationary Markov strategies g_1^i for γ_1^i periods, g_2^i for γ_2^i periods... and g_P^i for γ_P^i periods ($\sum_{p=1}^P \gamma_p^i = \gamma^i$). In the rest of the paper this strategy will be denoted by g^{i,γ_i} and called time averaged cycle punishment. To deter deviation aimed at improving the probability distribution over states time averaged cycle

punishments will be made independent of the initial states of the cycle. When a public randomization is played before starting to play a time strategy g^{i,γ_i}, the payoff is given by $\sum_k \beta^k U_i^{\gamma^i}(s, g^{i,k}) = z_i^{i,\gamma^i}(s)$. Let $\pi^1, ..., \pi^n$ be the n publicly randomized punishment strategies. Each cycle is repeated r_2 times. The overall length of the game is $R_2 = r_2 \sum_k \gamma_k^i = r_2 \gamma^i$. Let the $z_i^{i,R_2}(s)$ be the i's average payoffs received during the rewarding phase.

Lemma 8 *Assume that A1, A2 and A4 hold. There is an $\bar{\epsilon} > 0$ so that if $\epsilon \leq \bar{\epsilon}$ then there is a $\gamma^i(\epsilon)$ so that if $\gamma^i \geq \gamma^i(\epsilon)$ then for all $s', s'' \in S$ and $i \neq j$ we have:*

(i) $z_i^{i,\gamma^i}(s) - z_i^{j,\gamma^j}(s') \geq \kappa^i(\epsilon) > 0$

(ii) $z_i^{i,\gamma^i}(s) = z_i^{i,\gamma^i}(s')$

(iii) $m_i(s') < z_i^{i,\gamma^i}(s) < U_i^{\gamma^i}(\tilde{\sigma}, s'')$

(iv) $z_i^{i,\gamma^i}(.)$ *is strictly rational.*

(iv)No deviation inside the support of mixed strategy during the minmax phase are profitable.

PROOF: This proof is a variation on the proof the theorem 9 in Dutta (1995).

STEP 1.

Take $w_i^i = \min\{v_i : (v_i, v_j) \in F\}$ and $l_i^i = \max\{v_i : (v_i, v_j) \in F\}$. By assumption 4 there exist $z^i, z^j \in F^*$,where z^i, z^j are asymmetric payoffs. Take $v = U_i(\tilde{\sigma}, s)$. Let weights $\mu_1, \mu_2, \mu_3 > 0$ be such that $\sum \mu_k = 1$. Let W^i be the following convexification of l^i, z^i,v $W^i = \mu_1 w^i + \mu_2 z^i + \mu_3 v$.

Lemma 9 *(Dutta) There exists a W^i that verifies for $i = (1, ..., n)$:*

(i) strictly individual rationality $W_j^i > m_j$ for all $j = (1, ..., n)$

(ii) asymmetry $W_i^i < W_i^j$ for $i \neq j$

(iii) target domination $W_i^i < v_i$

PROOF(Dutta) : From lemma 3, it follows that l^i, z^i and v are convex combinations of payoffs generated by pure stationary Markov strategies. By lemma 5, W^i can be approximated by a cycle of pure stationary Markov strategies, denoted $g^{i,\gamma}$, consists in playing g_1^i for γ_1^iperiods, then g_2^i for $\gamma_2$periods and ending by playing g_K^i for γ_K^i periods. Let $\gamma^i =$

$\sum_{k=1}^{K} \gamma_k^i$.Choose γ_k^i such that γ_k^i / γ^i can be made arbitrarily close to the convexification weights.

□

Choose μ_1, μ_2, μ_3 such that the lemma above holds.

STEP 2.

Take ϵ' a positive real number so that we have:

$$\min\{W_i^j, v_i\} - \epsilon' > W_i^i + \epsilon' > W_i^i - \epsilon' > m_i + \epsilon'$$

. Let $\gamma^i(\epsilon')$ be an integer so that lemma 5 holds.So $\max_s W^{i,\gamma^i}(s) < \min_s b_i^{i,\gamma^i}(s)$ for all s where $b_i^{i,\gamma^i}(s)$ is the best γ^i-average payoffs when the initial states is s. Thus we can find weights $\mu^i(s), \mu^i(s^{'})$ such that for all $s, s' \in S$:

$$\mu^i(s) W_i^{i,\gamma^i}(s) + (1 - \mu^i(s)) b_i^{i,\gamma^i}(s) = \mu^i(s') W_i^{i,\gamma^i}(s') + (1 - \mu^i(s')) b_i^{i,\gamma^i}(s').$$

Let $z_i^{j,\gamma^i}(s') = \mu^i(s) W_i^{i,\gamma^i}(s) + (1-\mu^i(s)) b_i^{\gamma^i}(s)$, where $\mu^i(s)$ represents the weight involved by the public randomization.

For all s as $\epsilon' \to 0$,we have $\mu^i(s) \to 1$.Then if ϵ' is sufficiently small we have $z_i^{j,\gamma^i}(s') > z_i^{i,\gamma^i}(s)$. Let $\bar{\epsilon} \leq \epsilon'$ be the greater integer so that the previous inequality holds. For a $\epsilon \leq \bar{\epsilon}$ take $\gamma^i(\epsilon)$ be an integer so that lemma 5 and lemma 4 hold. It is clear that (i) and (ii).are satisfied. Also, by lemma 4 $v_i - \epsilon > m_i + \epsilon \geq m_i^{\gamma^i}(s)$. This fact proves (iii). Furthermore $z_i^{i,\gamma^i}(s) > W_i^i - \epsilon > m_i + \epsilon \geq m_i^{\gamma^i}(s')$. This proves (iv). The equilibrium strategies are pure stationary Markov strategies, except those played during the minmaxed phase. This imply that during the minmaxed phase deviations inside the support of mixed strategies are not directly perceptible. Nevertheless, Dutta shows that by modifying the public device during the reward phase, it is possible to deter deviations inside the support of mixed strategies during the minmax phase. This proves (v).

■

6.2 Transition phase and Perfect Threat

Late deviation may occur during the first punishment phases because the remaing time is not sufficient for conditions in lemma 8 to be satisfied. Nevertheless these deviations

can be deterred if the last periods of the game are dedicated to the play of perfect threat stategies. First we give a definition of a perfect threat for stochastic games is given. Then we show that under assumption 3 a stochastic game with sufficiently large horizon, there exists a perfect threat.

Definition 1 *The SPE strategies* $\hat{\alpha}, \check{\alpha}^1, ..., \check{\alpha}^n$ *in stochastic game* Γ^L *are called perfect threat strategies yielding* $M \geq 0$ *if for all* $s, s' \in S$ *and all* i*:* $U_i^L(\hat{\alpha}_L, s) - U_i^L(\check{\alpha}_L^i, s') \geq M$.

The perfect threat strategies consist of a reward strategy $\hat{\alpha}$ and a punishment strategy, $\check{\alpha}^i$, one for each player i. Suppose that the value of M is the expected gain from deviating during the first punishment phases. Because the cost of being punished by perfect threat strategies is greater than M, deviations during the first phases are deterred.

Lemma 10 *Assume that A3 holds. For all* $s \in S$ *and for all* $M > 0$ *there is a* $L(M)$ *such that if* $L \geq L(M)$ *then there are perfect threat strategies yielding* M*.*

PROOF :

Let L_1 and L_2 such denoted respectively the length of the transition phase and the last phase. Let $L = L_1 + L_2$. Let $\hat{\sigma}$ and $\check{\sigma}^i$ the almost stationary Markov strategies defined in A3. Let $X = (x_{\hat{\sigma}_p}, x_{\check{\sigma}_p^1}, ..., x_{\check{\sigma}_p^n})$

Notice that if players play an almost stationary Markov strategy then the transition function is a stationary Markov chain if no deviation occurs. By a standard probability theory result and by A3(i) , for any $x \in X$, the set $\mathcal{R}(x)$, is reached in finite time if players don't deviate from x. Then there exists a pure Markov stationary strategy $\forall x \in X$ so that the play reaches a state in $\mathcal{R}(x)$ in finite time. Denote those strategies $\hat{\sigma}(p)$ and $\check{\sigma}(p)$, and define $X(p) = (x_{\hat{\sigma}_p}, x_{\check{\sigma}_p^1}, ..., \check{\sigma}_p^n)$. Fix L_1 as follows:

$$L_1 = \max_{s \in S, x \in X(p)} \min\{t : \Pr_{x,s}(s^t \in \mathcal{R}(x)) = 1\}$$

Now we focus on the length of the last phase, L_2. Recall that $w_i^i = \min_{s,a} u_i(a, s)$ and $b_i^i = \max_{s,a} u_i(a, s)$. Fix $M' = \max_i R(b_i^i - w_i^i)$ and any integer M. Let $N = M + M'$.

We shall show that for all $s \in \mathcal{R}(\hat{\sigma})$ and $s' \in \mathcal{R}(\check{\sigma}^i)$ we can choose Q' sufficiently so that: $U_i^{\tilde{T}Q'}(\hat{\sigma}, s) - U_i^{\tilde{T}Q'}(\check{\sigma}, s') > N$. Recall that by assumption the $U_i^{\tilde{T}}(\hat{\sigma}, s) - U_i^{\tilde{T}}(\check{\sigma}, s') > 0$ for all $s \in \mathcal{R}(\hat{\sigma})$ and $s' \in \mathcal{R}(\check{\sigma}^i)$. For all $i = (1, ..., n)$, there exist $\bar{s} \in \mathcal{R}(\hat{\sigma})$ and $\tilde{s} \in \mathcal{R}(\check{\sigma}^i)$ such that $\bar{s} = \arg\min_{s \in \mathcal{R}}(\hat{\sigma}) U_i^{\tilde{T}}(\hat{\sigma}, s)$ and $\tilde{s} = \arg\max_{s \in \mathcal{R}(\check{\sigma}^i)} U_i^{\tilde{T}}(\check{\sigma}^i, s)$. Without loss of generality, suppose that i is the player for which $U_i^{\tilde{T}}(\hat{\sigma}, \bar{s}) - U_i^{\tilde{T}}(\check{\sigma}^i, \tilde{s})$ is the smallest. The lower bound of a sequence of $Q'\tilde{T}-$average reward strategy is $U_i^{\tilde{T}}(\hat{\sigma}, \bar{s})$ and the bound of a sequence of $Q'\tilde{T}-$average punishment strategy is $U_i^{\tilde{T}}(\check{\sigma}^i, \tilde{s})$. Then there is a value $Q'(N)$ such that if $Q' \geq Q'(N)$ then we have $Q'(U_i^{\tilde{T}}(\hat{\sigma}, \bar{s}) - U_i^{\tilde{T}}(\check{\sigma}^i, \tilde{s})) > N$.

Suppose that for some $x \in X$ so that $\mathcal{R}(x) \subset S$. Let Q'' be such that

$$b_i^i + [L_1 + (Q' + Q'')\tilde{T} - 1]\omega_i - L_1 w_i^i - (Q' + Q'')\tilde{T} \min_{s \in \mathcal{R}(\check{\sigma}^i)} U_i^{\tilde{T}}(\check{\sigma}^i, s) < 0$$

Such a choice of Q'' makes deviations during the transition phase unprofitable. If for all $x \in X$ so that $\mathcal{R}(x) = S$ then $Q'' = 0$. Finally, $(Q'(N) + Q'')\tilde{T} = Q(N)\tilde{T} = L_2$.

Consider the following strategies:

- No deviation during the reward phase : play $\hat{\sigma}(p)$ for L_1 periods. If i deviates during the transition phase then play the worst SPE until the end. Otherwise play $\hat{\sigma}$ for L_2 periods. Denote this strategy by $\hat{\alpha}$.

- Deviation during the reward phase : play $\check{\sigma}^i(p)$ for L_1 periods if no deviation. If i deviates during the transition phase then play his worst SPE until the end. Otherwise play $\check{\sigma}^i$ for L_2 periods. Denote this strategy by $\check{\alpha}^i$.

Finally we shall verify that the $\hat{\alpha}$ and $\check{\alpha}^i$ are SPE strategy in $\Gamma^L(s)$ for all $s \in S$. First, our choice of Q'' ensures that no deviation occurs during the transition phase. By lemma7 $\hat{\sigma}$ and $\check{\sigma}^i$ are SPE strategies during the last phase whatever Q. Then the strategies $\hat{\alpha}$ and $\check{\alpha}^i$ are SPE in the stochastic game repeated L times, whatever the initial state. Furthermore, it is easy to check that $U_i^L(\hat{\alpha}_L, s) - U_i^L(\check{\alpha}_L^i, s') \geq M$.

■

In the rest of the paper, the payoffs $\min_s U_i^{L_1}(\hat{\alpha}, s) + QU_i^{\tilde{T}}(\hat{\alpha}, \bar{s})$ and $\max_s U_i^{L_1}(\check{\alpha}^i, s) + QU_i^{\tilde{T}}(\check{\alpha}^i, \tilde{s})$ are denoted by $\hat{y}_i^L$ and by $\check{y}_j^{j,L}$. The proof of the previous lemma show that for a given value of M we have $\hat{y}_i^L - \check{y}_j^{j,L} > M$.

6.3 Proof of proposition 1.

The following algorithm defines the strategy that will be used to prove proposition 1.

- STEP 1 : Normal path $\tilde{\sigma}$.

  Play $\tilde{\sigma}$ until T' if no deviation. Then go to step 4.

  If i deviates from $\tilde{\sigma}$ at $t_0 < T - (R_1 + R_2)$ go to step 2. If i deviates from $\tilde{\sigma}$ at $t \geq T - (R_1 + R_2)$ go to step 5.

- STEP 2 : Minmax phase.

  Minmax i for R_1 periods. Then go to 3.

  If $j \neq i$ deviates at $t < T - (R_1 + R_2)$ then go to step 2.

  If $j \neq i$ deviates at $t \geq T - (R_1 + R_2)$ then go to step 5.

  Otherwitre $j \neq i$ deviates restart step 2 where j is minmaxed.

- STEP 3 : Reward phase.

  If i is the last player who deviates play π^i for γ periods and repeat it r_2 times. Let $\gamma \times r_2 = R_2$. Then go to step 4.

  If i deviates at $t < T - (R_1 + R_2)$ then go to step 2.

  If i deviates at $t \geq T - (R_1 + R_2)$ then go to step 5.

- STEP 4 : Transition phase followed by Good SPE phase.

  Play $\hat{\alpha}$ for L periods.

- STEP 5 : Transition phase followed by Bad SPE phase.

  If i is the last player who deviates play $\check{\alpha}^i$ for L periods.

The following lines give a proof of Proposition 1. Fix a $s \in S$. Take V^i the reward phase payoffs constructed in the first step in lemma 7. Choose an ϵ sufficiently small such that $\min\{W_i^j, v_i\} - \epsilon > V_i^i + \epsilon > V_i^i - \epsilon > m_i + \epsilon$. Recall that $w_i^i = \min_{s,a} u_i(a, s)$ and $b_i^i = \max_{s,a} u_i(a, s)$. Without loss of generality, the player who deviates from the normal

play is denoted by i and the date of deviation t. To simplify the notations the index of γ^j is omitted. In the rest of the proof we consider only $R_1 \geq R_1(\epsilon)$, where is the threeshold find in lemma 4. Take the smaller γ such that lemma 8 holds. Fix any integer any L such that $\check{y}_j^{j,L} < \hat{y}_j^L$ for all player j.(it is made possible by lemma 10).

To ensure that no deviation occurs, it suffices to shows that the inequalities 6.1 and 6.2 hold simultaneously. Assume that a deviation occurs at $t < T' - R$. Take $t = T' - (R_1 + r\gamma + \gamma)$. We first fix R_1. If the following inequality hold then no deviation ex post (once the state is realized) occurs during the reward phase: The left hand side is an upper bound on the deviation payoff and the right hand side is a lower bound on the no deviation payoffs.

$$b_i^i + R_1(m_i + \epsilon) + \gamma r \max_s z_i^{i,\gamma}(s) + (\gamma - 1)b_i^i + L\hat{y}_i^L < \gamma w_i^i + (R_1 + r\gamma) \min_s z_i^{i,\gamma}(s) + L\hat{y}_i^L \quad (6.1)$$

By our choice of ϵ and by lemma 8 we have: $\max_s z_i^{i,\gamma}(s) = \min_s z_i^{i,\gamma}(s) = z_i^{i,\gamma}$ and $m_i + \epsilon - z_i^{i,\gamma} < 0$. Then the deviation gain can not be greater than:

$$\Delta_1 = \gamma(b_1 - w_1) + R_1(m_i + \epsilon - z_i^{i,\gamma})$$

Fix R_1 such that Δ_1 is negative. In the following inequality, the left hand side is an upper bound of the payoff of j when he deviates from a punishment directed to i. The right hand side is a lower bound on the no deviation payoff.

$$b_j^j + R_1(m_j + \epsilon) + \gamma r \max_s z_j^{j,\gamma}(s) + (\gamma - 1)b_j^j + L\hat{y}_j^L < R_1 w_j^j + \gamma(r+1) \min_s z_j^{j,\gamma}(s) + L\hat{y}_j^L \quad (6.2)$$

We will show that for an appropriate value of R_2 this inequality is verified. The deviation gain can not be greater than:

$$\Delta_2 = R_1(m_j + \epsilon - w_j^j) + \gamma(b_j^j - \min_s z_i^{i,\gamma}(s)) - \gamma r(\kappa(\epsilon))$$

Where $\kappa(\epsilon) > 0$ by lemma 8. By r sufficiently large, Δ_2 is negative. Take such a r and deduce $R_2 = r \times \gamma$, where γ was fixed in the beginning of the proof, so that 6.2 holds.

Now consider any $T' > R_1 + R_2$.

We shall determine L so that $\check{y}^{1,L}$,...,$\check{y}^{n,L}, \hat{y}^L$ provide a perfect threat. Take R_1 and R_2 so that the inequalities 6.1 and 6.2 hold simultaneously. Let $M = (R_1 + R_2) \max_i (b_i^i - w_i^i)$.

Then apply lemma 10 for this value of M and deduce L. Suppose that i deviates late, i.e. at $t \geq T' - R$. Then $\check{\alpha}^i$ is played. The maximal gain from deviation is $\check{y}_i^{i,L}$. If he doesn't deviate he can get at least $\hat{y}_i^L$. Recall that by lemma 10 we have $L\hat{y}_i^L - L\check{y}_i^{i,L} > M$. Thus player i has no incentive to deviate late.

■

7 Limiting equilibrium payoffs.

These lines give the proof of the theorem 1. Recall that in the proof of the proposition 1 we have fixed L independently of T'. We give a characterization of the limit equilibrium payoffs.

Fix $v \in F^*$, $s \in S$ and $\epsilon > 0$. By lemma 5 there is $T_1'(\epsilon/2)$ such that if $T_1' \geq T_1'(\epsilon/2)$ then there is a $\sigma_{T'}$ such that $||U^{T'}(\sigma, s) - v|| \leq \epsilon/2$ so that $U^{T'}(\sigma, s) \in co\phi^*(s)$. Also, there exists a $\bar{\epsilon}'$ such that lemma 8 holds. Take $\bar{\epsilon} = \max\{\epsilon/2, \bar{\epsilon}\}$. Then by proposition 1 there is $T_2'(\bar{\epsilon})$ and L so that for all $T' \geq T_2'(\bar{\epsilon})$, $\sigma_{T'}\alpha_L \in P^{T'+L}(s)$. Because L is independent of T' then there is a $T_3'(\epsilon/2)$ such that if $T' \geq T_3'(\epsilon/2)$ with $||U^{T'+L}(\sigma_{T'}\alpha_L, s) - U^{T'}(\sigma, s)|| \leq \epsilon/2$.

For all $T^* \geq \max\{T_1'(\epsilon/2), T_2'(\bar{\epsilon}), T_3'(\epsilon/2)\} = T_1'(\epsilon)$ we have $||U^{T'}(\sigma, s) - v|| + ||U^{T'+L}(\sigma_{T'}\alpha_L, s) - U^{T'}(\sigma_{T'}, s)|| \leq \epsilon/2 + \epsilon/2$ and then $||v - U^T(\sigma_{T'}\alpha_L, s)|| \leq \epsilon$.

Now we prove the second point. Fix $s \in S$ and $\epsilon > 0$. Take any $u \in \liminf_{T\to\infty} P(s)$. There exist $T'(\epsilon/2)$ and L so that there are σ, a cycle of pure Markov strategy, and α that satisfy $||U^{T'+L}(\sigma_{T'}\alpha_L, s) - u|| \leq \epsilon/2$ and $||U^{T'+L}(\sigma_{T'}\alpha_L, s) - v|| \leq \epsilon/2$. Then $||v - u|| \leq \epsilon$.

■

8 References

References

[1] Abreu, D.(1983) : "Repeated Games with Discounting," Ph.D. Dissertation, Department of Economics, Princeton University.

[2] Abreu, D., Dutta, P.K., Smith, L.,(1994). "The folk theorem for repeated games: A NEU condition", Econometrica, 62, 939–948.

[3] Blackwell, D., "Discounted dynamic programming", Ann. Math. Statist. 36 (1965) 226-235.

[4] Benoît, J. P, Krishna, V, (1985), "Finitely repeated games", Econometrica,53, 905-922.

[5] Benoît, J.-P., Krishna, V, (1996), "The folk theorems for repeated games: A synthesis", Technical report, EconWPA.

[6] Bewley, T, Kohlberg, E, (1976), "The asymptotic theory of stochastic game", Mth. Operation Res 1 , 197-208.

[7] Dutta, P.K,. (1995), "A folk Theorem for Stochastic games", Journal of Economic Theory, 66, 1-32.

[8] Foata, D.and Fuchs, A, (2002), "Processus stochastiques", Dunod, Paris.

[9] Fudenberg, D, Maskin, E, (1986), "The Folk Theorem in repeated games with discounting and with incomplete information", Econometrica, 54, 533-554.

[10] Neyman, A, Sorin, S. (Eds), (2003), "Stochastic games and Applications", Proceedings of the NATO ASI on Stochastic Games, Kluwer.

[11] Segal, U., Sobel, J.,(2007), "Tit for tat: Foundations of preferences for reciprocity in strategic settings", Journal of Economic Theory Volume 136, Issue 1