12-1-2009

Recursive Methods in Discounted Stochastic Games: An Algorithm for Delta Approaching 1 and a Folk Theorem

Johannes Hörner

Takuo Sugaya

Satoru Takahashi

Nicolas Vieille

RECURSIVE METHODS IN DISCOUNTED STOCHASTIC GAMES: AN ALGORITHM FOR $\delta \to 1$ AND A FOLK THEOREM

By

Johannes Hörner, Takuo Sugaya, Satoru Takahashi and Nicolas Vieille

December 2009
Revised August 2010

COWLES FOUNDATION DISCUSSION PAPER NO. 1742

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Recursive Methods in Discounted Stochastic Games: An Algorithm for $\delta \to 1$ and a Folk Theorem*

Johannes Hörner†, Takuo Sugaya‡, Satoru Takahashi§ and Nicolas Vieille¶

August 11, 2010

Abstract

We present an algorithm to compute the set of perfect public equilibrium payoffs as the discount factor tends to one for stochastic games with observable states and public (but not necessarily perfect) monitoring when the limiting set of (long-run players') equilibrium payoffs is independent of the state. This is the case, for instance, if the Markov chain induced by any Markov strategy profile is irreducible. We then provide conditions under which a folk theorem obtains: if in each state the joint distribution over the public signal and next period's state satisfies some rank condition, every feasible payoff vector above the minmax payoff is sustained by a perfect public equilibrium with low discounting.

Keywords: stochastic games.

JEL codes: C72, C73

1 Introduction

Dynamic games are difficult to solve. In repeated games, finding some equilibrium is easy, as any repetition of a stage-game Nash equilibrium will do. This is not the case in stochastic

*This paper is the result of a merger between two independent projects: Hörner and Vieille developed the algorithm, and Sugaya and Takahashi established the folk theorem by a direct argument. We thank Katsuhiko Aiba, Eduardo Faingold, Drew Fudenberg, Yingni Guo, Tristan Tomala and Yuichi Yamamoto for helpful discussions.

†Yale University, 30 Hillhouse Ave., New Haven, CT 06520, USA, johannes.horner@yale.edu.

‡Princeton University, Fisher Hall, Princeton, NJ 08540, USA, tsugaya@princeton.edu.

§Princeton University, Fisher Hall, Princeton, NJ 08540, USA, satorut@princeton.edu.

¶HEC Paris, 78351 Jouy-en-Josas, France, vieille@hec.fr.

games. Even the most basic equilibria for such games, namely (stationary) Markov equilibria, in which continuation strategies depend on the current state only, often turn out to be challenging. Further complications arise once the assumption of perfect monitoring is abandoned. In that case, our understanding of equilibria in repeated games owes an invaluable debt to Abreu, Pearce and Stacchetti (1990), whose recursive techniques have found numerous applications. Parallel ideas have been developed in stochastic games by Mertens and Parthasarathy (1987, 1991) to establish the most general equilibrium existence results to date (see also Solan, 1998). Those ideas have triggered a development of numerical methods (Judd, Yeltekin and Conklin, 2003), whose use is critical for applications in which postulating a high discount factor appears too restrictive.

In repeated games, more tractable characterizations have been achieved in the case of low discounting. Fudenberg and Levine (1994, hereafter FL) and Fudenberg, Levine and Maskin (1994, hereafter FLM) show that the set of (perfect public) equilibrium payoffs can be characterized by a family of static constrained optimization programs. Based on this and other insights, they derive sufficient conditions for a folk theorem for repeated games with imperfect public monitoring. The algorithm developed by FL has proved to be useful, as it allows to identify the equilibrium payoff set in interesting cases in which the sufficient conditions for the folk theorem fail. This is the case, for instance, in the partnership game of Radner, Myerson and Maskin (1986). Most importantly, FL's algorithm can accommodate both long-run and short-run players, which is essential for many applications, especially in macroeconomics and political economy, in which consumers or voters are often modeled as non-strategic (see Mailath and Samuelson, 2006, Chapters 5 and 6).

This paper extends these results to stochastic games. More precisely, it provides an algorithm that characterizes the set of perfect public equilibria in stochastic games with finite states, signals and actions, in which states are observed, monitoring is imperfect but public, under the assumption that the limiting set of equilibrium payoffs (of the long-run players) is independent of the initial state (this is the case, in particular, if the Markov chain over states defined by any Markov strategy profile is irreducible). This algorithm is a natural extension of FL, and indeed, reduces to it if there is only one state. The key to this characterization lies in the linear constraints to impose on continuation payoffs. In FL, each optimization program is indexed by a direction $\lambda \in \mathbf{R}^I$ that specifies the weights on the payoffs of the I players. Trivially, the boundary point v of the equilibrium payoff set that maximizes this weighted average in the direction λ is such that, for any realized signal y, the continuation payoff $w(y)$ must satisfy the property that

$\lambda \cdot (w(y) - v) \leq 0$. Indeed, one of the main insights of FL is that, once discounting is sufficiently low, attention can be restricted to this linear constraint, for each λ, so that the program itself becomes linear in w, and hence considerably more tractable. In a stochastic game, it is clear that the continuation payoff must be indexed not only by the public signal, but also by the realized state, and since the stage games that correspond to the current and to the next state need not be the same, there is little reason to expect this property to be preserved for each pair of states. Indeed, it is not. One might then wonder whether the resulting problem admits a linear characterization at all. The answer is affirmative.

When all players are long-run, this characterization can be used to establish a folk theorem under assumptions that parallel those invoked by FLM. In stochastic games, note that state transitions might be affected by actions taken, so that, because states are observed, they already provide information about players' past actions. Therefore, it is natural to impose rank conditions at each state on how actions affect the joint distribution over the future state and signal. This is weaker than requiring such conditions on signals alone, and indeed, it is easy to construct examples where the folk theorem holds without any public signals (see Sections 3.2 and 5).

This folk theorem also generalizes Dutta's (1995) folk theorem for stochastic games with perfect monitoring. Unlike ours, Dutta's ingenious proof is constructive, extending ideas developed by Fudenberg and Maskin (1986) for the case of repeated games with perfect monitoring. However, his assumptions (except for the monitoring structure) are the same as ours. In independent and simultaneous work, Fudenberg and Yamamoto (2010) provide a different, direct proof of the folk theorem for stochastic games with imperfect public monitoring under irreducibility, without a general characterization of the equilibrium payoff set. Their rank assumptions are stronger, as discussed in Section 5.

Finally, our results also imply the average cost optimality equation from dynamic programming, which obtains here as a special case where there is a single player. The average cost optimality equation is widely used in operations research, for instance in routing, inventory, scheduling and queuing problems, and our results might thus prove useful for game-theoretic extensions of such problems, as in inventory or queuing games.

Of course, stochastic games are also widely used in economics. They play an important role in industrial organization (among many others, see Ericson and Pakes, 1995). It is hoped that methods such as ours might help provide integrated analyses of questions whose treatment had to be confined to simple environments so far, such as the role of imperfect monitoring

(Green and Porter, 1984) and of business cycles (Rotemberg and Saloner, 1986) in collusion, for instance. Rigidities and persistence play an important role in macroeconomics as well, giving rise to stochastic games. See Phelan and Stacchetti (2001), for example. In Section 6, we shall apply our results to a simple political economy game, and establish a version of the folk theorem when some players are short-run.

2 Notation and Assumptions

We introduce stochastic games with public signals. At each stage, the game is in one state, and players simultaneously choose actions. Nature then stochastically determines the current reward (or flow payoff) profile, the next state and a public signal, as a function of the current state and the action profile. The sets S of possible states, I of players, A^i of actions available to player i, and Y of public signals are assumed finite.[1]

Given an action profile $a \in A := \times_i A^i$ and a state $s \in S$, we denote by $r(s,a) \in \mathbf{R}^I$ the reward profile in state s given a, and by $p(t,y|s,a)$ the joint probability of moving to state $t \in S$ and of getting the public signal $y \in Y$. (As usual, we can think of $r^i(s,a)$ as the expectation given a of some realized reward that is a function of a private outcome of player i and the public signal only).

We assume that at the end of each period, the only information publicly available to all players consists of nature's choices: the next state together with the public signal. When properly interpreting Y, this includes the case of perfect monitoring and the case of publicly observed rewards. Note however that this fails to include the case where actions are perfectly monitored, yet states are not disclosed. In such a case, the natural "state" variable is the (common) posterior belief of the players on the underlying state.

Thus, in the stochastic game, in each period $n = 1, 2, \ldots$, the state is observed, the stage game is played, and the corresponding public signal is then revealed. The stochastic game is parameterized by the initial state s_1, and it will be useful to consider all potential initial states simultaneously. The public history at the beginning of period n is then $h_n = (s_1, y_1, \ldots, s_{n-1}, y_{n-1}, s_n)$. We set $H_1 := S$, the set of initial states. The set of public histories at the beginning of period n is therefore $H_n := (S \times Y)^{n-1} \times S$, and we let $H := \bigcup_{n \geq 1} H_n$ denote the set of all public histories. The private history for player i at the beginning of period n is a sequence $h_n^i =$

[1]For notational convenience, the set of available actions is independent of the state. See, however, footnote 11. All results extend beyond that case.

$(s_1, a_1^i, y_1, \ldots, s_{n-1}, a_{n-1}^i, y_{n-1}, s_n)$, and we similarly define $H_1^i := S$, $H_n^i := (S \times A^i \times Y)^{n-1} \times S$ and $H^i := \bigcup_{n \geq 1} H_n^i$. Given a stage $n \geq 1$, we denote by s_n the state, a_n the realized action profile, and y_n the public signal in period n. We will often use the same notation to denote both these realizations and the corresponding random variables.

A (behavior) strategy for player $i \in I$ is a map $\sigma^i : H^i \to \Delta(A^i)$. Every pair of initial state s_1 and strategy profile σ generates a probability distribution over histories in the obvious way and thus also generates a distribution over sequences of the players' rewards. Players seek to maximize their payoff, that is, the average discounted sum of their rewards, using a common discount factor $\delta < 1$. Thus, the payoff of player $i \in I$ if the initial state is s_1 and the players follow the strategy profile σ is defined as

$$\sum_{n=1}^{\infty} (1-\delta)\delta^{n-1} \mathbf{E}_{s_1,\sigma}[r^i(s_n, a_n)].$$

We shall consider a special class of equilibria. A strategy σ^i is public if it depends on the public history only, and not on the private information. That is, a public strategy is a mapping $\sigma^i : H \to \Delta(A^i)$. A *perfect public* equilibrium (hereafter, PPE) is a profile of public strategies such that, given any period n and public history h_n, the strategy profile is a Nash equilibrium from that period on. Note that this class of equilibria includes Markov equilibria, in which strategies only depend on the current state and period. In what follows though, a *Markov* strategy for player i will be a public strategy that is a function of states only, i.e. a function $S \to \Delta(A^i)$.[2]

Note also that the set of PPE payoffs is a function of the current state only, and does not otherwise depend on the public history, nor on the period. Perfect public equilibria are sequential equilibria, but it is easy to construct examples showing that the converse is not generally true. What we characterize, therefore, is a subset of the sequential equilibrium payoffs.

We denote by $E_\delta(s) \subset \mathbf{R}^I$ the (compact) set of PPE payoffs of the game with initial state $s \in S$ and discount factor $\delta < 1$. All statements about convergence of, or equality between sets are understood in the sense of the Hausdorff distance $d(A, B)$ between sets A, B.

Because both state and action sets are finite, it follows from Fink (1964) and Takahashi (1964) that a (perfect public) equilibrium always exists in this set-up.

Our main result does not apply to all finite stochastic games. Some examples of stochastic games, in particular those involving absorbing states, exhibit remarkably complex asymptotic

[2]In the literature on stochastic games, such strategies are often referred to as stationary strategies.

properties. See, for instance, Bewley and Kohlberg (1976) or Sorin (1986). We shall come back to the implications of our results for such games. Our main theorem makes use of the following assumption.

Assumption A: The limit set of PPE payoffs is independent of the initial state: for all $s, t \in S$,

$$\lim_{\delta \to 1} d(E_\delta(s), E_\delta(t)) = 0.$$

This is an assumption on endogenous variables. A stronger assumption on exogenous variables that implies Assumption **A** is *irreducibility*: For any pure Markov strategy profile $(a_s)_{s \in S} \in A^S$, the Markov chain over S with transition function

$$q(t|s) := p(\{t\} \times Y | s, a_s)$$

is irreducible. Actually, it is not necessary that every Markov strategy gives rise to an irreducible Markov chain. It is clearly sufficient if there is some state that is accessible from every other state regardless of the Markov strategy.

Another class of games that satisfy Assumption **A**, although they do not satisfy irreducibility, is the class of alternating-move games (see Lagunoff and Matsui, 1997 and Yoon, 2001). With two players, for instance, such a game can be modeled as a stochastic game in which the state space is $A^1 \cup A^2$, where $a^i \in A^i$ is the state that corresponds to the last action played by player i, when it is player $-i$'s turn to move. (Note that this implies perfect monitoring by definition, as states are observed.)

Note also that, by redefining the state space to be $S \times Y$, one may further assume that only states are disclosed. That is, the class of stochastic games with public signals is no more general than the class of stochastic games in which only the current state is publicly observed. However, the Markov chain over $S \times Y$ with transition function $\tilde{q}(t, z|s, y) := p(t, z|s, a_s)$ need not be irreducible even if q is.

3 An Algorithm to Compute Equilibrium Payoffs

3.1 Preliminaries: Repeated Games

As our results generalize the algorithm of FL, it is useful to start with a reminder of their results, and examine, within a specific example, what difficulties a generalization to stochastic games raises.

Recall that the set of (perfect public) equilibrium payoffs must be a fixed point of the Bellman-Shapley operator (see Abreu, Pearce and Stacchetti, 1990). Define the one-shot game $\Gamma_\delta(w)$, where $w : Y \to \mathbf{R}^I$, with action sets A^i and payoff

$$v = (1-\delta)r(a) + \delta \sum_{y \in Y} p(y|a)w(y). \tag{1}$$

Note that if v is an equilibrium payoff vector of the repeated game, associated with action profile α and continuation payoff vectors $w(y)$, as a function of the initial signal, then α must be a Nash equilibrium of $\Gamma_\delta(w)$, with payoff v. Conversely, if we are given a function w such that $w(y) \in E_\delta$ for all y, and a Nash equilibrium α of $\Gamma_\delta(w)$ with payoff v, then we can construct an equilibrium of the repeated game with payoff v in which the action profile α is played in the initial period.

Therefore, the analysis of the repeated game can be reduced to that of the one-shot game. The constraint that the continuation payoff lies in the (unknown) set E_δ complicates the analysis significantly. FL's key observation is that this constraint can be replaced by linear constraints for the sake of asymptotic analysis (as $\delta \to 1$). If $v \in E_\delta$ is an equilibrium payoff of the one-shot game $\Gamma_\delta(w)$, then, subtracting δv on both sides and dividing through by $1-\delta$,

$$v = r(a) + \sum_{y \in Y} p(y|a)x(y), \tag{2}$$

where, for all y,

$$x(y) := \frac{\delta}{1-\delta}(w(y) - v), \text{ or } w(y) = v + \frac{1-\delta}{\delta}x(y).$$

Thus, provided that the equilibrium payoff set is convex, v is also in $E_{\tilde{\delta}}$ for all $\tilde{\delta} > \delta$, because we can use as continuation payoff vectors $\tilde{w}(y)$ the re-scaled vectors $w(y)$ (see Figure 1). Conversely, provided that the normal vector to the boundary of E_δ varies continuously with the boundary point, then any set of payoff vectors $w(y)$ that lie in one of the half-spaces defined by this normal vector (i.e. such that $\lambda \cdot (w(y) - v) \leq 0$, or equivalently, $\lambda \cdot x(y) \leq 0$) must also lie in E_δ

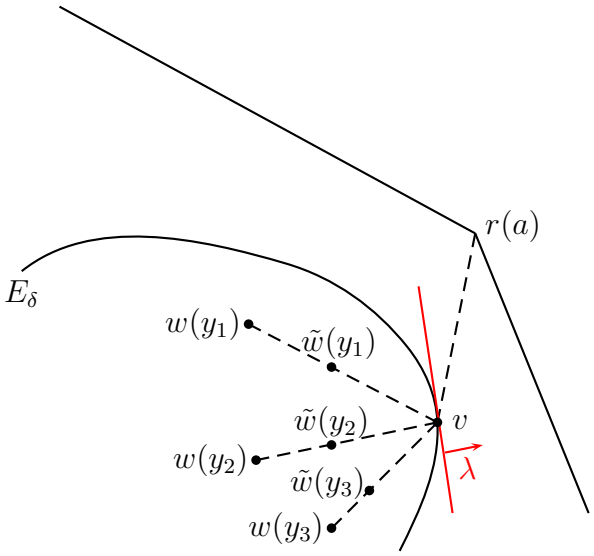

Figure 1: Continuation payoffs as a function of the discount factor

for discount factors close enough to one. In particular, if we seek to identify the payoff v that maximizes $\lambda \cdot v$ on E_δ for δ close enough to 1, given $\lambda \in \mathbf{R}^I$, it suffices to compute the score

$$k(\lambda) := \sup_{x,v} \lambda \cdot v,$$

such that v be a Nash equilibrium payoff of the game $\Gamma(x)$ whose payoff function is given by (2), and subject to the linear constraints $\lambda \cdot x(y) \leq 0$ for all y. Note that the discount factor no longer appears in this program. We thus obtain a half-space $\mathcal{H}(\lambda) := \{v \in \mathbf{R}^I : \lambda \cdot v \leq k(\lambda)\}$ containing $\lim_{\delta \to 1} E_\delta$.

This must be true for all vectors $\lambda \in \mathbf{R}^I$. Let $\mathcal{H} := \bigcap_{\lambda \in \mathbf{R}^I} \mathcal{H}(\lambda)$. It follows, under appropriate dimensionality assumptions, that the equilibrium payoff set can be obtained as the intersection of these half-spaces (see FL, Theorem 3.1):

$$\mathcal{H} = \lim_{\delta \to 1} E_\delta.$$

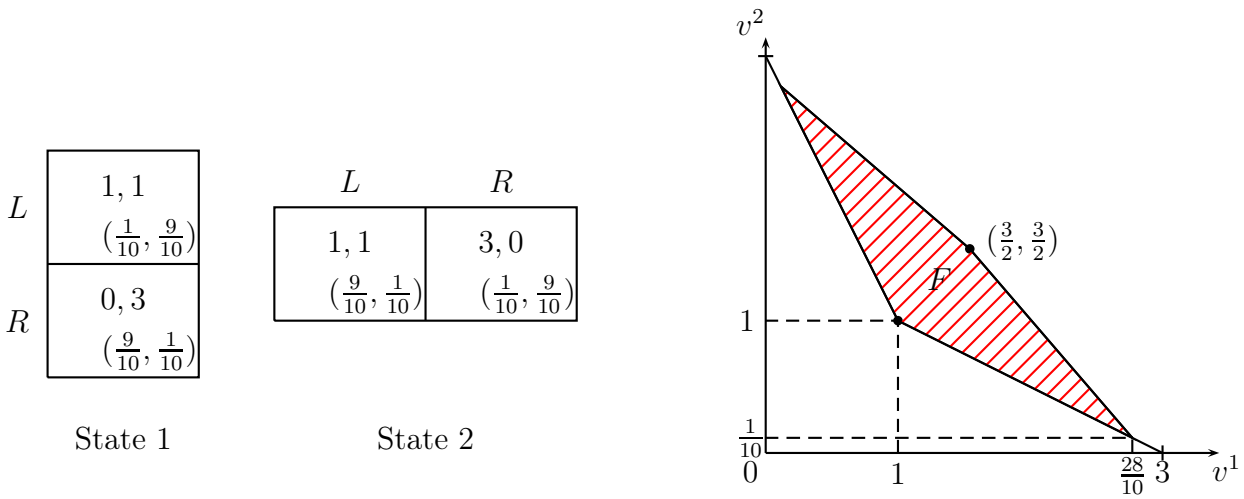

Figure 2: Rewards and Transitions in Example 1

3.2 A Stochastic Game

Our purpose is to come up with an algorithm for stochastic games that would generalize FL's algorithm. To do so, we must determine the appropriate constraints on continuation payoffs. Let us attempt to adapt FL's arguments to a specific example

There are two states, $i = 1, 2$, and two players. Each player only takes an action in his own state: player i chooses L or R in state i. Actions are not observable ($Y = \varnothing$), but affect transitions, so that players learn about their opponents' actions via the evolution of the state. If action L (R) is taken in state i, then the next state is again i with probability p^L (p^R). Let us pick here $p^L = 1 - p^R = 1/10$. Rewards are given in Figure 2 (transition probabilities to states 1 and 2, respectively, are given in parenthesis). Throughout, we refer to this game as Example 1.

Player i has a higher reward in state $j \neq i$, independently of the action. Moreover, by playing L, which yields him the higher reward in his own state, he maximizes the probability to switch states. Thus, playing the efficient action R requires intertemporal incentives, which are hard to provide absent public signals. Constructing an equilibrium in which L is not always played appears challenging, but not impossible: playing R in state i if and only if the state was $j \neq i$ in the previous two periods (or since the beginning of the game if fewer periods have elapsed) is an equilibrium for some high discount factor ($\delta \approx .823$). So there exist equilibrium payoffs above 1.

In analogy with (1), we may now decompose the payoff vector in state $s = 1, 2$ as

$$v_s = (1 - \delta) r(s, a_s) + \delta \sum_t p(t|s, a_s) w_t(s), \tag{3}$$

where t is the next state, $w_t(s)$ is the continuation payoff then, and $p(t|s, a_s)$ is the probability of transiting from s to t given action a_s at state s. Fix $\lambda \in \mathbf{R}^I$. If v_s maximizes the score $\lambda \cdot v_s$ in all states $s = 1, 2$, then the continuation payoff in state t gives a lower score than v_t, independently of the initial state: for all t,

$$\lambda \cdot (w_t(s) - v_t) \leq 0. \tag{4}$$

Our goal is to eliminate the discount factor. Note, however, that if we subtract δv_s on both sides of (3), and divide by $1 - \delta$, we obtain

$$v_s = r(s, a_s) + \sum_t p(t|s, a_s) \frac{\delta}{1-\delta} (w_t(s) - v_s), \tag{5}$$

and there is no reason to expect $\lambda \cdot (w_t(s) - v_s)$ to be negative, unless $s = t$ (compare with (4)). Unlike the set of limiting payoffs as $\delta \to 1$, the set of feasible rewards does depend on the state (in state 1, it is the segment $[(1, 1), (0, 3)]$; in state 2, the segment $[(1, 1), (3, 0)]$; see right panel of Figure 2), and so the score $\lambda \cdot w_t(s)$ in state t might exceed the maximum score achieved by v_s in state s. Thus, defining x by

$$x_t(s) := \frac{\delta}{1-\delta} (w_t(s) - v_s),$$

we know that $\lambda \cdot x_s(s) \leq 0$, for all s, but not the sign of $\lambda \cdot x_t(s)$, $t \neq s$. On the one hand, we cannot restrict it to be negative: if $\lambda \cdot x_2(1) \leq 0$, then, because also $\lambda \cdot x_1(1) \leq 0$, by considering $\lambda = (1, 0)$, player 1's payoff starting from state 1 cannot exceed his highest reward in that state (i.e., 1). Yet we know that some equilibria yield strictly higher payoffs. On the other hand, if we impose no restrictions on $x_t(s)$, $s \neq t$, then we can set v_s as high as we wish in (5) by picking $x_t(s)$ large enough. The value of the program to be defined would be unbounded. What is the missing constraint?

We do know that (4) holds for all pairs (s, t). By adding up these inequalities for $(s, t) = (1, 2)$ and $(2, 1)$, we obtain

$$\lambda \cdot (w_1(2) + w_2(1) - v_1 - v_2) \leq 0, \text{ or, rearranging, } \lambda \cdot (x_1(2) + x_2(1)) \leq 0. \tag{6}$$

Equation (6) has a natural interpretation in terms of *s-blocks*, as defined in the literature on Markov chains (see, for instance, Nummelin, 1984). When the Markov chain (induced by the players' strategies) is communicating, as it is in our example, we might divide the game into the

subpaths of the chain between consecutive visits to a given state s. The score achieved by the continuation payoff once state s is re-visited on the subpath $(s_1, \ldots, s_L)$ (where $s_1 = s_L = s$) cannot exceed the score achieved by v_s, and so the difference in these scores, as measured by the sum $\sum_{l=1}^{L-1} \lambda \cdot x_{s_{l+1}}(s_l)$, must be negative. Note that the irreducibility assumption also guarantees that the limit set of feasible payoffs F (as $\delta \to 1$) is independent of δ, as shown in the right panel of Figure 2.[3]

To conclude, we obtain the program

$$\sup_{\alpha, x, v} \lambda \cdot v,$$

over payoff vectors $v \in \mathbf{R}^2$, $\alpha = (\alpha_s)_{s=1,2}$, and $\{x_s(t) \in \mathbf{R}^2 : s, t = 1, 2\}$ such that, in each state s, α_s is a Nash equilibrium of the game with payoff $r(s, a) + \sum_t p(t|s, a) x_t(s)$, and such that $\lambda \cdot x_1(1) \leq 0, \lambda \cdot x_2(2) \leq 0$, and $\lambda \cdot (x_1(2) + x_2(1)) \leq 0$. Note that this program already factors in our assumption that equilibrium payoffs can be taken to be independent of the state.

It will follow from the main theorem of the next section that this is the right program. Perhaps it is a little surprising that the constraints involve unweighted sums of vectors $x_t(s)$, rather than, say, sums that are weighted by the invariant measure under the equilibrium strategy. Lemma 1 below will provide a link between the constraints and such sums. What should come as no surprise, though, is that generalizing these constraints to more than two states will involve considering all cycles, or permutations, over states (see constraint **(ii)** below).

3.3 The Characterization

Given a state $s \in S$ and a map $x : S \times Y \to \mathbf{R}^{S \times I}$, we denote by $\Gamma(s, x)$ the one-shot game with action sets A^i and payoff function

$$r(s, a_s) + \sum_{t \in S} \sum_{y \in Y} p(t, y|s, a_s) x_t(s, y),$$

where $x_t(s, y) \in \mathbf{R}^I$ is the t-th component of $x(s, y)$.

[3]This set can be computed by considering randomizations over pure Markov strategies, see Dutta (1995), Lemma 1.

Given $\lambda \in \mathbf{R}^I$, we denote by $\mathcal{P}(\lambda)$ the maximization program

$$\sup \lambda \cdot v,$$

where the supremum is taken over all $v \in \mathbf{R}^I$ and $x : S \times Y \to \mathbf{R}^{S \times I}$ such that

(i) For each s, v is a Nash equilibrium payoff of the game $\Gamma(s, x)$;

(ii) For each $T \subseteq S$, for each permutation $\phi : T \to T$ and each map $\psi : T \to Y$, one has $\lambda \cdot \sum_{s \in T} x_{\phi(s)}(s, \psi(s)) \leq 0$.

Denote by $k(\lambda) \in [-\infty, +\infty]$ the value of $\mathcal{P}(\lambda)$. We will prove that the feasible set of $\mathcal{P}(\lambda)$ is non-empty, so that $k(\lambda) > -\infty$ (Proposition 1), and that the value of $\mathcal{P}(\lambda)$ is finite, so that $k(\lambda) < +\infty$ (Section 3.4).

We define $\mathcal{H}(\lambda) := \{v \in \mathbf{R}^I : \lambda \cdot v \leq k(\lambda)\}$, and set $\mathcal{H} := \bigcap_{\lambda \in \mathbf{R}^I} \mathcal{H}(\lambda)$. Note that $\mathcal{H}$ is convex. Let S^1 denote the set of $\lambda \in \mathbf{R}^I$ of norm 1.[4] Observe that $\mathcal{H}(0) = \mathbf{R}^I$, and that $\mathcal{H}(\lambda) = \mathcal{H}(c\lambda)$ for every $\lambda \in \mathbf{R}^I$ and $c > 0$. Hence $\mathcal{H}$ is also equal to $\bigcap_{\lambda \in S^1} \mathcal{H}(\lambda)$.

Our main result is a generalization of FL's algorithm to compute the limit set of payoffs as $\delta \to 1$.

Theorem 1 (Main Theorem) *Assume that $\mathcal{H}$ has non-empty interior. Under Assumption **A**, $E_\delta(s)$ converges to $\mathcal{H}$ as $\delta \to 1$, for any $s \in S$.*

Note that, with one state only, our optimization program reduces to the algorithm of FL. The proof of Theorem 1 is organized in two propositions, stated below and proved in appendix. Note that these propositions do not rely on Assumption **A**.

Proposition 1 *For every $\delta < 1$, we have the following.*

1. *$k(\lambda) \geq \min_{s \in S} \max_{w \in E_\delta(s)} \lambda \cdot w$ for every $\lambda \in S^1$.*
2. *$\mathcal{H} \supseteq \bigcap_{s \in S} E_\delta(s)$.*

We note that it need not be the case that $E_\delta(s) \subseteq \mathcal{H}$ for each $s \in S$.

Proposition 2 *Assume that $\mathcal{H}$ has non-empty interior, and let Z be any compact set contained in the interior of $\mathcal{H}$. Then $Z \subseteq E_\delta(s)$, for every $s \in S$ and δ large enough.*

[4] Throughout, we use the Euclidean norm.

The logic of the proof of Proposition 2 is inspired by FL and FLM, but differs in some important respects. We here give a short and overly simplified account of the proof, that nevertheless contains some basic insights.

Let a payoff vector $v \in Z$, and a direction λ be given. Since v is interior to $\mathcal{H}$, one has $\lambda \cdot v < k(\lambda)$, and there thus exists $x = (x_t(s,y))$ such that v is a Nash equilibrium payoff of the one-shot game $\Gamma(s,x)$, and all inequality constraints on x hold with a strict inequality.

For high δ, we use x to construct equilibrium continuation payoffs w adapted to v in the discounted game, with the interpretation that $x_t(s,y)$ is the normalized (continuation) *payoff increment*, should (t,y) occur. Since we have no control over the sign of $\lambda \cdot x_t(s,y)$, the one-period argument that is familiar from repeated games does not extend to stochastic games. To overcome this issue, we will instead rely on large blocks of stages of fixed size. Over such a block, and thanks to the inequalities **(ii)** satisfied by x, we will prove that the sum of payoff increments is negative. This in turn will ensure that the continuation payoff at the end of the block is below v in the direction λ.

Since $\mathcal{H}$ is convex, it follows from these two propositions that

$$\mathcal{H} = \lim_{\delta \to 1} \bigcap_{s \in S} E_\delta(s).$$

This statement applies to all finite stochastic games with observable states and full-dimensional $\mathcal{H}$, whether they satisfy Assumption **A** or not. Theorem 1 then follows, given Assumption **A**.

3.4 Finiteness of $\mathcal{P}(\lambda)$

It is instructive, and useful for the sequel, to understand why the value of $\mathcal{P}(\lambda)$ is finite, for each possible choice of λ. To see this, we rely on the next lemma, which we also use at other places.

Lemma 1 *Let q be an irreducible transition function over a finite set R, with invariant measure μ. Then, for each $\varnothing \neq T \subseteq R$, and every one-to-one map $\phi : T \to T$, there is $\pi_{T,\phi} \geq 0$, such that the following holds. For every $(x_t(s)) \in \mathbf{R}^{R \times R}$, one has*

$$\sum_{s \in R} \mu(s) \sum_{t \in R} q(t|s) x_t(s) = \sum_{\varnothing \neq T \subseteq R} \sum_{\phi : T \to T} \pi_{T,\phi} \sum_{s \in T} x_{\phi(s)}(s),$$

where the sum ranges over all one-to-one maps $\phi : T \to T$.

In words, for any payoff increments $(x_t(s))$, the expectation of these increments (with respect to the invariant measure) is equal to some fixed conical combination of their sum over cycles. This provides a link between the constraints **(i)** and **(ii)**.

Proof. We exploit an explicit formula for μ, due to Freidlin and Wentzell (1991). Given a state $s \in R$, an s-graph is a directed graph g with vertex set R, and with the following two properties:

- any state $t \neq s$ has outdegree one, while s has outdegree zero;
- the graph g has no cycle.

Equivalently, for any $t \neq s$, there is a unique longest path starting from t, and this path ends in s. We identify such a graph with its set of edges. The weight of any such graph g is defined to be $q(g) := \prod_{(t,u)\in g} q(u|t)$. Let $G(s)$ denote the set of s-graphs. Then

$$\mu(s) = \frac{\sum_{g\in G(s)} q(g)}{\sum_{t\in R}\sum_{g\in G(t)} q(g)}.$$

Thus, one has

$$\sum_{s\in R} \mu(s) \sum_{t\in R} q(t|s)x_t(s) = \frac{\sum_{s\in R}\sum_{g\in G(s)}\sum_{t\in R} q(g)q(t|s)x_t(s)}{\sum_{t\in R}\sum_{g\in G(t)} q(g)}.$$

Let Ω be the set of triples (s, g, t), such that $g \in G(s)$. Define an equivalence relation $\sim$ over Ω by $(s, g, t) \sim (s', g', t')$ if $g \cup \{(s \to t)\} = g' \cup \{(s' \to t')\}$. Observe that for each $(s, g, t) \in \Omega$, the graph $g \cup \{(s \to t)\}$ has exactly one cycle (which contains $s \to t$), and all vertices of S have outdegree 1.

Let $C \subseteq \Omega$ be any equivalence class for $\sim$. Let $s_1 \to s_2 \to \cdots \to s_k \to s_1$ denote the unique, common, cycle of all $(s, g, t) \in C$. Define $T := \{s_1, \ldots, s_k\}$, and denote by $\phi : T \to T$ the map which associates to any state $u \in T$ its successor in the cycle.

Observe that the product $q(g)q(t|s)$ is independent of $(s, g, t) \in C$, and we denote it $\rho^C_{T,\phi}$. It is then readily checked that

$$\sum_{(s,g,t)\in C} q(g)q(t|s)x_t(s) = \rho^C_{T,\phi} \sum_{u\in T} x_{\phi(u)}(u).$$

The result follows by summation over equivalence classes. ■

Fix a direction $\lambda \in \mathbf{R}^I$. To show that the value of $\mathcal{P}$ is finite, take any feasible point (v, x) in $\mathcal{P}(\lambda)$ and a Markov strategy $\alpha = (\alpha_s)$ such that, for each $s \in S$, α_s is a Nash equilibrium of $\Gamma(s, x)$, with payoff v. Let $R \subseteq S$ be an arbitrary recurrent set of the Markov chain induced by α, and denote by $\mu \in \Delta(R)$ the invariant measure over R. For each $s, t \in R$, let $f(s, t)$ be a signal y that maximizes $\lambda \cdot x_t(s, y)$, and denote $q(t|s) = p(\{t\} \times Y|s, \alpha_s)$. Then Lemma 1 implies that

$$\begin{aligned}
\lambda \cdot v &\leq \lambda \cdot \sum_{s \in R} \mu(s) r(s, \alpha_s) + \lambda \cdot \sum_{s \in R} \mu(s) \sum_{t \in R} q(t|s) x_t(s, f(s, t)) \\
&= \lambda \cdot \sum_{s \in R} \mu(s) r(s, \alpha_s) + \sum_{\emptyset \neq T \subseteq R} \sum_{\phi : T \to T} \pi_{T,\phi} \left(\lambda \cdot \sum_{s \in T} x_{\phi(s)}(s, f(s, \phi(s))) \right) \\
&\leq \lambda \cdot \sum_{s \in R} \mu(s) r(s, \alpha_s)
\end{aligned}$$

for some $\pi_{T,\phi} \geq 0$, each $\varnothing \neq T \subseteq R$ and each one-to-one map $\phi : T \to T$.

4 Connection to Known Results

Our characterization includes as special cases the characterization obtained by FL for repeated games, as well as the average cost optimality equation from dynamic programming in the case of one player. In this section, we explain these connections in more detail.

4.1 Another Generalization of FL's Algorithm

The specific form of the optimization program $\mathcal{P}(\lambda)$ is intriguing, and calls for some discussion. We here elaborate upon Section 3.2. For simplicity, we assume full monitoring.

Perhaps a natural generalization of FL's algorithm to a stochastic set-up would have been the following. Given a direction $\lambda \in \mathbf{R}^I$, and for each initial state $s \in S$, consider the highest PPE payoff $v_s \in \mathbf{R}^I$ in the direction λ, when starting from s. Again, for each initial state s, there is a mixed profile α_s, and continuation PPE payoffs $w_t(s, a)$, to be interpreted as the continuation payoff in the event that the next state turns out to be t, and such that:

(a) For each s, α_s is a Nash equilibrium with payoff v_s of the game with payoff function $(1-\delta)r(s,a)+\delta\sum_t p(t|s,a)w_t(s,a)$.

(b) For any two states $s,t\in S$, and any action profile $a\in A$, $\lambda\cdot w_t(s,a)\leq\lambda\cdot v_t$.

Mimicking FL's approach, we introduce the program, denoted $\tilde{\mathcal{P}}(\lambda,\delta)$, $\sup\min_s\lambda\cdot v_s$, where the supremum is taken over all $((v_s),w,\alpha)$ such that **(a)** and **(b)** hold.[5] As we discussed in Section 3.2, and unlike in the repeated game framework, the value of $\tilde{\mathcal{P}}(\lambda,\delta)$ *does* depend on δ. The reason is that, when setting $x_t(s,a)=\frac{\delta}{1-\delta}(w_t(s,a)-v_s)$, the inequality $\lambda\cdot x_t(s,a)\leq 0$ need not hold (note that $x_t(s,a)$ involves $w_t(s,a)$ and v_s, while **(b)** involves $w_t(s,a)$ and v_t).

To obtain a program that does not depend on δ, we relax the program $\tilde{\mathcal{P}}(\lambda,\delta)$ as follows. Observe that, for any sequence $s_1,\ldots,s_k$ of states such that $s_k=s_1$, and for each $a_j\in A$ $(1\leq j<k)$, one has

$$\begin{aligned}\lambda\cdot\sum_{j=1}^{k-1}x_{s_{j+1}}(s_j,a_j)&=\frac{\delta}{1-\delta}\left(\sum_{j=1}^{k-1}\lambda\cdot\left(w_{s_{j+1}}(s_j,a_j)-v_{s_j}\right)\right)\\&=\frac{\delta}{1-\delta}\left(\sum_{j=1}^{k-1}\lambda\cdot\left(w_{s_{j+1}}(s_j,a_j)-v_{s_{j+1}}\right)\right)\leq 0,\end{aligned}$$

where the second equality holds since $s_1=s_k$, and the final inequality holds by **(b)**. These are precisely the averaging constraints which appear in our linear program $\mathcal{P}(\lambda)$.

That is, the program $\mathcal{P}(\lambda)$ appears as a discount-independent relaxation of the program $\tilde{\mathcal{P}}(\lambda,\delta)$.[6] This immediately raises the issue of whether this is a "meaningful" relaxation of the program. We here wish to suggest that this is the case, by arguing somewhat informally that the two programs $\mathcal{P}(\lambda)$ and $\tilde{\mathcal{P}}(\lambda,\delta)$ have asymptotically the same value, as $\delta\to 1$. To show this claim, we start with a feasible point (v,x,α) in $\mathcal{P}(\lambda)$, and we construct a feasible point $((v_s),w,\alpha)$ in $\tilde{\mathcal{P}}(\lambda,\delta)$ such that v_s-v is of the order of $(1-\delta)c$ for each $s\in S$ and some real number c.[7]

[5] Taking the minimum (or maximum) over s in the objective function has no effect asymptotically under Assumption **A**. We choose the minimum for the convenience of the discussion.

[6] Note that $((v_s),x)$ does not define a feasible point in $\mathcal{P}(\lambda)$, since feasibility in $\mathcal{P}(\lambda)$ requires that payoffs from different initial states coincide. This is straightforward to fix. Let $\bar{s}$ be a state that minimizes $\lambda\cdot v_s$. Set $v:=v_{\bar{s}}$, and $\bar{x}_t(s,a):=x_t(s,a)+v_{\bar{s}}-v_s$ for each s,a,t. Then $(v,\bar{x})$ is a feasible point in $\mathcal{P}(\lambda)$.

[7] Since the constant c depends on x, this does not exactly suffice to prove our claim. Plainly, the discussion below is not meant to be a substitute for a proof of Theorem 1.

To keep the discussion straightforward, we assume that transitions are independent of actions.[8] Set first $\tilde{w}_t(s,a) := v + \frac{1-\delta}{\delta} x_t(s,a)$. Note that, for each state $s \in S$, the profile α_s is an equilibrium of the one-shot game with payoff function $(1-\delta)r(s,a) + \delta \sum_{t\in S} p(t|s)\tilde{w}_t(s,a)$. The desired continuation payoff vector w will be obtained by adding an action-independent vector to $\tilde{w}$. That is, we will set $w_t(s,a) := \tilde{w}_t(s,a) + C_t(s)$, for some vectors $C_t(s) \in \mathbf{R}^I$. For each choice of $(C_t(s))$, the profile α_s is still a Nash equilibrium of the one-shot game when continuation payoffs are given by w instead of $\tilde{w}$. And it yields a payoff of

$$v_s := (1-\delta)r(s,\alpha_s) + \delta \sum_{t\in S}\sum_{a\in A} \alpha_s(a)p(t|s)w_t(s,a) = v + \delta \sum_{t\in S} p(t|s)C_t(s).$$

We thus need to check that there exist vectors $C_t(s)$, with a norm of the order of $1-\delta$, such that the inequality $\lambda \cdot w_t(s,a) \leq \lambda \cdot v_t$ holds for each (s,a,t). Setting $c_t(s) := \lambda \cdot C_t(s)$, basic algebra shows that the latter inequalities are satisfied as soon as the inequality

$$c_t(s) - \delta \sum_{u\in S} p(u|t)c_u(t) \leq l_t(s) \qquad (\mathcal{S})$$

holds for any pair (s,t) of states, where $l_t(s) := \min_{a\in A} \lambda \cdot (v - \tilde{w}_t(s,a))$. We are left to show that such values of $c_t(s)$ can be found.

Feasibility of (v,x,α) in $\mathcal{P}(\lambda)$ implies that $\sum_{s\in T} l_{\phi(s)}(s) \geq 0$ for every $T \subseteq S$ and every permutation ϕ over T (This is a simple rewriting of condition (ii)). As shown in the Appendix (see Claim 3), this implies that there exists a vector $l^* = (l^*_t(s)) \in \mathbf{R}^{S\times S}$ such that $l^*_t(s) \leq l_t(s)$ for every $s,t \in S$ and $\sum_{s\in T} l^*_{\phi(s)}(s) = 0$ for every $T \subseteq S$ and every permutation ϕ over T.

Given $\delta < 1$ and $t \in S$, we denote by $\mu_{\delta,t} \in \Delta(S)$ the expected, δ-discounted occupancy measure of the Markov chain with initial state t. That is, $\mu_{\delta,t}(s)$ is the expected discounted frequency of visits to state s, when starting from state t. Formally,

$$\mu_{\delta,t}(s) := (1-\delta)\sum_{n=1}^{+\infty} \delta^{n-1}\mathbf{P}_t\left(\mathbf{s}_n = s\right),$$

where $\mathbf{P}_t\left(\mathbf{s}_n = s\right)$ is the probability that the Markov chain visits state s in stage n.

[8] The proof in the general case is available upon request.

Elementary algebra then shows that the vector $c_t(s) := \mathbf{E}_{\mu_{\delta,t}}[l^*_{\mathbf{s}}(s)]$ solves

$$c_t(s) - \delta \sum_{u \in S} p(u|t) c_u(t) = l^*_t(s),$$

for each s, a, t,[9] and is therefore a solution to $(\mathcal{S})$.

We conclude by arguing briefly that the norm of $C_t(s)$ is of the order of $1-\delta$, as desired. Because $l_t(s) = \min_{a \in A} \lambda \cdot (v - \tilde{w}_t(s,a)) = -\frac{1-\delta}{\delta} \max_{a \in A} \lambda \cdot x_t(s,a)$, $l_t(s)$ is of the order of $1-\delta$. It follows from the proof of Claim 3 that l^* is a solution for a linear program whose constraints are linear in $l = (l_t(s))$. Thus l^* and $c_t(s) = \mathbf{E}_{\mu_{\delta,t}}[l^*_{\mathbf{s}}(s)]$ are also of the order of $1-\delta$. It then suffices to choose $C_t(s)$ of the order of $1-\delta$ such that $\lambda \cdot C_t(s) = c_t(s)$.

4.2 Dynamic Programming

It is sometimes argued that the results of Abreu, Pearce and Stacchetti (1990) can be viewed as generalizations of dynamic programming to the case of multiple players. In the absence of any payoff-relevant state variable, this claim is difficult to appreciate within the context of repeated games. By focusing on one-player games, we show that, indeed, our characterization reduces to the optimality equation of dynamic programming. We here assume irreducibility. Irreducibility implies Assumption **A**, which in turn implies that the set of limit points of the family $\{v_\delta(s)\}_{\delta<1}$ as $\delta \to 1$, is independent of s, where $v_\delta(s)$ is the value of the problem with initial state s.

Corollary 1 *In the one-player case with irreducible transition probabilities, the set $\mathcal{H}$ is a singleton $\{v^*\}$, with $v^* = \lim_{\delta \to 1} v_\delta(s)$. Moreover, there is a vector $x^* \in \mathbf{R}^S$ such that*

$$v^* + x^*_s = \max_{a_s \in A} \left(r(s, a_s) + \sum_{t \in T} p(t|s) x^*_t \right) \tag{7}$$

holds for each s, and $v = v^*$ is a unique value solving (7) for some $x \in \mathbf{R}^S$.

This statement is the so-called Average Cost Optimality Equation, see Hoffman and Karp (1966), Sennott (1998) or Kallenberg (2002). To get some intuition for this corollary, note first that, with one player, signals become irrelevant, and we might ignore them. Consider the direction $\lambda = 1$. Furthermore, to maximize his payoff, we should increase the values of $x_t(s)$ as

[9] The computation uses the identity $\mathbf{E}_{\mu_{\delta,t}}[f(\mathbf{s})] = (1-\delta) f(t) + \delta \sum_{u \in S} p(u|t) \mathbf{E}_{\mu_{\delta,u}}[f(\mathbf{s})]$, which holds for any map $f : S \to \mathbf{R}$.

much as possible. So conjecture for a moment that all the constraints **(ii)** bind: for all $T \subseteq S$ and permutations $\phi : T \to T$, $\sum_{s \in T} x_{\phi(s)}(s) = 0$. Let us then set $x_t^* := x_t(\bar{s})$, for some fixed state $\bar{s}$. Then note that, for all $s, t \in S$,

$$x_t(s) = -x_{\bar{s}}(t) - x_s(\bar{s}) = x_t(\bar{s}) - x_s(\bar{s}) = x_t^* - x_s^*,$$

where the first two equalities use the binding constraints. Because a_s is a Nash equilibrium of the game $\Gamma(s, x)$ with payoff v^*, we have

$$v^* = \max_{a_s \in A} \left(r(s, a_s) + \sum_{t \in T} p(t|s) x_t(s) \right).$$

Using $x_t(s) = x_t^* - x_s^*$ gives the desired result. The proof below verifies the conjecture, and supplies missing steps.

Proof. By Proposition 1 (first item), the common set of limit points of $\{v_\delta(s)\}_{\delta<1}$ is a subset of $\mathcal{H}$, so that $\mathcal{H} \neq \varnothing$. We first argue that $\mathcal{H}$ is a singleton.

The set $\mathcal{H}$ is uniquely characterized by the two values $k(\lambda)$, $\lambda \in \{-1, +1\}$. Since $\mathcal{H} \neq \varnothing$, one has $k(1) \geq -k(-1)$.

Note now that any pair (v, x) that is feasible in $\mathcal{P}(\lambda, \alpha)$ is also feasible in $\mathcal{P}(\lambda, a)$, for any $a \in A^S$ such that $\alpha_s(a_s) > 0$ for each s. Consequently, one need only look at pure Markov strategies in order to compute $k(+1)$ and $k(-1)$.

Let $a \in A^S$ be any such strategy. Let (v^+, x^+) be a feasible pair in $\mathcal{P}(+1, a)$, and (v^-, x^-) be a feasible pair in $\mathcal{P}(-1, a)$. By Lemma 1, one has

$$v^+ = \sum_{s \in S} \mu(s) r(s, a_s) + \sum_{\varnothing \neq T \subseteq S} \sum_{\phi : T \to T} \pi_{T,\phi} \left(\sum_{s \in T} x^+_{\phi(s)}(s, a_s) \right),$$

and a similar formula relates v^- and x^-. Since $\sum_{s \in T} x^+_{\phi(s)}(s, a_s) \leq 0$, while $\sum_{s \in T} x^-_{\phi(s)}(s, a_s) \geq 0$ for each T, ϕ, it follows that $v^+ \leq v^-$. Therefore, $k(+1) \leq -k(-1)$, and $\mathcal{H}$ is a singleton. This implies in particular that $v^* := \lim_{\delta \to 1} v_\delta(s)$ exists, for each $s \in S$.

We will use the following claim.

Claim 1 *If a vector* $(x_t(s)) \in \mathbf{R}^{S\times S}$ *satisfies* $\sum_{s\in T} x_{\phi(s)}(s) = 0$ *for all* $T \subseteq S$ *and all permutations* $\phi : T \to T$*, and*

$$v^* \leq r(s, a_s) + \sum_{t\in S} p(t|s, a_s)x_t(s) \tag{8}$$

for some $a = (a_s) \in A^S$*, and all* $s \in S$*, then the inequality (8) holds with equality, for each* s*.*

Proof of the Claim. Assume to the contrary that the inequality (8) is strict for some $\bar{s} \in S$. Let $\mu_a \in \Delta(S)$ be the invariant measure of the Markov chain induced by a over S. By Lemma 1,

$$v^* < \sum_{s\in S} \mu_a(s)r(s, a_s).$$

But this implies that, for all δ high enough, the δ-discounted payoff induced by a exceeds v_δ, which is a contradiction. ■

There is $a^* \in A^S$ and $(x_t(s))$ such that (v^*, x) is feasible in $\mathcal{P}(+1, a^*)$. Pick a vector $x^* = (x^*_t(s)) \in \mathbf{R}^{S\times S}$ such that $x^*_t(s) \geq x_t(s)$ for all $s, t \in S$, and $\sum_{s\in T} x^*_{\phi(s)}(s) = 0$ for all (T, ϕ).[10] We claim that, for each $s \in S$, one has

$$v^* = \max_{a_s\in A} \left(r(s, a_s) + \sum_{t\in T} p(t|s)x^*_t(s) \right). \tag{9}$$

Observe first that, for each given s,

$$v^* = r(s, a^*_s) + \sum_{t\in T} p(t|s)x_t(s) \leq r(s, a^*_s) + \sum_{t\in T} p(t|s)x^*_t(s).$$

Thus, the left-hand side in (9) does not exceed the right-hand side.

Let now a state $s \in S$ and an action $a_s \in A$ be given, that satisfy the inequality

$$v^* \leq r(s, a_s) + \sum_{t\in T} p(t|s)x^*_t(s). \tag{10}$$

Then, by Claim 1, applied to $a := (a^*_{-s}, a_s)$, (10) holds with equality. This proves that (9) holds for each $s \in S$.

[10] See Claim 3 in Section 9.5 for the existence of x^*.

Since $\sum_{s\in T} x^*_{\phi(s)}(s) = 0$ for all (T,ϕ), there is a vector $x^* \in \mathbf{R}^S$ such that $x^*_t(s) = x^*_t - x^*_s$, for each (s,t).

Uniqueness of v^* follows at once. Assume indeed that (7) holds for some (v,x). Set $\tilde{x}_t(s) := x_t - x_s$ for each $s,t \in S$. Then the pair $(v,\tilde{x})$ is feasible in $\mathcal{P}(+1)$ and in $\mathcal{P}(-1)$ as well. Hence, $-k(-1) \leq v \leq k(+1)$, so that $v = v^*$. ■

4.3 Interpretation of the Variables $x_t(s,y)$

The variables $x_t(s,y)$ are not continuation payoffs *per se*. Rather, they are payoff differences that account both for the signal and the possible change of state. In the case of a repeated game, they reduce to a variable of the signal alone (in the notation of FL, they are then equal to $\frac{\delta}{1-\delta}(w(y)-v)$). This variable reflects how the continuation payoff adjusts, from the current to the following period, to provide the appropriate incentives, as a function of the realized signal. In the case of dynamic programming, these variables collapse to a function $x_t(s)$. This is the relative value function, as it is known in stochastic dynamic programming, and it captures the value of the Markov decision process in state t relative to state s. It can be further decomposed into a difference $\hat{x}(s) - \hat{x}(t)$, for some function $\hat{x}$ that only depends on the current state.

While there is no reason to expect the system of inequalities **(ii)** to simplify in general, there are some special cases in which it does. For instance, given our discussion above, one might suspect that the payoff adjustments required by the provision of incentives, on one hand, and by the state transitions, on the other, can be disentangled whenever transitions are uninformative about actions, conditional on the signals. Indeed, both in the case of action-independent transitions, studied in Section 7, and in the case of perfect monitoring, one can show that these variables can be separated as $\hat{x}_t(s) + \tilde{x}(s,y)$, for some function $\hat{x}$ that only depends on the current and the next state, and some $\tilde{x}$ that only depends on the current state and the realized signal.

5 The Folk Theorem

FLM establish a folk theorem for repeated games with imperfect public monitoring when the signal distribution satisfies some rank condition. In this section, we extend their folk theorem to stochastic games. We derive our folk theorem by investigating the programs $\mathcal{P}(\lambda)$ under a similar rank condition and relating scores $k(\lambda)$ to feasible sets and to minmax payoffs.

In this section, we do not rely on Assumption **A**. Instead, it is more convenient to impose state independence on feasible sets and minmax values when necessary.

Let $F_\delta(s)$ be the convex hull of the set of feasible payoffs of the game with initial state $s \in S$ and discount factor $\delta < 1$. The set $F_\delta(s)$ is compact, and converges to $F(s)$ as $\delta \to 1$, where $F(s)$ is the set of the *limit-average* feasible payoffs with initial state s (see, for instance, Dutta, 1995, Lemma 2). Let also $m^i_\delta(s)$ be player i's minmax payoff in the game with initial state s and discount factor δ, defined as

$$m^i_\delta(s) := \min_{\sigma^{-i}} \max_{\sigma^i} \sum_{n=1}^{\infty} (1-\delta)\delta^{n-1} \mathbf{E}_{s_1,\sigma}[r^i(s_n, a_n)],$$

where the minimum is taken over public strategies σ^{-i}. As $\delta \to 1$, $m^i_\delta(s)$ converges to $m^i(s)$, where $m^i(s)$ is player i's limit-average minmax payoff with initial state s (see Mertens and Neyman, 1981 and Neyman, 2003).

Define the intersection of the sets of feasible and individually rational payoffs in state s:

$$F^* := \bigcap_{s \in S} \{v \in F(s) : v^i \geq m^i(s) \ \forall i \in I\}.$$

For a given state $s \in S$ and a Markov strategy $\alpha^{-i} = (\alpha^j)_{j \neq i}$, let $\Pi^i(s, \alpha^{-i})$ be the $|A^i| \times |S \times Y|$ matrix whose $(a^i, (t, y))$-th component is given by $p(t, y|s, a^i, \alpha^{-i})$. For $s \in S$ and $\alpha = (\alpha^i)$, we stack two matrices vertically:

$$\Pi^{ij}(s, \alpha) := \begin{pmatrix} \Pi^i(s, \alpha^{-i}) \\ \Pi^j(s, \alpha^{-j}) \end{pmatrix}.$$

An action profile $\alpha \in \times_i \Delta(A^i)$ has *individual full rank for player i in state s* if $\Pi^i(s, \alpha^{-i})$ has rank $|A^i|$; the profile α has *pairwise full rank for players i and j in state s* if $\Pi^{ij}(s, \alpha)$ has rank $|A^i| + |A^j| - 1$. Note that $|A^i| + |A^j| - 1$ is the highest possible rank since $\Pi^{ij}(s, \alpha)$ always has at least one non-trivial linear relation among its row vectors.

Assumption F1: Every pure action profile has individual full rank for every player in every state.

Assumption F2: For each state s and pair (i, j) of players, there exists a mixed action profile that has pairwise full rank for players i and j in state s.

The assumptions are the obvious generalizations of the assumptions of individual and pairwise full rank made by FLM. Note that Assumptions **F1** and **F2** are weaker than the rank assumptions of Fudenberg and Yamamoto (2010). Fudenberg and Yamamoto require that players can statistically identify each others' deviations via actual signals y, whereas we allow players to make inferences from the observed state as well.

Example 1 in Section 3.2 provides a useful illustration of the difference. In this example, there are no public signals, so Fudenberg and Yamamoto's rank assumptions are not satisfied. On the other hand, Assumptions **F1** and **F2** are satisfied if $p^L \neq p^R$.[11] If $p^L = p^R$, then incentives cannot be provided for players to play R, so that the unique PPE payoff is $(1, 1)$.

With the above assumptions, we characterize $k(\lambda)$ in terms of feasible and minmax payoffs only. Let e^i denote the i-th coordinate basis vector in $\mathbf{R}^I$.

Lemma 2 *Under Assumptions **F1-F2**, one has the following.*

1. *If $\lambda \in S^1$ and $\lambda \neq -e^i$ for any i, then $k(\lambda) = \min_s \max_{w \in F(s)} \lambda \cdot w$.*
2. *$k(-e^i) = -\max_s m^i(s)$ for any i.*

While this lemma characterizes the value of the optimization program for each direction λ under Assumptions **F1-F2**, the algorithm can be adapted to the purpose of computing feasible and minmax payoffs (in public strategies) without these assumptions. In the first case, it suffices to ignore the incentive constraints **(i)** in the program $\mathcal{P}(\lambda)$ and to take the intersection of the resulting half-spaces. In the second case, it suffices to focus on the incentives of the minmaxed player i, and to take as direction the coordinate vector $-e^i$. The proofs follow similar lines, and details are available from the authors.

Combined with Proposition 2, Lemma 2 implies the following folk theorem, which extends both the folk theorem for repeated games with imperfect public monitoring by FLM and the folk theorem for stochastic games with observable actions by Dutta (1995).[12]

Theorem 2 (Folk Theorem) *Under Assumptions **F1-F2**, it holds that $\mathcal{H} = F^*$. In particular, if $F(s) = F$ and $m^i(s) = m^i$ for all $s \in S$, and $F^* = \{v \in F : v^i \geq m^i \ \forall i \in I\}$ has non-empty interior, then $E_\delta(s)$ converges to F^* as $\delta \to 1$, for any $s \in S$.*

[11] When the set of available actions depends on the state, as in Example 1, the definitions of full rank must be adjusted in the obvious way. Namely, we say that α has individual full rank for player i at state s if the rank of $\Pi^i(s, \alpha^{-i})$ is no less than the number of actions available to player i at state s. A similar modification applies to pairwise full rank.

[12] Note that Dutta (1995, Theorem 9.3) shows that full-dimensionality can be weakened to payoff asymmetry if mixed strategies are observable, an assumption that makes little sense under imperfect monitoring.

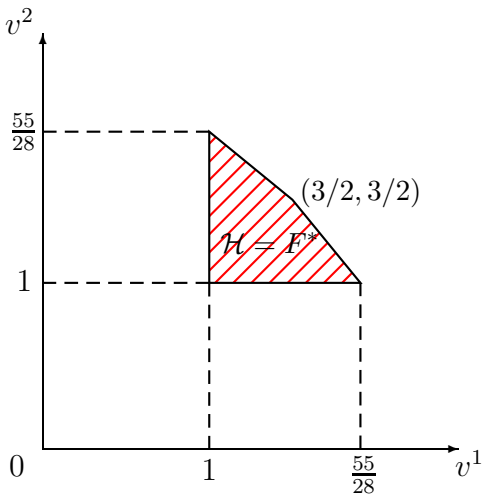

Figure 3: The limit set of PPE payoffs in Example 1

It is known that, under irreducibility, $F(s)$ and $m^i(s)$ are independent of state s. In Example 1 in Section 3.2, irreducibility is satisfied if $p^L, p^R \neq 1$. Dutta (1995) provides somewhat weaker assumptions on the transition function that guarantee the state-independence property. But these assumptions are not necessary: the limit set of feasible payoffs is independent of the initial state as long as at least one of these probabilities is strictly less than one, and the minmax payoff is always independent of the initial state.

To illustrate this folk theorem, let for instance $p^L = 1 - p^R = 1/10$ in Example 1, so that playing action L gives player i both his highest reward in that state and the highest probability of transiting to the other state, in which his reward is for sure higher than in his own state. The folk theorem holds here. The set of PPE payoffs is then the convex hull of $(1, 1)$, $(3/2, 3/2)$, $(55/28, 1)$ and $(1, 55/28)$. See Figure 3.

6 Short-run Players

It is trivial to extend the algorithm to the case in which some players are short-run. Following FL, suppose that players $i = 1, \ldots, L$, $L \leq I$, are long-run players, whose objective is to maximize the average discounted sum of rewards, with discount factor $\delta < 1$. Players $j \in SR := \{L + 1, \ldots, I\}$ are short-run players, each representative of which plays only once. For each state $s \in S$, let

$$B(s) : \times_{i=1}^{L} \Delta(A^i) \to \times_{j=L+1}^{I} \Delta(A^j)$$

be the correspondence that maps any mixed action profile $(\alpha^1, \ldots, \alpha^L)$ for the long-run players to the corresponding static equilibria for the short-run players. That is, for each $\alpha \in \text{graph}B(s)$,

	P	NP
H	$\tau, 1-\tau-c$	$0,0$
C	$1,-c$	$0,0$

Figure 4: A Political Game

and each $j > L$, α^j maximizes $r^j(s, \cdot, \alpha^{-j})$. The characterization goes through if we "ignore" the short-run players and simply modify **(i)** by requiring that v be a Nash equilibrium payoff of the game $\Gamma(s, x)$ for the long-run players, achieved by some $\alpha_s \in \text{graph}B(s)$ for each s.

6.1 An Example

We now provide an illustration of the algorithm that attempts to tread the thin line between accessibility and triviality. Consider the following game, loosely inspired by Dixit, Grossman and Gul (2000) and Phelan (2006).

There are two parties. In any given period, one of the two parties is in power. Party $i = 1, 2$ is in power in state i. Only the party in power and the households take actions. Households can produce (P) at cost c with value one, or not produce (NP). There is a continuum of households, and we treat them therefore as one short-run player. The government in power can either honor (H) or confiscate (C). Honoring means choosing a fixed tax rate τ and getting therefore revenues $\tau\mu^i$, where μ^i is the fraction of households who produce in state i. By confiscating, the government appropriates all output. This gives rise to the payoff matrix given by Figure 4.

It is assumed that $1 - \tau > c > 0$. Actions are not observed, but parties that honor are more likely to remain in power. More precisely, if the state is i, and the realized action is H, the state remains i with probability p^H; if the action is C, it remains the same with probability p^L, with $0 < p^L < p^H < 1$. We call this game Example 2.

Note that, given the households' preferences, the best-reply correspondence in state i is

$$B(i)(\alpha^i) = \begin{cases} [0,1] & \text{if } \alpha^i = c/(1-\tau), \\ \{0\} & \text{if } \alpha^i < c/(1-\tau), \\ \{1\} & \text{if } \alpha^i > c/(1-\tau), \end{cases}$$

where α^i is the probability that party i plays H. The feasible set F is independent of the initial state, and equal to the convex hull of the payoff vectors $(0,0)$, $((1-c)/2, (1-c)/2)$, $(\bar{v}, 0)$ and

$(0, \bar{v})$, where

$$\bar{v} := \frac{1-p^L}{2-\bar{p}-p^L}(1-c),\ \bar{p} := \frac{c}{1-\tau}p^H + \left(1-\frac{c}{1-\tau}\right)p^L.$$

These expressions are intuitive. Note that the highest symmetric payoff involves both parties confiscating as much as possible, subject to the constraint that the households are still willing to produce. The payoff from confiscating at this rate is $1-c$, and since time is equally spent between both parties, the resulting payoff is $((1-c)/2, (1-c)/2)$. Consider next the case in which one party always confiscates, so that households never produce in that state, while the other party confiscates at the highest rate consistent with all the households producing. The invariant distribution assigns probability $(1-p^L)/(2-\bar{p}-p^L)$ to that state, and the asymmetric payoff follows. Finally, the minmax payoff of each party is zero, which is an equilibrium payoff.

Theorem 1 applies to this game. Let us compute the equilibrium payoff set as $\delta \to 1$. The optimization program with weights $\lambda = (\lambda^1, \lambda^2)$ involves eight variables $x_t^i(s)$, $i = 1, 2$, $s, t = 1, 2$, satisfying the constraints

$$\lambda \cdot x_1(1) \leq 0,\ \lambda \cdot x_2(2) \leq 0,\ \lambda \cdot (x_2(1) + x_1(2)) \leq 0, \tag{11}$$

in addition to the requirement that α^i be a Nash equilibrium of the game $\Gamma(i, x)$, $i = 1, 2$.

Consider a vector $\lambda > 0$. Note that the constraints (11) must bind: Indeed, note that, because player i does not make a choice in state $-i$, the Nash equilibrium requirements (i.e. constraints **(i)** in program $\mathcal{P}(\lambda)$) give us at most three constraints per player (his preference ordering in the state in which he takes an action, and the fact that he gets the same payoff in the other state). In addition to the three constraints (11), this gives us nine constraints, and there are eight variables $x_t^i(s)$. One can check that, if one of the three constraints is slack, we can increase the payoff of a player by changing the values of $x_t^i(s)$, while satisfying all binding constraints. Observe now that we must have $\mu^i \in \{0, 1\}$, $i = 1, 2$. Indeed, suppose that $\mu^i \in (0, 1)$, for some i, so that $\alpha^i \geq c/(1-\tau)$. Then we can increase μ^i and decrease $x_i^i(i), x_i^i(-i)$ so as to keep player i's payoff constant, while keeping him indifferent between both actions (which he must be since $\mu^i \in (0, 1)$). Given these observations, this now becomes a standard problem that can be solved by enumeration. Note, however, that we do not need to consider the case $\alpha^1 > c/(1-\tau), \alpha^2 > c/(1-\tau)$: if this were the case, we could decrease both these probabilities in such a way as to maintain the relative time spent in each state the same, while increasing both players' payoffs in their own state (because $\mu^i = 1$ is feasible as long as $\alpha^i \geq c/(1-\tau)$).

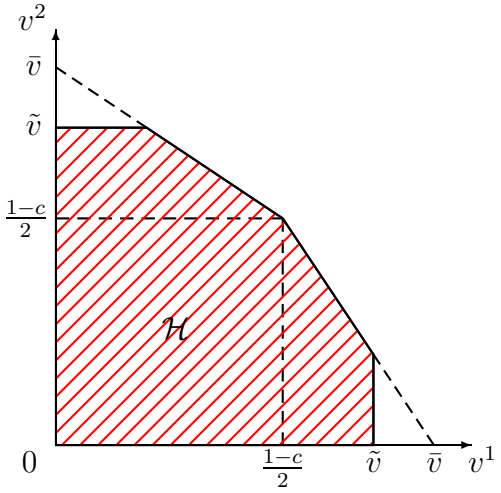

Figure 5: The limit set of equilibrium payoffs in Example 2

It is not hard to guess that the optimal action profile for $\lambda^1 = \lambda^2 > 0$ is to set $\alpha^i = c/(1-\tau)$, $i = 1, 2$ (and, as mentioned, $\mu^i = 1$), and we obtain the highest feasible symmetric payoff. If we consider a coordinate direction, say $\lambda^1 > 0, \lambda^2 = 0$, it is intuitive that households will not produce in state 2, and party 2 will confiscate with sufficient probability for such behavior to be optimal, and party 1 will confiscate at the highest rate consistent with households producing. Party 1 must not be willing to either confiscate for sure (which increases his reward when he does) or honor for sure (which increases the fraction of time spent in his state, when his reward is positive), and this means that his payoff cannot exceed

$$\tilde{v} := \min \left\{ \frac{p^H - \tau p^L - (1-\tau)}{p^H - p^L}, \frac{(1-p^L)(p^H - \tau p^L)}{(2-p^L)(p^H - p^L)} \right\},$$

assuming this is positive. It follows that the limit equilibrium payoff set is given by $\mathcal{H} = \{v \in F : 0 \leq v^i \leq \tilde{v} \ \forall i = 1, 2\}$ if $\tilde{v} > 0$, and the singleton payoff $(0, 0)$ otherwise. Note that, from the first argument defining $\tilde{v}$, the payoff set shrinks to $(0, 0)$ as $p^H \to p^L$, as is to be expected: if confiscation cannot be statistically detected, households will prefer not to produce. See Figure 5 for an illustration in the case in which $\tilde{v} > (1-c)/2$.

6.2 A Characterization with Full Rank

The equilibrium payoff set obtained in Example 2 is reminiscent of the results of FL for repeated games with long-run and short-run players. We provide here an extension of their result to the case of stochastic games, but maintain full rank assumptions for long-run players.

One of the main insights of FL, which is an extension of Fudenberg, Kreps and Maskin

(1990), is that achieving some payoffs may require the long-run players to randomize on the equilibrium path, so as to induce short-run players to take particular actions. This means that long-run player i must be indifferent between all actions in the support of his mixture, and so continuation payoffs must adjust so that his payoff cannot exceed the one he would achieve from the action in this support that yields him the lowest payoff.

For a direction $\lambda \neq \pm e^i$, the pairwise full rank condition ensures that the above requirement is not restrictive, that is, we only need to care about the feasibility given that short-run players take static best responses. However, for the coordinate directions $\lambda = \pm e^i$, since we cannot adjust long-run player i's continuation payoff without affecting the score, the optimal mixed action of player i is determined taking into account both the effect on the short-run players' actions and the cost of letting player i take that action.

The same applies here. We first deal with non-coordinate directions. To incorporate short-run players' incentives, we define $F_\delta(s)$ as the convex hull of the set of long-run players' feasible payoffs of the game with initial state s and discount factor δ, where we require that the players play actions from graph$B(t)$ whenever the current state is t. The set $F(s)$ of long-run players' limit-average feasible payoffs is defined similarly. The set $F_\delta(s)$ converges to $F(s)$ as $\delta \to 1$.

Assumption F3: For each state s and pair (i, j) of long-run players, every mixed action profile in graph$B(s)$ has pairwise full rank for players i and j in state s.

Similarly to Lemma 2, we have the following.

Lemma 3 *Under Assumption **F3**, one has the following.*

1. *If $\lambda \in S^1$ and $\lambda \neq \pm e^i$ for any i, then $k(\lambda) = \min_{s \in S} \max_{w \in F(s)} \lambda \cdot w$.*
2. *$\mathcal{H} = \left\{ v \in \bigcap_{s \in S} F(s) : -k(-e^i) \leq v^i \leq k(e^i)\ \forall i = 1, \ldots, L \right\}$.*

Theorem 1 applies here. Namely, if Assumption **A** is satisfied and $\mathcal{H}$ has non-empty interior, then $E_\delta(s)$ converges to $\mathcal{H}$ as $\delta \to 1$, for every $s \in S$.

Lemma 3 leaves open how to determine $k(\pm e^i)$. Under Assumption **F3**, since every mixed action profile has individual full rank, we can control each long-run player's payoffs and induce him to play any action. This implies that, without loss of generality, in program $\mathcal{P}(\pm e^i)$, we can ignore the incentives of long-run players other than player i without affecting the score.

We can obtain a further simplification if Assumption **F3** is strengthened to full monitoring. In this case, $k(e^i)$ is equal to the value of the problem $\mathcal{Q}^i_+$:

$$\sup v^i,$$

where the supremum is taken over all $\alpha = (\alpha_s)_{s\in S} \in \times_s \mathrm{graph}B(s)$, $v^i \in \mathbf{R}$, and $\bar{x}^i : S \to \mathbf{R}^S$ such that

(i) For each s,

$$v^i = \min_{a^i_s \in \mathrm{support}(\alpha^i_s)} \left(r^i(s, \alpha^{-i}_s, a^i_s) + \sum_{t\in S} p(t|s, a^i_s, \alpha^{-i}_s)\bar{x}^i_t(s) \right);$$

(ii) For each $T \subseteq S$, for each permutation $\phi : T \to T$, one has $\sum_{s\in T} \bar{x}^i_{\phi(s)}(s) \leq 0$.

Here v^i is equal to the minimum of player i's payoffs, where a^i_s is taken over the support of player i's mixed action α^i_s. If v^i were larger than the minimum, then we would not be able to provide incentives for player i to be indifferent between all actions in the support of α^i_s. Note that this program is simpler than $\mathcal{P}(e^i)$ in that $\bar{x}^i$ depends on the current and next states, but is independent of realized actions.

Similarly, $-k(-e^i)$ is equal to the solution for the problem $\mathcal{Q}^i_-$:

$$\inf v^i,$$

where the infimum is taken over all $\alpha \in \times_s \mathrm{graph}B(s)$, $v^i \in \mathbf{R}$, and $\bar{x}^i : S \to \mathbf{R}^S$ such that

(i) For each s,

$$v^i = \max_{a^i_s \in A^i} \left(r^i(s, \alpha^{-i}_s, a^i_s) + \sum_{t\in S} p(t|s, a^i_s, \alpha^{-i}_s)\bar{x}^i_t(s) \right);$$

(ii) For each $T \subseteq S$, for each permutation $\phi : T \to T$, one has $\sum_{s\in T} \bar{x}^i_{\phi(s)}(s) \geq 0$.

See also Section 4.2 for the discussion of the related, one-player case.

7 Action-independent Transitions

Our characterization can be used to obtain qualitative insights. We here discuss the case in which the evolution of the state is independent of players' actions. To let the transitions

probabilities of the states vary, while keeping the distributions of signals fixed, we assume that the transition probabilities can be written as a product $p(t|s) \times \pi(y|s,a)$. In other words, and in any period, the next state t and the public signal y are drawn independently, and the action profile only affects the distribution of y.

In this setup, it is intuitively clear that the (limit) set of PPE payoffs should only depend on p through its invariant measure, μ. If the signal structure is rich enough and the folk theorem holds, this is straightforward. Indeed, the (limit) set of feasible payoffs is equal to $\sum_{s\in S} \mu_s r(s, \Delta(A))$, and a similar formula holds for the (limit) minmax.

We prove below that this observation remains valid even when the folk theorem fails to hold. The proof is based on our characterization, and providing a direct proof of this fact does not seem to be an easy task.

In this subsection, we fix a signal structure $\pi : S \times A \to \Delta(Y)$, and we let the transition probability p over states vary. To stress the dependence, and given a (unit) direction $\lambda \in S^1$, and a mixed profile α, we denote by $\mathcal{P}_p(\lambda, \alpha)$ the optimization program $\mathcal{P}(\lambda, \alpha)$, when the transition probability over states is set to $p : S \to \Delta(S)$. We also denote by $k_p(\lambda, \alpha)$ the value of $\mathcal{P}_p(\lambda, \alpha)$, and by $\mathcal{H}(p)$ the intersection of all half-spaces $\{v \in \mathbf{R}^I : \lambda \cdot v \leq \sup_\alpha k_p(\lambda, \alpha)\}$ obtained by letting λ vary.

Proposition 3 *Let p and q be two irreducible transition functions over S, with the same invariant measure μ. Then $\mathcal{H}(p) = \mathcal{H}(q)$.*

This implies that, for the purpose of studying the limit equilibrium payoff set in the case of transitions that are action-independent, one might as well assume that the state is drawn i.i.d. across periods. In that case, the stochastic game can be viewed as a repeated game, in which player i's actions in each period are maps from S to A^i. The program $P(\lambda, \alpha)$ can then be shown to be equivalent to the corresponding program of FL. One direction is a simple change of variable. The other direction relies on Lemma 1.

8 Concluding Comments

This paper shows that some of the methods developed for the study of repeated games can be generalized to stochastic games, and can be applied to obtain qualitative insights as $\delta \to 1$.

This, of course, leaves many questions unanswered. First, as mentioned, the results derived here rely on $\delta \to 1$. While not much is known in the general case for repeated games either, there

still are a few results available in the literature (on the impact of the quality of information, for instance). It is then natural to ask whether those results can be extended to stochastic games.

Within the realm of asymptotic analysis, it also appears important to generalize our results to broader settings. The characterization of the equilibrium payoff set and the folk theorem established in this paper rely on some strong assumptions. The set of actions, signals and states are all assumed to be finite. We suspect that the characterization can be generalized to the case of richer action and signal sets, but extending the result to richer state spaces raises significant challenges. Yet this is perhaps the most important extension: even with finitely many states, beliefs about those states affect the players' incentives when states are no longer common knowledge, as is often the case in applications in which players have private information, and these beliefs must therefore be treated as state variables themselves.

Finally, there is an alternative possible asymptotic analysis that is of significant interest. As is the case with public signals (see the concluding remarks of FLM), an important caveat is in order regarding the interpretation of our limit results. It is misleading to think of $\delta \to 1$ as periods growing shorter, as transitions are kept fixed throughout the analysis. If transitions are determined by physical constraints, then these probabilities should be adjusted as periods grow shorter. As a result, feasible payoffs will not be independent of the initial state. It remains to be seen whether such an analysis is equally tractable.

9 Proofs

9.1 Proof of Proposition 1

Let $\delta < 1$ be given. To show the first statement, for each state s, choose an equilibrium payoff v_s that maximizes the inner product $\lambda \cdot w$ for $w \in E_\delta(s)$. Consider a PPE σ with payoffs (v_s). For each initial state s, let $\alpha_s = \sigma(s)$ be the mixed moves played in stage 1, and $w_t(s,y)$ be the continuation payoffs following y, if in state t at stage 2. Note that $\lambda \cdot w_t(s,y) \leq \lambda \cdot v_t$ for $y \in Y$ and $s,t \in S$. By construction, for each $s \in S$, α_s is a Nash equilibrium, with payoff v_s, of the game with payoff function

$$(1-\delta)r(s,a) + \delta \sum_{t \in S, y \in Y} p(t,y|s,a) w_t(s,y).$$

Set $x_t(s,y) := \frac{\delta}{1-\delta}(w_t(s,y) - v_s)$. Observe that α_s is a Nash equilibrium of the game $\Gamma(s,x)$, with payoff v_s. Next, let $\bar{s}$ be a state that minimizes $\lambda \cdot v_s$, and $\tilde{x}_t(s,y) := x_t(s,y) + v_{\bar{s}} - v_s$ for each $s,t \in S$, $y \in Y$. Then α_s is a Nash equilibrium of $\Gamma(s,\tilde{x})$, with payoff $v_{\bar{s}}$. On the other hand, for each T, ϕ, ψ, one has

$$\begin{aligned}\sum_{s \in T} \lambda \cdot x_{\phi(s)}(s,\psi(s)) &= \frac{\delta}{1-\delta} \sum_{s \in T} \lambda \cdot (w_{\phi(s)}(s,\psi(s)) - v_s) \\ &= \frac{\delta}{1-\delta} \sum_{s \in T} \lambda \cdot (w_{\phi(s)}(s,\psi(s)) - v_{\phi(s)}) \leq 0,\end{aligned}$$

and hence

$$\lambda \cdot \sum_{s \in T} \tilde{x}_{\phi(s)}(s,\psi(s)) = \sum_{s \in T} \lambda \cdot x_{\phi(s)}(s,\psi(s)) + \sum_{s \in T} (\lambda \cdot v_{\bar{s}} - \lambda \cdot v_s) \leq 0.$$

Therefore, $(v_{\bar{s}}, \tilde{x})$ is feasible in $\mathcal{P}(\lambda)$. Thus

$$k(\lambda) \geq \lambda \cdot v_{\bar{s}} = \min_{s \in S} \max_{w \in E_\delta(s)} \lambda \cdot w.$$

The second statement follows immediately from the first statement.

9.2 Proof of Proposition 2

Since Z is a compact set contained in the interior of $\mathcal{H}$, there exists $\eta > 0$ such that

$$Z_\eta := \{z \in \mathbf{R}^I : d(z, Z) \leq \eta\}$$

is also a compact set contained in the interior of $\mathcal{H}$. We start with a technical statement.

Lemma 4 *There are $\varepsilon_0 > 0$ and a bounded set $K \subset \mathbf{R}^I \times \mathbf{R}^{S\times Y\times S\times I}$ of (v, x) such that the following holds. For every $z \in Z_\eta$ and $\lambda \in S^1$, there exists $(v, x) \in K$ such that (v, x) is feasible in $\mathcal{P}(\lambda)$ and $\lambda \cdot z + \varepsilon_0 < \lambda \cdot v$.*

Proof. Given $z \in Z_\eta$ and since Z_η is contained in the interior of $\mathcal{H}$, one has $\lambda \cdot z < k(\lambda)$ for every $\lambda \in S^1$. Therefore, there exists a feasible pair (v, x) in $\mathcal{P}(\lambda)$ such that $\lambda \cdot z < \lambda \cdot v$. The conclusion of the lemma states that (v, x) can be chosen within a bounded set, independently of λ and of z.[13]

Choose $\tilde{\varepsilon} > 0$ such that $\lambda \cdot z + \tilde{\varepsilon} < \lambda \cdot v$, and define

$$\tilde{x}_t(s, y) = x_t(s, y) - \tilde{\varepsilon}\lambda, \quad \tilde{v} = v - \tilde{\varepsilon}\lambda$$

for each $s, t \in S$ and $y \in Y$. Observe that for each $s \in S$, α_s is a Nash equilibrium of the game $\Gamma(s, \tilde{x})$, with payoff $\tilde{v}$. Note in addition that $\sum_{s\in T} \lambda \cdot \tilde{x}_{\phi(s)}(s, \psi(s)) \leq -\tilde{\varepsilon}|T| < 0$ for each T, ϕ and ψ. Therefore, for every $\tilde{z}$ close enough to z, for every $\tilde{\lambda}$ close enough to λ, the pair $(\tilde{v}, \tilde{x})$ is feasible in $\mathcal{P}(\tilde{\lambda})$ and $\tilde{\lambda} \cdot \tilde{z} < \tilde{\lambda} \cdot \tilde{v}$. The result then follows, by compactness of Z_η and of S^1. ■

In the sequel, we let $\kappa_0 > 0$ be such that $\|x\| \leq \kappa_0$ and $\|z - v\| \leq \kappa_0$ for every $(v, x) \in K$ and $z \in Z_\eta$. Choose $n \in \mathbf{N}$ such that $\varepsilon_0(n-1)/2 > 2\kappa_0|S|$. Set

$$\varepsilon := \varepsilon_0 \frac{n-1}{2} - 2\kappa_0|S| > 0.$$

Next, choose $\bar{\delta} < 1$ to be large enough so that $(n/2)^2(1-\delta) \leq |S|$ and $1 - \delta^{n-1} \geq (n-1)(1-\delta)/2$ for every $\delta \geq \bar{\delta}$. Finally, set

$$\kappa := \frac{2(n-1)}{\bar{\delta}^{n-1}}\kappa_0.$$

[13] Note however that, for given z and λ, the set of feasible pairs (v, x) in $\mathcal{P}(\lambda)$ such that $\lambda \cdot z < \lambda \cdot v$ is typically unbounded.

Given a map $w : H_n \to \mathbf{R}^I$ which associates to any history of length n a payoff vector, we denote by $\Gamma^n(s, w; \delta)$ the δ-discounted, $(n-1)$-stage game, with final payoffs w.

The following proposition is essential for the proof of Proposition 2.

Proposition 4 *For every direction $\lambda \in S^1$, every $z \in Z_\eta$ and every discount factor $\delta \geq \bar{\delta}$, there exist continuation payoffs $w : H_n \to \mathbf{R}^I$ such that:*

C1 *For each s, z is a PPE payoff of the game $\Gamma^n(s, w; \delta)$.*

C2 *One has $\|w(h) - z\| \leq (1-\delta)\kappa$ for every $h \in H_n$.*

C3 *One has $\lambda \cdot w(h) \leq \lambda \cdot z - (1-\delta)\varepsilon$ for every $h \in H_n$.*

Proof. Let λ, z and δ be given as stated. Pick a mixed profile $\alpha = (\alpha_s)_{s \in S}$ and $(v, x) \in K$ that is feasible in $\mathcal{P}(\lambda)$ such that, for each $s \in S$, α_s is a Nash equilibrium of $\Gamma(s, x)$, with payoff v, and $\lambda \cdot z + \varepsilon_0 < \lambda \cdot v$.

Given $s, t \in S$, and $y \in Y$, set

$$\phi_t(s, y) = v + \frac{1-\bar{\delta}}{\bar{\delta}} x_t(s, y).$$

For each history h of length not exceeding n, we define $w(h)$ by induction on the length of h. The definition of w follows FL. If $h = (s_1)$, we set $w(h) = z$. For $k \geq 1$ and $h_{k+1} \in H_{k+1}$, we set

$$w(h_{k+1}) = \frac{\delta - \bar{\delta}}{\delta(1-\bar{\delta})} w(h_k) + \frac{\bar{\delta}(1-\delta)}{\delta(1-\bar{\delta})} \left(\phi_{s_{k+1}}(s_k, y_k) + \frac{w(h_k) - v}{\bar{\delta}} \right) \tag{12}$$

where h_k is the restriction of h_{k+1} to the first k stages.

Let $k \geq 1$, and $h_k \in H_k$. As in FL, using (12), α_{s_k} is a Nash equilibrium of the one-shot game with payoff function

$$(1-\delta) r(s_k, a) + \delta \sum_{t \in S} \sum_{y \in Y} p(t, y | s_k, a) w(h_k, y, t),$$

with payoff $w(h_k)$. By the one-shot deviation principle, it follows that Markov strategy profile α is a PPE of the game $\Gamma^n(s_1, w; \delta)$, with payoff z (for each initial state s_1). This proves **C1**.

We now turn to **C2**. Let $h_n \in H_n$ be given and let h_k ($k \leq n$) denote the restriction of h_n to the first k stages. Observe that

$$\begin{aligned} w(h_{k+1}) - v &= \frac{\delta - \bar{\delta}}{\delta(1-\bar{\delta})}\left(w(h_k) - v\right) + \frac{\bar{\delta}(1-\delta)}{\delta(1-\bar{\delta})}\left(\phi_{s_{k+1}}(s_k, y_k) - v + \frac{1}{\bar{\delta}}\left(w(h_k) - v\right)\right) \\ &= \frac{1}{\delta}\left(w(h_k) - v\right) + \frac{1-\delta}{\delta} x_{s_{k+1}}(s_k, y_k) \end{aligned}$$

for any $k \leq n-1$. Therefore,

$$w(h_n) - v = \frac{1}{\delta^{n-1}}(z - v) + \frac{1-\delta}{\delta}\sum_{k=1}^{n-1} \frac{1}{\delta^{n-1-k}} x_{s_{k+1}}(s_k, y_k),$$

and one gets

$$w(h_n) - z = \frac{1-\delta^{n-1}}{\delta^{n-1}}(z - v) + \frac{1-\delta}{\delta^{n-1}}\sum_{k=1}^{n-1} \delta^{k-1} x_{s_{k+1}}(s_k, y_k). \tag{13}$$

Hence

$$\begin{aligned} \|w(h_n) - z\| &\leq \frac{1-\delta^{n-1}}{\delta^{n-1}}\|z - v\| + \frac{1-\delta}{\delta^{n-1}}\sum_{k=1}^{n-1} \delta^{k-1}\|x_{s_{k+1}}(s_k, y_k)\| \\ &\leq \frac{2(1-\delta^{n-1})}{\delta^{n-1}}\kappa_0 \\ &\leq \frac{2(n-1)(1-\delta)}{\bar{\delta}^{n-1}}\kappa_0 = (1-\delta)\kappa. \end{aligned}$$

This proves **C2**.

We finally prove that **C3** holds as well. The proof makes use of the following lemma.

Lemma 5 *Let $x_1, \ldots, x_m \in [-1, 1]$ be such that $\sum_{k=1}^{m} x_k \leq 0$. Then $\sum_{k=1}^{m} \delta^{k-1} x_k \leq \frac{(1-\delta^{m/2})^2}{1-\delta}$.*

Proof. If m is even, the sum $\sum_{k=1}^{m} \delta^{k-1} x_k$ is highest if $x_k = 1$ for $k \leq \frac{m}{2}$, and $x_k = -1$ for $k > \frac{m}{2}$. If m is odd, the sum is maximized by setting $x_k = 1$ for $k < \frac{m+1}{2}$, $x_k = 0$ for $k = \frac{m+1}{2}$ and $x_k = -1$ for $k > \frac{m+1}{2}$. In both cases, the sum is at most $\frac{1-2\delta^{m/2}+\delta^m}{1-\delta}$. ■

Set $m_0 = 0$, $l_1 + 1 := \min\{k \geq 1 : s_m = s_k \text{ for some } m > k\}$ ($\min \varnothing = +\infty$), and $m_1 + 1 := \max\{k \leq n : s_k = s_{l_1+1}\}$. Next, as long as $l_j < +\infty$, define $l_{j+1} + 1 := \min\{k \geq m_j + 1 : s_m =$

s_k for some $m > k\}$ and $m_{j+1} + 1 := \max\{k \leq n : s_k = s_{l_{j+1}+1}\}$. Let J the largest integer j with $l_j < +\infty$.

Since $s_{l_j+1} = s_{m_j+1}$, one has $\lambda \cdot \sum_{k=l_j+1}^{m_j} x_{s_{k+1}}(s_k, y_k) \leq 0$ for each $j \leq J$. By Lemma 5, this implies

$$\begin{aligned}\sum_{j=1}^{J} \sum_{k=l_j+1}^{m_j} \delta^{k-1} \lambda \cdot x_{s_{k+1}}(s_k, y_k) &\leq \frac{\kappa_0}{1-\delta} \sum_{j=1}^{J} \delta^{l_j} (1 - \delta^{(m_j - l_j)/2})^2 \\ &= \frac{\kappa_0}{1-\delta} \sum_{j=1}^{J} (\delta^{l_j/2} - \delta^{m_j/2})^2.\end{aligned}$$

The sum which appears in the last line is of the form

$$\sum_{j=1}^{J} (u_j - v_j)^2,$$

with $1 \geq u_1 \geq v_1 \geq \cdots \geq u_J \geq v_J \geq \delta^{n/2}$. Such a sum is maximized when $J = 1$, $u_1 = 1$ and $v_1 = \delta^{n/2}$. It is then equal to $(1 - \delta^{n/2})^2$. On the other hand, there are at most $|S|$ stages k with $m_j < k \leq l_{j+1}$ for some j. Therefore,

$$\begin{aligned}\frac{1-\delta}{\delta^{n-1}} \sum_{k=1}^{n-1} \delta^{k-1} \lambda \cdot x_{s_{k+1}}(s_k, y_k) &\leq |S| \kappa_0 \frac{1-\delta}{\delta^{n-1}} + \frac{\kappa_0}{\delta^{n-1}} (1 - \delta^{n/2})^2 \\ &\leq |S| \kappa_0 \frac{1-\delta}{\delta^{n-1}} + \frac{\kappa_0}{\delta^{n-1}} \left(\frac{n}{2}\right)^2 (1-\delta)^2 \\ &\leq 2|S| \kappa_0 \frac{1-\delta}{\delta^{n-1}}.\end{aligned}$$

Substituting this inequality into (13), one obtains

$$\begin{aligned}\lambda \cdot w(h_n) &\leq \lambda \cdot z + \frac{1-\delta^{n-1}}{\delta^{n-1}} \lambda \cdot (z - v) + \frac{1-\delta}{\delta^{n-1}} \sum_{k=1}^{n-1} \delta^{k-1} \lambda \cdot x_{s_{k+1}}(s_k, y_k) \\ &\leq \lambda \cdot z - \varepsilon_0 \frac{n-1}{2} \frac{1-\delta}{\delta^{n-1}} + 2|S| \kappa_0 \frac{1-\delta}{\delta^{n-1}} \\ &= \lambda \cdot z - \varepsilon \frac{1-\delta}{\delta^{n-1}} \leq \lambda \cdot z - \varepsilon(1-\delta)\end{aligned}$$

as claimed. ■

Let $\bar{\bar{\delta}} < 1$ be large enough so that $(1-\bar{\bar{\delta}})\kappa \leq \eta$ and $(1-\bar{\bar{\delta}})\kappa^2 \leq 2\varepsilon\eta$. The next lemma exploits the smoothness of Z_η.

Lemma 6 *For every $z \in Z_\eta$ and every $\delta \geq \bar{\bar{\delta}}$, there exists a direction $\lambda \in S^1$ such that, if $w \in \mathbf{R}^I$ satisfies $\|w - z\| \leq (1-\delta)\kappa$ and $\lambda \cdot w \leq \lambda \cdot z - (1-\delta)\varepsilon$, then one has $w \in Z_\eta$.*

Proof. By the definition of Z_η, for each $z \in Z_\eta$, there exists $z_0 \in Z$ such that $\|z - z_0\| \leq \eta$. Let $\lambda := (z - z_0)/\|z - z_0\|$. (If $z_0 = z$, then take any unit vector.) Then, for any w, one has

$$\begin{aligned}\|w - z_0\|^2 &= \|z - z_0\|^2 + 2(z - z_0) \cdot (w - z) + \|w - z\|^2 \\ &\leq \|z - z_0\|^2 - 2(1-\delta)\varepsilon\|z - z_0\| + (1-\delta)^2\kappa^2.\end{aligned}$$

The last expression is a quadratic form, which is maximized when $\|z - z_0\| = 0$ or $\|z - z_0\| = \eta$. Therefore,

$$\|w - z_0\|^2 \leq \max\{(1-\delta)^2\kappa^2, \eta^2 - 2(1-\delta)\varepsilon\eta + (1-\delta)^2\kappa^2\} \leq \eta^2,$$

because $\delta \geq \bar{\bar{\delta}}$. Thus $\|w - z_0\| \leq \eta$, hence $w \in Z_\eta$. ■

We here prove Proposition 2. Fix any $\delta \geq \max\{\bar{\delta}, \bar{\bar{\delta}}\}$. For any $z \in Z_\eta$, we construct a public strategy $\sigma : H \to \times_i \Delta(A^i)$ and continuation payoffs $w : H \to Z_\eta$ inductively as follows. Set $w(h) = z \in Z_\eta$ for any $h = (s_1) \in H_1$. For $k \geq 1$ and $h \in H_{(n-1)(k-1)+1}$, given that $w(h) \in Z_\eta$, by Proposition 4, there exist continuation payoffs $w_h : H_n \to \mathbf{R}^I$ that satisfy **C1** (that is, there exists a PPE of $\Gamma^n(s_{(n-1)(k-1)+1}, w_h; \delta)$, with payoff $w(h)$), **C2** and **C3**. Let σ prescribe the PPE of $\Gamma^n(s_{(n-1)(k-1)+1}, w_h; \delta)$ for the block of periods between $(n-1)(k-1)+1$ and $(n-1)k$. For any $\tilde{h} \in H_{(n-1)k+1}$ whose restriction to the first $(n-1)(k-1)+1$ periods is equal to h, let $w(\tilde{h}) = w_h({}_n\tilde{h})$, where ${}_n\tilde{h}$ is the restriction of $\tilde{h}$ to the last n periods. It follows from **C2**, **C3** and Lemma 6 that $w(\tilde{h}) \in Z_\eta$. By the one-shot deviation principle and **C1**, for any initial state s, the constructed strategy σ is a PPE of the whole (infinite-horizon) game, with payoff z. Thus $Z \subset Z_\eta \subseteq E_\delta(s)$ for any $s \in S$.

9.3 Proof of Lemma 2

We prove the two statements in turn. We start with the first one, and consider the following two cases: $\lambda \neq \pm e^i$ for any i, and $\lambda = e^i$.

Suppose that $\lambda \neq \pm e^i$ for any i. Let $\delta < 1$ be given. For each state s, choose a feasible payoff $\gamma_s \in F_\delta(s)$ that maximizes the inner product $\lambda \cdot \gamma_s$. Let $a_s \in A$ be the profile of moves played in stage 1, and $w_t(s,y) \in F_\delta(t)$ be the continuation payoffs following y, if in state t at stage 2. Note that $\lambda \cdot w_t(s,y) \leq \lambda \cdot \gamma_t$ for $y \in Y$, $s,t \in S$. By construction, for each $s \in S$, $\gamma_s = (1-\delta) r(s,a_s) + \delta \sum_{t \in S, y \in Y} p(t,y|s,a_s) w_t(s,y)$.

Fix $\varepsilon > 0$ arbitrarily. From Assumption **F2** and Lemma 6.2 of FLM, there exists an open and dense set of profiles each of which has pairwise full rank for all pairs of players. Therefore, for each $s \in S$, there exist $\hat{\gamma}_s$, $\hat{\alpha}_s$, and $\hat{w}_t(s,y)$ such that $\lambda \cdot \hat{w}_t(s,y) \leq \lambda \cdot \hat{\gamma}_t$ for $y \in Y$, $t \in S$, $\hat{\gamma}_s = (1-\delta) r(s,\hat{\alpha}_s) + \delta \sum_{t \in S, y \in Y} p(t,y|s,\hat{\alpha}_s) \hat{w}_t(s,y)$, $\lambda \cdot \hat{\gamma}_s \geq \lambda \cdot \gamma_s - \varepsilon$, and $\hat{\alpha}_s$ has pairwise full rank for all pairs of players in state s.

Similarly to the proof of Proposition 1, there exist $\tilde{x}_t(s,y)$ ($y \in Y$, $s,t \in S$) and $v \in \mathbb{R}^I$ such that $\lambda \cdot v = \min_s \lambda \cdot \hat{\gamma}_s$,

$$v = r(s,\hat{\alpha}_s) + \sum_t p(t,y|s,\hat{\alpha}_s) \tilde{x}_t(s,y) \ (s \in S),$$

and

$$\lambda \cdot \sum_{s \in T} \tilde{x}_{\phi(s)}(s, \psi(s)) \leq 0,$$

for each $T \subseteq S$, for each permutation $\phi : T \to T$ and each map $\psi : T \to Y$. For each $s \in S$, although $\hat{\alpha}_s$ is not a Nash equilibrium of $\Gamma(s,\tilde{x})$, since $\hat{\alpha}_s$ satisfies pairwise full rank, there exist $\hat{x}_t(s,y)$ for $y \in Y$ and $t \in S$ such that $\hat{\alpha}_s$ is a Nash equilibrium of $\Gamma(s, \tilde{x} + \hat{x})$ with payoff v, and such that $\lambda \cdot \hat{x}_t(s,y) = 0$ for each $y \in Y$ and $t \in S$. With $x_t(s,y) := \tilde{x}_t(s,y) + \hat{x}_t(s,y)$, the payoff vector v is a Nash equilibrium payoff of $\Gamma(s,x)$, and one has

$$\lambda \cdot \sum_{s \in T} x_{\phi(s)}(s, \psi(s)) \leq 0$$

for each $T \subseteq S$, each permutation $\phi : T \to T$, and each map $\psi : T \to Y$.

It follows that

$$k(\lambda) \geq \lambda \cdot v = \min_{s \in S} \lambda \cdot \hat{\gamma}_s \geq \min_{s \in S} \max_{w \in F_\delta(s)} \lambda \cdot w - \varepsilon.$$

Since $\varepsilon > 0$ and $\delta < 1$ are arbitrary,

$$k(\lambda) \geq \sup_{\delta < 1} \min_{s \in S} \max_{w \in F_\delta(s)} \lambda \cdot w \geq \min_{s \in S} \max_{w \in F(s)} \lambda \cdot w.$$

Conversely, we have

$$\max_{w\in F(s)} \lambda\cdot w = \sup_{\alpha}\sum_{t\in S}\mu_{\alpha,s}(t)\lambda\cdot r(t,\alpha_t),$$

where $\mu_{\alpha,s}(t)$ is the long-run frequency of state t under the Markov chain induced by α with initial state s, and where the supremum is taken over all Markov strategy profiles $\alpha=(\alpha_t)_{t\in S}$. Thus

$$\min_{s\in S}\max_{w\in F(s)} \lambda\cdot w \geq \sup_{\alpha}\min_{\mu_\alpha}\sum_{t\in S}\mu_{\alpha}(t)\lambda\cdot r(t,\alpha_t),$$

where the minimum is taken over all invariant measures of the Markov chain induced by α.

By Lemma 1, for each α, one has

$$\sum_{t\in S}\mu_{\alpha}(t)\lambda\cdot r(t,\alpha_t)\geq k(\lambda,\alpha).$$

for any invariant measure μ_α. Hence,

$$\min_{s\in S}\max_{w\in F(s)} \lambda\cdot w \geq \sup_{\alpha} k(\lambda,\alpha)=k(\lambda).$$

Suppose next that $\lambda=e^i$ for some i. By replacing Assumption **F2** with Assumption **F1** and Lemma 6.2 of FLM with Lemma 6.3 of FLM, a similar argument establishes $k(e^i)=\min_{s\in S}\max_{w\in F(s)} w^i$. This concludes the proof of the first statement.

Consider $\lambda=-e^i$ for some i. Let $\delta<1$ be given. There exists a minmaxing (public) profile σ^{-i} such that

$$m^i_\delta(s)=\max_{a^i_s\in A^i}\left((1-\delta)r^i(s,a^i_s,\alpha^{-i}_s)+\delta\sum_{t\in S,y\in Y}q(t|s,a^i_s,\alpha^{-i}_s)w^i_t(s,y)\right),$$

for every $s\in S$, where α^{-i}_s is the mixed moves played in stage 1 (as a function of the initial state) and $w^i_t(s,y)$ is a continuation payoff for player i when player i takes a best-response against σ^{-i} from stage 2 onward. Note that $w^i_t(s,y)\geq m^i_\delta(t)$ for each t. Therefore, by replacing Assumption **F2** with Assumption **F1** and Lemma 6.2 of FLM with Lemma 6.3 of FLM, a similar argument establishes $k(-e^i)\geq -\max_s m^i(s)$.

Conversely, we have

$$m^i(s)\leq \inf_{\alpha^{-i}}\sup_{\alpha^i}\sum_{t\in S}\mu_{\alpha,s}(t)r^i(t,\alpha_t),$$

where α^i and α^{-i} are Markov strategies and $\mu_{\alpha,s}(t)$ is the long-run frequency of state t under the Markov chain induced by α with initial state s. Thus

$$\max_{s\in S} m^i(s) \leq \inf_{\alpha^{-i}} \sup_{\alpha^i} \max_{\mu_\alpha} \sum_{t\in S} \mu_\alpha(t) r^i(t, \alpha_t),$$

where the maximum is taken over all invariant measures of the Markov chain induced by α.

Let $(\tilde{\alpha}^i, \alpha^{-i})$ be an arbitrary Markov strategy, and $\varepsilon > 0$. Then there exists (v, x) that is feasible in $\mathcal{P}(-e^i, (\tilde{\alpha}^i, \alpha^{-i}))$ and $-v_i > k(-e^i, \tilde{\alpha}^i, \alpha^{-i}) - \varepsilon$. It follows from the incentive constraints in $\mathcal{P}(-e^i, (\tilde{\alpha}^i, \alpha^{-i}))$ that, for each $s \in S$, one has

$$v^i \geq r^i(s, \alpha_s^i, \alpha_s^{-i}) + \sum_{t\in S, y\in Y} p(t, y|s, \alpha_s^i, \alpha_s^{-i}) x_t^i(s, y).$$

By Lemma 1, it thus follows that for each $(\tilde{\alpha}^i, \alpha^{-i})$,

$$\sum_{s\in S} \mu_{\alpha^i,\alpha^{-i}}(s) r^i(s, \alpha_s^i, \alpha_s^{-i}) \leq v_i < -k(-e^i, \tilde{\alpha}^i, \alpha^{-i}) + \varepsilon,$$

for any α^i and any invariant measure $\mu_{\alpha^i,\alpha^{-i}}$. Since $(\tilde{\alpha}^i, \alpha^{-i})$ and $\varepsilon > 0$ are arbitrary,

$$\max_{s\in S} m^i(s) \leq \inf_{\alpha^{-i}} \inf_{\tilde{\alpha}^i} (-k(-e^i, \tilde{\alpha}^i, \alpha^{-i})) = -k(-e^i).$$

This concludes the proof of the second statement.

9.4 Proof for the Characterization with Full Monitoring

We first show that $\mathcal{P}(e^i)$ and $\mathcal{Q}_+^i$ are equivalent. Take any (α, v, x) feasible in $\mathcal{P}(e^i)$. We have

$$v^i = r^i(s, a_s^i, \alpha_s^{-i}) + \sum_{t\in T} \sum_{a_s^{-i}\in A^{-i}} \alpha_s^{-i}(a_s^{-i}) p(t|s, a_s^i, a_s^{-i}) x_t^i(s, a_s^i, a_s^{-i})$$

for every $s \in S$ and $a_s^i \in \text{support}(\alpha_s^i)$. For each s, set

$$\tilde{x}_t^i(s) := \max_{a_s^i \in \text{support}(\alpha_s^i)} \max_{a_s^{-i}\in A^{-i}} x_t^i(s, a_s^i, a_s^{-i}),$$

and

$$\tilde{v}_s^i := \min_{a_s^i \in \text{support}(\alpha_s^i)} \left(r^i(s, a_s^i, \alpha_s^{-i}) + \sum_{t \in T} p(t|s, a_s^i, \alpha_s^{-i}) \tilde{x}_t^i(s) \right).$$

Next, set $\bar{x}_t^i(s) := \tilde{x}_t^i(s) + v^i - \tilde{v}_s^i$. Since $\tilde{v}_s^i \geq v^i$ for every $s \in S$, we have that $(\alpha, v^i, \bar{x}^i)$ is feasible in $\mathcal{Q}_+^i$.

Conversely, take any $(\alpha, v^i, \bar{x}^i)$ feasible in $\mathcal{Q}_+^i$. For each $s \in S$ and $a_s^{-i} \in A^{-i}$, we set

$$x_t^i(s, a_s^i, a_s^{-i}) := \bar{x}_t^i(s) + v^i - \left(r^i(s, a_s^i, \alpha_s^{-i}) + \sum_{t \in T} p(t|s, a_s^i, \alpha_s^{-i}) \tilde{x}_t^i(s) \right)$$

if $a_s^i \in \text{support}(\alpha_s^i)$ (note that the right-hand side is independent of a_s^{-i}), and we let $x_t^i(s, a_s^i, a_s^{-i})$ be sufficiently small if $a_s^i \notin \text{support}(\alpha_s^i)$. Since $v^i \leq r^i(s, a_s^i, \alpha_s^{-i}) + \sum_{t \in T} p(t|s, a_s^i, \alpha_s^{-i}) \tilde{x}_t^i(s)$ for every $a_s^i \in \text{support}(\alpha_s^i)$, we have that $(\alpha, v^i, v^{-i}, x^i, x^{-i})$ is feasible in $\mathcal{P}(e^i)$. (v^{-i} and x^{-i} are chosen to give incentives for players $-i$ to play α^{-i}.)

The equivalence between $\mathcal{P}(-e^i)$ and $\mathcal{Q}_-^i$ follows similarly.

9.5 Proof of Proposition 3

We let a unit direction $\lambda \in S^1$, and a mixed profile α be given. We prove that for each feasible pair (v, x) in $\mathcal{P}_p(\lambda, \alpha)$, there is $z : S \times Y \to \mathbf{R}^{S \times I}$ such that (v, z) is a feasible pair in $\mathcal{P}_q(\lambda, \alpha)$. This implies $k_p(\lambda, \alpha) = k_q(\lambda, \alpha)$, and hence the result. The proof uses the following two observations.

Denote by P (resp. Q) the $S \times S$ matrix whose (s,t)-th entry is $p(t|s)$ (resp. $q(t|s)$).

Claim 2 *The ranges of the two linear maps (with matrices) $Id - P$ and $Id - Q$ are equal.*

Proof of the Claim. Assume that q is aperiodic. Let be given a vector $\xi \in \mathbf{R}^S$ in the range of $Id - P$. Hence, there is $\psi \in \mathbf{R}^S$, such that $\xi = \psi - P\psi$. Hence, $\langle \mu, \xi \rangle = 0$. Since q is an irreducible transition function with invariant measure μ, the vector $Q^n \xi$ converges as $n \to \infty$ to the constant vector, whose components are all equal to $\langle \mu, \xi \rangle = 0$. Moreover, the convergence takes place at an exponential speed. Therefore, the vector $\tilde{\psi} := \sum_{n=0}^{+\infty} Q^n \xi$ is well-defined, and solves $\tilde{\psi} = \xi + Q\tilde{\psi}$. This proves the result.[14] ■

[14]The proof in the case where q is periodic is similar. As in the aperiodic case, the infinite sum in $\tilde{\psi}$ is well-defined.

Claim 3 *Let a vector $c = (c_t(s)) \in \mathbf{R}^{S\times S}$ be given, such that the inequality $\sum_{s\in T} c_{\phi(s)}(s) \leq 0$ holds for every $T \subseteq S$, and every permutation ϕ over T. Then there exists a vector $c^* = (c^*_t(s)) \in \mathbf{R}^{S\times S}$ such that (i) $c^* \geq c$ (componentwise) and (ii) $\sum_{s\in T} c^*_{\phi(s)}(s) = 0$, for every $T \subseteq S$ and every permutation ϕ over T.*

Proof of the Claim. Consider the linear program $\sup \sum_{T,\phi} \left(\sum_{s\in T} g_{\phi(s)}(s) \right)$, where the sum ranges over all (T, ϕ) such that $T \subseteq S$ and ϕ is a permutation over T, and the supremum is taken over all g such that (i) $g \geq c$ and (ii) $\sum_{s\in T} g_{\phi(s)}(s) \leq 0$, for every $T \subseteq S$ and every permutation ϕ over T. This program is bounded and c is a feasible point, hence there is an optimal solution, c^*. We claim that the value of the program is 0. Assume to the contrary that

$$\sum_{s\in T} c^*_{\phi(s)}(s) < 0 \tag{14}$$

for some T and some ϕ. It must then be the case that for each $s \in T$, there exists some set $T_s \subseteq S$ containing s, and some permutation ϕ_s over T_s with $\phi_s(s) = \phi(s)$, and such that the constraint $\sum_{t\in T_s} c^*_{\phi_s(t)}(t) \leq 0$ is binding. (Otherwise indeed, a small increase in the value of $c^*_{\phi(s)}(s)$ would preserve feasibility of c^*, and improve upon c^*.) In particular,

$$\sum_{t\in T_s, t\neq s} c^*_{\phi_s(t)}(t) = -c^*_{\phi(s)}(s).$$

When summing over $s \in T$, and by (14), one obtains

$$\sum_{s\in T} \sum_{t\in T_s, t\neq s} c^*_{\phi_s(t)}(t) > 0. \tag{15}$$

We claim that the directed graph over S with edge set $\bigcup_{s\in T} \bigcup_{t\in T_s, t\neq s} (t, \phi_s(t))$ (where any edge that appears more than once is repeated) is a collection of disjoint cycles. Hence (15) is in contradiction with the fact that c^* is feasible.

To see this last claim, observe that for each state $s \in T$, the edges $(t, \phi_s(t))$, where $t \in T_s$, $t \neq s$, form a directed path from $\phi(s)$ to s. Hence, the union over s of these paths is a union of disjoint cycles. ■

We turn to the proof of Proposition 3. Let (v, x) be a feasible pair in $\mathcal{P}_p(\lambda, \alpha)$. For $(s, t) \in S$, we set $c_t(s) := \max_{y\in Y} \lambda \cdot x_t(s, y)$. Apply Claim 3 to obtain $(c^*_t(s))$. Since $\sum_{s\in T} c^*_{\phi(s)}(s) = 0$, for

every $T \subseteq S$ and every permutation ϕ over T, there is a vector $\bar{c}^* \in \mathbf{R}^S$, such that $c_t^*(s) = \bar{c}_t^* - \bar{c}_s^*$ for every $s, t \in S$.

Next, by Claim 2, there is $\bar{d} \in \mathbf{R}^S$, such that $(Id - P)\bar{c}^* = (Id - Q)\bar{d}$. For $s, t \in S$, set $d_t(s) := \bar{d}_t - \bar{d}_s$. Observe that, by construction, one has $\sum_{t \in S} p(t|s)c_t^*(s) = \sum_{t \in S} q(t|s)d_t(s)$ for each $s \in S$.

Finally, we set

$$z_t^i(s, y) := \lambda^i d_t(s) + \sum_{u \in S} p(u|s) \left(x_u^i(s, y) - \lambda^i c_u^*(s) \right),$$

for any i, s, t, y. We claim that (v, z) is feasible in $\mathcal{P}_q(\lambda, \alpha)$, as desired.

By construction, one has

$$\lambda \cdot z_t(s, y) = d_t(s) + \sum_{u \in S} p(u|s) \left(\lambda \cdot x_u(s, y) - c_u^*(s) \right) \leq d_t(s),$$

hence the vector z satisfies all linear constraints in $\mathcal{P}_q(\lambda, \alpha)$.

On the other hand, and for each y, one has

$$\begin{aligned} \sum_{t \in S} q(t|s) z_t^i(s, y) &= \sum_{u \in S} p(u|s) x_u^i(s, y) + \lambda^i \left(\sum_{t \in S} q(t|s) d_t(s) - \sum_{u \in S} p(u|s) c_u^*(s) \right) \\ &= \sum_{t \in S} p(t|s) x_t^i(s, y). \end{aligned}$$

Hence the equality

$$\sum_{t \in S, y \in Y} q(t|s) \pi(y|s, a) z_t^i(s, y) = \sum_{t \in S, y \in Y} p(t|s) \pi(y|s, a) x_t^i(s, y)$$

holds for any $a \in S$. This concludes the proof of Proposition 3.

References

[1] Abreu, D., D. Pearce, and E. Stacchetti (1990). "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, **58**, 1041–1063.

[2] Bewley, T. and E. Kohlberg (1976). "The Asymptotic Theory of Stochastic Games," *Mathematics of Operations Research*, **1(3)**, 197–208.

[3] Dixit, A., Grossman G. M. and F. Gul (2000). "The Dynamics of Political Compromise," *Journal of Political Economy*, **108**, 531–568.

[4] Dutta, P. K. (1995). "A Folk Theorem for Stochastic Games," *Journal of Economic Theory*, **66**, 1–32.

[5] Ericson, R. and A. Pakes (1995). "Markov Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, **62**, 53–82.

[6] Fink, A. M. (1964). "Equilibrium in a stochastic n-person game," *Journal of Science of the Hiroshima University, Series A-I*, **28**, 89–93.

[7] Freidlin, M. I. and A. D. Wentzell (1991). *Random Perturbations of Dynamical Systems*, Springer-Verlag, New York.

[8] Fudenberg, D. and E. Maskin (1986). "The Folk Theorem for Repeated Games with Discounting or with Incomplete Information," *Econometrica* , **54**, 533–554.

[9] Fudenberg, D. and D. Kreps and E. Maskin (1990). "Repeated Games with Long-run and Short-run Players," *Review of Economic Studies*, **57**, 555–573.

[10] Fudenberg, D. and D. Levine (1994). "Efficiency and Observability with Long-Run and Short-Run Players, *Journal of Economic Theory*, **62**, 103–135.

[11] Fudenberg, D., D. Levine, and E. Maskin (1994). "The Folk Theorem with Imperfect Public Information," *Econometrica*, **62**, 997–1040.

[12] Fudenberg, D. and Y. Yamamoto (2010). "The Folk Theorem for Irreducible Stochastic Games with Imperfect Public Monitoring," mimeo, Harvard University.

[13] Green, E. J. and R. H. Porter (1984). "Noncooperative Collusion under Imperfect Price Information," *Econometrica*, **52**, 87–100.

[14] Hoffman, A. J. and R. M. Karp (1966). "On Nonterminating Stochastic Games," *Management Science*, **12**, 359–370.

[15] Judd, K. L., S. Yeltekin and J. Conklin (2003). "Computing Supergames Equilibria," *Econometrica*, **71**, 1239–1254.

[16] Kallenberg, L. (2002). "Finite State and Action MDPs," in *Handbook of Markov Decision Processes Methods and Applications*, E. A. Feinberg and A. Shwartz (eds.), Kluwer Academic, 21–88.

[17] Lagunoff, R. and A. Matsui (1997). "Asynchronous Choice in Repeated Coordination Game," *Econometrica*, **65**, 1467–1477.

[18] Mailath, G. J. and L. Samuelson (2006). *Repeated Games and Reputations: Long-Run Relationships*, Oxford University Press, New York.

[19] Mertens, J.-F. and A. Neyman (1981). "Stochastic Games," *International Journal of Game Theory,* **10**, 53–66.

[20] Mertens, J.-F. and T. Parthasarathy (1987). "Equilibria for Discounted Stochastic Games," C.O.R.E. Discussion Paper 8750.

[21] Mertens, J.-F. and T. Parthasarathy (1991). "Nonzero-sum stochastic games." In: T. E. S. Raghavan, T. S. Ferguson, T. Parthasarathy and O. J. Vrieze, Editors, *Stochastic Games and Related Topics*, Kluwer Academic, Boston.

[22] Neyman, A. (2003). "Stochastic Games: Existence of the Minmax," in *Stochastic Games and Applications*, A. Neyman and S. Sorin (eds.), NATO ASI series, Kluwer Academic Publishers, 173–193.

[23] Nummelin, E. (1984). *General irreducible Markov Chains and Non-Negative Operators*, Cambridge University Press, Cambridge.

[24] Phelan, C. (2006). "Public Trust and Government Betrayal," *Journal of Economic Theory*, **130**, 27–43.

[25] Phelan, C. and E. Stacchetti (2001). "Sequential Equilibria in a Ramsey Tax Model," *Econometrica*, **69**, 1491–1518.

[26] Radner, R., R. Myerson and E. Maskin (1986). "An Example of a Repeated Partnership Game with Discounting and with Uniformly Inefficient Equilibria," *Review of Economic Studies*, **53**, 59–69.

[27] Rotemberg, J. J. and G. Saloner (1986). "A Supergame-Theoretic Model of Price Wars during Booms," *American Economic Review*, **76**, 390–407.

[28] Sennott, L. I. (1998). *Stochastic Dynamic Programming and the Control of Queuing Systems*, Wily Interscience, New York.

[29] Solan, E. (1998). "Discounted Stochastic Games," *Mathematics of Operations Research*, **23**, 1010–1021.

[30] Sorin, S. (1986). "Asymptotic Properties of a Non-Zero Sum Stochastic Game," *International Journal of Game Theory*, **15**, 101–107.

[31] Takahashi, M. (1962). "Stochastic Games with Infinitely Many Strategies," *Journal of Science of the Hiroshima University, Series A-I*, **26**, 123–124.

[32] Yoon, K. (2001). "A Folk Theorem for Asynchronously Repeated Games," *Econometrica*, **69**, 191–200.